Understanding the Interplay Between Thermodynamics and Surface Kinetics in the Growth of Dilute Nitride Alloys From First Principles

Hazem Abu Farsakh

Understanding the Interplay Between Thermodynamics and Surface Kinetics in the Growth of Dilute Nitride Alloys From First Principles

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

Hazem Abu Farsakh

Promotionskommission:

Vorsitzender	Prof. Dr. phil. Klaus Lischka
Gutachter	Prof. Dr. rer. nat. Wolf Gero Schmidt
Gutachter	Prof. Dr. rer. nat. Jörg Neugebauer
Beisitzer	Dr. rer. nat. Eva Rauls

Tag der Einreichung: 17.05.2010
Tag der mündlichen Prüfung: 06.07.2010

To my wife Lara and my daughter Jasmin

Acknowledgments

This work would not have been possible without a number of people to whom I would like to express my sincere thanks.

I would like first to express my deepest gratitude to my supervisor Prof. Dr. Jörg Neugebauer for giving me the excellent opportunity to work in this exciting project and for providing an excellent working environment. Jörg has always offered his full support, guidance, encourage, and understanding. Throughout many fruitful discussions with him I have gained a lot of knowledge and learned how to challenge obstacles in scientific research.

I would like to thank Dr. Martin Albrecht and his group at the Institut für Kristallzüchtung in Berlin for the experimental collaboration and the fruitful discussions.

I would like to thank my colleagues at the department of computational materials design at the Max-Planck-Institut für Eisenforschung in Düsseldorf for their encouragement and kind support. In particular, I would like to thank Sixten Boeck who introduced me to the field of code development. I would like to thank Blazej Grabowski for providing his help in many things specially when I first arrived to Germany. I would also like to thank Fritz Körmann and Ugur Aydin for their efforts in translating the summary to German.

I would like to thank my parents who supported my interest in making my PhD study in a place far away from them. I am grateful for the love and support of my parents, sisters and brothers. I have missed you all a lot during this period. I am specially deeply grateful to my wife Lara for her love, understanding, and endless care and support during this work. Thank you for being always beside me. Finally, I thank my daughter Jasmin for the joyful time I spent with her.

Düsseldorf, May 2010

Zusammenfassung

Seit einigen Jahren gibt es großes Interesse an “dilute nitride alloys” - Gruppe III-V Halbleitern mit geringer Beimischung an Stickstoff. Die Mischung von konventionellen III-V Halbleitern mit N führt zu einer starken Verbiegung der Bandlücke und eröffnet damit einen viel versprechenden Ansatz, die optischen Eigenschaften zu kontrollieren. Diese einzigartige Eigenschaft macht solche Legierungen interessant für verschiedenste Anwendungen von Infrarot-Lasern bis hin zu Tandem-Photovoltaik-Solarzellen. Während ursprünglich die Realisierung solcher Legierungen als einfach angenommen wurde, stellte sich die praktische Realisierung mit Hilfe von Kristallwachstums-Techniken als überaus schwierig heraus. Ein großes Problem stellt die extrem geringe Löslichkeit von N in III-V Halbleitern im thermodynamischen Gleichgewicht dar. Weitere Probleme entstehen durch die starke räumliche Verteilung sowie der Tendenz zur Phasenseparation und 3D Inselwachstum. Während eine Vielzahl experimenteller Anstrengungen entsprechenden Problemlösungen gewidmet ist, blieb ein Verständnis der grundlegenden Mechanismen dieser Wachstumsschwierigkeiten bisher aus.

Um die N-Konzentration zu erhöhen, die Qualität zu verbessern sowie die lokale Anordnung der Atome zu kontrollieren, ist ein theoretisches Verständnis der grundlegenden physikalischen Mechanismen, die das Wachstum der III-V verdünnten Stickstoff Legierungen auf atomarer Ebene beeinflussen, erforderlich. Hochmoderne *ab-initio* Simulationen bieten einen leistungsfähigen und viel versprechenden Zugang um dieses Ziel zu erreichen. Zu diesem Zwecke sind großskalige Simulationen solcher Mehr-Komponenten-Legierungen wichtig. Jedoch sind für Wachstumssimulationen unter realistischen Systemsgrößen die ausschließlich *ab-initio* basierten Methoden zu rechenintensiv. Das Hauptziel dieser Arbeit ist daher die Entwicklung der notwendigen Methoden und Konzepte um eine realistische Beschreibung durch Simulationen von kleineren Teilsystemen zu garantieren und mit diesen ein genaues Verständnis und Vorhersagen zu erzielen.

Im ersten Teil behandeln wir das Problem der beschränkten Löslichkeit von Stickstoff und diskutieren Ansätze um diese zu erhöhen. Insbesondere konzentrieren wir uns auf die Möglichkeiten, die durch die erhöhte Löslichkeit an der Oberfläche gegeben sind und wie diese in der Praxis umgesetzt werden können. Eine genaue Beschreibung der N-Konzentration als Funktion der Wachstumsbedingungen sowie der Löslichkeitsgrenzen wird durch die Berechnung detaillierter Phasen-Stabilitäts-Diagramme basierend auf der *ab-initio* Thermodynamik und in Kombination mit Monte-Carlo-Simulationen für die N-Substitutionsmöglichkeiten in GaAs und InAs Kristallen sowie den Ober-/Suboberflächen erzielt. Im Gegensatz zum Kristall finden wir, dass die Löslichkeit von N an den Oberflächen wesentlich größer ist sowie ein komplexeres Verhalten als Funktion der Wachstumsbedingungen aufweist. Ausgehend von diesen Resultaten

werden optimale Wachstumsbedingungen für die Stickstoff-Beimischung und kinetische Pfade der Adatomdiffusion vorhergesagt, um eine maximale Stickstoff-Konzentration zu erzielen.

Anschließend wird die Oberflächen-Kinetik analysiert. In einem ersten Schritt konzentrieren wir uns darauf, ein komplettes Verständnis der N-Adatom-Diffusion auf der GaAs Oberfläche zu erhalten, was sich als besonders herausforderndes Problem darstellt. Das komplizierte Diffusionsverhalten führt zunächst zu einem Zusammenbruch der konventionellen Herangehensweisen und üblichen Methoden zur Diffusions-Analyse. Ausgehend von einer gründlichen Analyse identifizieren wir die zugrunde liegenden Probleme und entwickeln entsprechende Lösungen. Im zweiten Schritt betrachten wir mögliche kinetische Pfade für die Inkorporation von Stickstoff auf und unterhalb der Oberfläche. Für diese Oberflächen werden die relevanten Diffusionspfade im Detail studiert sowie die damit verbundenen Barrieren und die korrespondierenden Reaktionsraten unter typischen Wachstumstemperaturen bestimmt. Basierend auf diesen Resultaten und mit Hilfe von kinetischen Monte-Carlo-Simulationen konnte die Diffusion sowie der Einbau von Stickstoffatomen vollständig analysiert und verstanden werden. Eine wichtige Schlussfolgerung dieser Ergebnisse ist, dass im Gegensatz zum bisherigen Verständnis, der Einbau von Stickstoff in der Suboberfläche nicht unter typischen MBE Wachstumsbedingungen realisierbar ist, da der Einbau auf der Oberfläche kinetisch deutlich bevorzugt ist.

Die in dieser Arbeit erzielten Ergebnisse ermöglichen die Ableitung wichtiger Schlussfolgerungen: Durch das Aufzeigen der aktiven Einbaumechanismen und basierend auf den Löslichkeits-Phasen-Diagrammen lassen sich die Wachstumsbedingungen identifizieren, die im Experiment realisiert werden sollten um den Stickstoff-Einbau zu maximieren. Eine zweite wichtige Schlussfolgerung, die sich aus dem hier gewonnenen Verständnis der Wechselwirkung zwischen In- und den N-Atomen in quarternären Stickstoff-Legierungen ergibt, ist die Identifizierung der Triebkräfte für die bisher unverstandene Korrelationen und Antikorrelationen der chemischen Zusammensetzung zu identifizieren. Auf Grund der in dieser Arbeit durchgeführten Simulationen gelang es damit erstmals einen detaillierten Einblick in die Wachstumsprozesse von dilute nitrides zu erhalten. Abschließend sei angemerkt, dass die hier entwickelten und verwendeten Methoden allgemein anwendbar sind – das hier studierte Legierungssystem stellt einen Prototyp für stark gitterfehlangepasste Mehr-Komponenten-Systeme dar, die mit konventionellen theoretischen Ansätzen nicht bzw. nur sehr approximativ beschreibbar sind.

Summary

Dilute nitride alloys, obtained by incorporating a small amount of nitrogen in III-V semiconductors, have attracted a considerable interest in the past few years. Alloying of conventional III-V semiconductors with N results in a large band-gap bowing and thus offers an interesting route to control optical properties. This unique property made this alloy a promising material for several applications ranging from infrared and mid-infrared laser diodes to multi-junction photovoltaic solar cells. While originally expected to be easy to realize, the actual fabrication of such alloys by employing crystal growth techniques has been challenging in practice. A major challenge for the realization of these alloys is the extremely low equilibrium N solubility in III-V semiconductors. Other challenges include large spatial composition fluctuations and the tendency for phase separation and 3D island growth. While numerous experimental efforts have been devoted to overcome these challenges, an understanding of the basic reasons leading to these growth difficulties is so far lacking.

In order to enhance N incorporation, improve quality, and control the local order of atoms, a theoretical understanding of the basic physical mechanisms that control the growth of III-V dilute nitride alloys at the atomic level is therefore crucial. State-of-the-art first-principles (*ab-initio*) simulations provide a powerful and a promising approach to achieve this goal. To accurately describe these multi-component alloys, large-scale simulations are highly desirable. However, performing growth simulations for a realistic system size fully based on *ab-initio* methods is computationally prohibitive. The main objective in this thesis is therefore the development of the necessary methodologies and concepts that allow us to overcome this challenge and obtain an accurate understanding and predictions based on small-scale atomistic simulations.

In the first part we address the issue of limited N solubility and discuss approaches to increase it. Particularly, we focus on the possibilities offered by enhanced surface solubility and how it can be employed in practice. By calculating the detailed stability phase diagrams for N substitutions in bulk GaAs and InAs and at surfaces and subsurfaces using *ab-initio* thermodynamics combined with Monte-Carlo simulations we provide an accurate measure of N concentration as a function of growth conditions and determine the solubility limits. Contrary to bulk, we find that the solubility of N at surfaces is significantly larger and shows a complex behavior as a function of growth conditions. Based on these results, optimal growth conditions for N incorporation and the kinetic pathway that should be employed to achieve maximum N concentration are predicted.

After that we focus on surface kinetics. Specifically, in a first step we concentrate on achieving a complete understanding of the diffusion of the N adatoms at the GaAs surface, which turns out to be a particularly challenging problem. The complex diffusion behavior leads for the first

time to the breakdown of the conventional approaches and methods usually employed to analyze diffusion. Based on a careful analysis we identify the underlying reasons and develop appropriate techniques to overcome this problem. In a second step we consider the possible kinetic pathways for surface/subsurface incorporation of N. We examine these pathways in detail and determine the associated barriers and the corresponding reaction rates under typical growth temperatures. Based on these results and with the help of kinetic Monte Carlo simulations we are able to fully analyze and understand N diffusion and their incorporation mechanisms. In contrast to what was previously believed, we find that the subsurface incorporation of N cannot be realized under typical MBE growth conditions, while their incorporation at the topmost surface layer is easily achievable.

Important consequences are realized based on these results. First, by revealing the functional N incorporation mechanism and based on the solubility phase diagrams, the growth conditions that should be employed in practice to achieve the highest N concentration are identified. A second consequence is that by additionally understanding the interaction behavior between In and N atoms in the quaternary dilute nitride alloys we are able to identify the driving mechanisms for the puzzling compositional correlations and anitcorrelations. This work allows to achieve a full understanding of the growth of dilute nitride alloys for the first time. We note that the methods derived and employed here are general – the alloy system studied here thus provides a prototype for understanding other highly-mismatched multi-component systems that are difficult to describe by conventional theoretical approaches.

Contents

Chapter 1

Introduction

Over the last decades, a tremendous growth of internet and telecommunications has taken place. For instance, the worldwide growth of internet usage in a 9-years period (2000-2009) is $\sim 400\,\%$ [1]. This huge growth drives an increasing demand for highly efficient technologies for data transmission. Since their development, fiber-optics have revolutionized telecommunications industry. Light waves have an enormous potential for data transmission due to their high frequency, leading to orders of magnitude higher data transfer rates than that of any electrical cable. For the "last-mile" connections to homes and offices (local-area networks (LAN)) and metropolian-areas network connections (MAN) the $1.3 - 1.55\,\mu$m wavelength range is appealing to minimize attenuation (i.e. intensity loss) and/or chromatic dispersion in the used silica optical fibers [2], which has triggered the research and development of laser diodes operating in this infrared wavelength range.

While InP-based infrared lasers have been used for these applications, the fabrication of GaAs-based lasers has been particularly promising to achieve a lower cost and a superior performance. Although the emission of GaAs lies outside the interesting range of wavelengths for these lasers, it can be tuned by employing band gap engineering using modern growth techniques. In particular, it has been found that substituting a low fraction ($x < 5\%$) of arsenic atoms by nitrogen atoms in GaAs results in a large band gap bowing, i.e., a rapid decrease of the band gap with increasing N concentration [3, 4, 5]. The powerful handle over the optical properties of GaAs alloys with dilute N concentrations ($GaAs_{1-x}N_x$) has made them appealing also for other applications such as using them in high efficiency multi-junction photovoltaic solar cells for space satellites [3, 6, 7, 8].

Apart from $GaAs_{1-x}N_x$, other highly mismatched alloys have also prompted a great deal of interest. The common characteristic of this new class of alloys is that the incorporation of low fractions of atoms with large size or electronegativity mismatch to the host atoms results in dramatic changes in their physical properties. Examples include II-VI_{1-x}-O_x [9, 10] and other III-V dilute nitride alloys (III-V_{1-x}-N_x) [11, 12, 3, 13, 6] such as $InAs_{1-x}N_x$ alloys. $InAs_{1-x}N_x$ alloys are relevant for mid-infrared laser diodes operating in the $2 - 5\,\mu$m range [14, 12, 11], which have several applications in medicine and in highly sensitive trace gas spectroscopy e.g. for monitoring atmospheric pollution [15].

The band gap bowing in dilute nitrides is illustrated in Fig. 1.1, which shows the band gap versus the lattice constant of a few selected semiconductor alloys. The solid lines indicate the

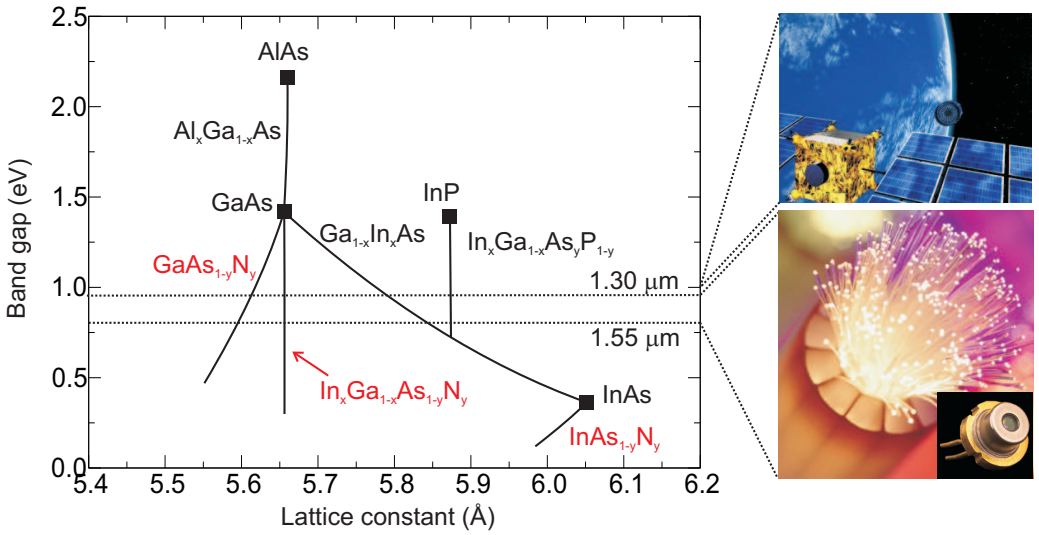


Figure 1.1: Illustration of the band gap bowing in dilute nitride alloys. The plot shows the band gap versus the lattice constant of selected semiconductors. The solid lines indicate the change with impurity concentration. The horizontal dotted lines show the band gap region relevant for infrared laser diodes ($1.3 - 1.55\,\mu$m) for fiber optics-based data transmission. The pictures on the right hand side illustrate some applications of dilute nitrides; upper: photovoltaic solar cells for space satellites, lower: fiber optics data transmission, with a laser diode shown.

effect of the incorporation of impurities as indicated. For the $GaAs_{1-x}N_x$ material system, the incorporation of N in GaAs results in a strong reduction of the band gap, as shown. For example, only 1 % of N in GaAs results in a strikingly large reduction of 180 meV in the fundamental band gap energy [16, 17]. We note that the phenomena of the large band gap bowing has been explained in terms of the band anti-crossing model [18, 19], which has also succeeded to explain the electronic structure in many highly mismatched systems [9, 20].

Since matching the lattice constant between the grown material and the substrate is important in the epitaxial growth e.g. to prevent strain in the grown films, it is more beneficial to grow $GaAs_{1-x}N_x$ lattice matched to GaAs. The incorporation of N atoms, which are smaller than As atoms, causes a tensile strain in the neighboring GaAs bonds, driving for a smaller lattice constant. As can be seen from Fig. 1.1, a compensation/cancellation of the strain can be realized by substituting a fraction of gallium atoms in $GaAs_{1-x}N_x$ alloys by indium atoms, having a larger atomic radius. This makes it possible to tune the lattice constant and the band gap independently. Therefore, $In_yGa_{1-y}As_{1-x}N_x$ alloys that are lattice matched to GaAs with the band gap in the infrared region can be realized. The growth of these quaternary $In_yGa_{1-y}As_{1-x}N_x$ alloys was first proposed by Kondow et al. [21, 22], which has generated an enormous interest since then.

1.1 Progress and challenges in dilute nitride alloys

The growth of high quality dilute nitride alloys has been hampered by several challenges. A major challenge is the extremely low N solubility in GaAs, which was first estimated to be

$\sim 10^{-7}$ % at 1000 K [23, 24]. The low equilibrium N solubility is attributed to the large lattice mismatch between GaAs and GaN of about 20% [4]. Substituting an As atom by a N causes a large tensile strain in the alloy around the N atom, which makes it an energetically highly unfavorable process.

Other challenges include compositional fluctuations and inhomogeneities, interface undulations, a tendency for 3D island growth, and phase separation [25, 26, 27]. All of these mechanisms may cause a degradation of material quality. A specific example of these problems is seen in the compositional anticorrelation between In and N in the quaternary $In_yGa_{1-y}As_{1-x}N_x$ alloys: While bulk thermodynamic considerations predict that forming In-N bonds is energetically more favorable due to strain compensation, experimental results show that the as-grown alloys contain more Ga-N bonds than In-N bonds [28, 29, 30]. A major difference of these quaternary alloys to the commonly employed ternary alloys is that four types of bonds between cations and anions are possible: Ga-As, Ga-N, In-As, and In-N. Thus, the macroscopic compositions (x and y) do *not* solely determine the types of bonds. The existence of this short range ordering is undesired as it strongly and uncontrollably affects the physical properties of these quaternary alloys [31, 32].

Experimentally, tremendous efforts have been made so far in order to overcome these challenges. The growth temperature has been found to be the most critical parameter: Lowering the temperature ($420°C < T < 450°C$) is essential to increase N solubility and to prevent other challenges [33]. Low growth temperatures can be realized by employing molecular-beam epitaxy (MBE) as a powerful growth technique that also allows to carefully control the composition and the growth rate of heterostructures. Furthermore, it has been found that the solubility of N can be increased by using a plasma N source, which supplies high energy N molecules/ions/atoms. Combining these approaches, N concentrations up to a few percent have been realized [34, 35, 36]. On the other hand, the use of these growth procedures has been found to lead to the introduction of defects. In order to improve material quality, thermal annealing after growth (post-growth annealing) at high temperatures is commonly employed. It has been found that the post-growth annealing causes a reorganization of atoms, leading to the preferred formation of In-N pairs [37, 29, 38, 39, 40]. While the thermal annealing improves the optical properties, e.g. by improving the photoluminescence efficiency by a factor of up to 20 and reducing the spectral linewidth, it has been found to cause an undesired blueshift of the photoluminescence [41, 40]. Further experiments revealed that the blueshift increases with lowering the growth temperature [40, 42]. The blueshift in the photoluminescence spectra has been also extensively studied and it is attributed mainly to the atomic rearrangements and redistribution [43, 27, 39].

Finally, we note that the 1.3 μm range InGaAsN edge-emitting lasers [44, 45] and vertical-cavity surface-emitting lasers (VCSELs) [46, 47, 48] have been successfully demonstrated. However, when attempting to increase the emission wavelength a degradation of the optical efficiency and laser performance has been observed. The loss in optical activity has been related to the higher required indium and nitrogen contents, of 40 % and 4 %, respectively. Hence, the progress for extending the emission wavelength towards the 1.55 μm range has been slow.

1.2 Surface kinetics and thermodynamics: Enhancing N solubility

The experimental findings discussed above indicate that the solubility of and the structural properties of dilute nitrides are determined via a delicate interplay between thermodynamics and surface kinetics. It is therefore essential to understand this interplay in order to optimize and control the growth of these alloys and to improve their properties. In order to explain this, a detailed understanding of the kinetic mechanisms for incorporating N in GaAs is crucial.

Early theoretical considerations pointed out that the solubility of impurities can be enhanced *near* surfaces [49]. One reason for this is the better atomic relaxations and hence better strain relief around the impurities at a surface compared to bulk. A second and more important reason is the tendency of semiconductor surfaces to reconstruct, i.e, atoms and bonds at the surface rearrange to reduce the surface energy. The resulting surface reconstruction leads to an intrinsic compressive/tensile strain in the surface and in the subsurface layers. Therefore, an effective strain cancellation can be realized, lowering the energy of the impurity and hence increasing its solubility. For incorporating N atoms in GaAs the effect of the surface on the solubility is expected to be substantial due to the large size-mismatch. Consequently, by employing different kinetic regimes for N incorporation different solubility limits are realized.

In order to explain this, a basic qualitative description is provided in the schematic picture in Fig. 1.2. It shows the energy of a N atom as a function of the incorporation layer (vertical depth), in a typical reconstructed GaAs(001) surface. As can be derived from Fig. 1.2, (i) the energy of the N atom when incorporated at the surface layers (E_1, E_2) is lower (i.e. more favorable) than that when incorporated in the bulk (E_3), and moreover, (ii) the lowest energy for the N atom (i.e. most favorable) is in the subsurface layer (here in the second anion layer, E_2), that is, the solubility of N is highest in the subsurface layer. In order to discuss the role of kinetics we refer to the mobility of the atoms. The mobility of an atom in a certain potential energy surface can be determined through its thermal energy. Let us define the threshold thermal energy as the maximum energy barrier that the N atom can overcome in an average rate that is higher than the growth rate of a monolayer. We show in Fig. 1.2, for the three energy minima points (E_1, E_2, and E_3), three threshold energies $E_{\mathrm{th1}}, \ldots, E_{\mathrm{th3}}$ of the N atom represented by three horizontal lines. These threshold energies correspond to three growth temperatures $T_1, \ldots, T_3$, with $T_1 > T_2 > T_3$. Based on this, if the threshold energy is above the left or right energy barrier the atom is considered to be mobile in the corresponding direction, and vice versa. If such a vertical transport is enabled the corresponding surface layer can equilibrate.

Based on this, three possible incorporation scenarios can be identified from Fig. 1.2. The first one is when the growth takes place at the temperature T_1 corresponding to E_{th1}. In this case, all surface and bulk layers are accessible by the N atom, and hence a full thermodynamic equilibrium is realized. The resulting N concentration is therefore determined by its energy in the bulk, i.e., it will be very low. The second case is when the growth takes place at the intermediate temperature T_2 corresponding to E_{th2}. Here only the topmost surface and the subsurface layers (first and second anion layers) are accessible by the N atom, while an equilibration with the bulk

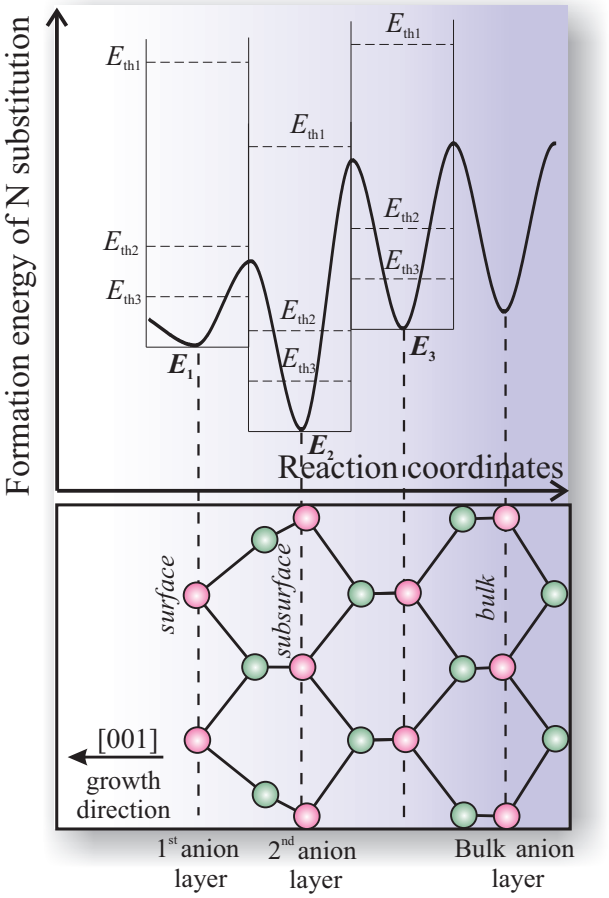


Figure 1.2: Qualitative description of how the *enhanced* solubility at semiconductor surfaces can be utilized to overcome the limited thermodynamic equilibrium bulk solubility. The figure schematically shows the energy of substitutional incorporation of a N atom as a function of the incorporation layer, in a typical (001) reconstructed surface. The horizontal lines are three selected thermal energies of the N atoms, which in turn determine their mobility. The maximal concentration of N can be achieved if the incorporation of N in the second anion layer is operational and can be frozen-in during growth (see text).

layers is prevented by the high energy barriers. More specifically, the N atoms preferentially get incorporated at the subsurface layer (E_2). During growth, an additional layer will be added to the surface, the subsurface layer becomes a bulk-like layer and the energy of the N atoms changes from E_2 to E_3. Since for this site the barrier from E_3 to E_2 is higher than the threshold energy E_{th2}, the N atoms will be prevented from going back to their ground state in E_2. The resulting N concentration is thus determined through the subsurface energy E_2 rather than by the bulk energy E_3, and will be frozen-in during growth, leading to an increased solubility compared to the equilibrium bulk solubility. The last case is realized at temperatures below T_3 corresponding to E_{th3}. Here only the topmost surface layer is accessible to the N atoms, and thus the equilibration is limited to it, therefore determining its solubility. The solubility in this case is lower than that in the second case. Therefore, the highest solubility is realized when the growth takes place at a thermal energy sufficient to overcome the barrier from E_1 to E_2 but too low to overcome the barrier from E_3 to E_2.

Based on this discussion, the concept of enhanced solubility at surfaces can be used in order to overcome the limited solubility of N in bulk GaAs. More specifically, the highest N solubility can be achieved if their incorporation in the subsurface region is operational and can be frozen-in during growth. Therefore, in order to employ this for the system it is required to: (i) identify the solubility limits in the bulk, surface, and subsurface layers, and (ii) identify the kinetic barriers that allow/limit the different mechanisms.

1.3 Objectives and the structure of this work

In order to overcome the growth challenges of dilute nitrides in achieving high N concentrations and high quality heterostructures for device applications an understanding of the mechanisms that control their growth and the interplay between them is required. Previously, only few theoretical studies were dedicated to calculate the solubility of N in GaAs bulk and at surfaces with the aim to understand the role of surface enhanced solubility in the experimentally achieved concentrations [50, 51]. However, these studies were rather simplified and approximate and failed to explain the experimental N concentrations in dilute nitrides under realistic conditions. Furthermore, surface and adatom kinetics, which have a very important role in the growth of these alloys, have not been studied or understood. As a result, the operational incorporation mechanism (surface, subsurface, or bulk) has been a controversial issue in determining N solubility. For instance, the solubility of N was first calculated assuming subsurface incorporation [50], and later it was calculated assuming bulk incorporation [51].

A promising approach to achieve a detailed understanding of the growth of dilute nitrides is to perform theoretical simulations based on first-principles (*ab-initio*) calculations. However, for these complex multi-component systems, large-scale *ab-initio* simulations for a realistic system size are currently out of reach. Nevertheless, in order to address this challenging problem, we develop in this work an approach that allows to understand the basic mechanisms controlling N solubility and the growth of these alloys on the atomic scale by combining surface and bulk thermodynamics with growth kinetics. The main aims for this work can be summarized as:

- First, to calculate the thermodynamic solubility of N in the bulk, surface, and subsurface of GaAs and InAs as a function of growth conditions, and taking into account the various surface reconstructions. This allows to identify the optimal growth conditions under which the highest N concentrations in the surface and subsurface layers can be achieved.

- Second, to identify N adatom kinetics at surfaces by determining the energy barriers for adatom diffusion and for the *s*urface and subsurface incorporation, and the kinetic barriers that prevent the equilibration with the bulk during growth. This allows to determine the functional incorporation mechanisms of N in practice and under typical MBE growth conditions, and to understand other growth related challenges of the ternary and quaternary systems.

- Finally, to identify the driving mechanisms for the puzzling compositional properties of quaternary dilute nitride alloys during and after annealing, specially the In-N anti-correlation, which allows to control them.

The thesis is organized as follows. In the second chapter we introduce key ideas and concepts of the *ab-initio* approach in material science based on the density functional theory (DFT). After that, in our search for a computationally less-expensive approach that can be applied for large-scale simulations, in Chapter 3 we introduce the density functional-based tight-binding (DFTB) approach and its implementation. Based on test calculations for the systems of interest

this method was not found to provide sufficient accuracy. Hence, the DFTB approach was not utilized further in this work, and a full DFT approach was employed to guarantee the required accuracy. In Chapter 4, we briefly discuss the employed computational details and provide the calculated bulk and cohesive properties for the material systems of interest.

In Chapter 5 we focus on the thermodynamics and solubility of N at surfaces. Being the initial step in the epitaxial growth of dilute nitrides, we first provide an overview of the surface reconstructions of the clean GaAs and InAs (001) surfaces, and calculate the surface phase diagrams. Next, based on extensive calculations for N substitutions, we analyze the effect of surface reconstruction on N solubility. Based on these results, we calculate the stability phase diagrams of N substitutions at GaAs and InAs (001) surface and subsurface layers and in the bulks, taking into account all relevant surface reconstructions. The results predict a significant enhancement of solubility at surface layers. By analyzing the phase diagrams and employing Monte Carlo simulations we develop a scheme that allows to calculate N concentration resulting from each of the incorporation mechanisms at typical growth temperatures as a function of the growth conditions. The solubility of N at surfaces turns out to be a complex function of growth conditions, in contrast to that in bulk. From these results we identify the optimal growth conditions that allow for maximal incorporation for the all considered mechanisms.

In Chapter 6 we focus on surface kinetics. First, we study the diffusion of N adatoms on the $\beta 2(2\times 4)$ reconstructed (001) surface. The results show that the diffusion behavior is challenging and leads to the failure of the conventional approach usually employed. A detailed analysis of this exceptionally complex diffusion behavior is provided. Based on this analysis we develop an approach that allows us to reveal the adatom diffusion energy landscape. After that and with the help of kinetic Monte Carlo simulations we develop a simplified and efficient approach that allows to determine the activation barriers for diffusion. In a next step we address the energy barriers for surface and subsurface substitutional incorporation of N. We propose possible reaction pathways for possible incorporation mechanisms and identify the associated energy barriers. According to these results we are able to identify the functional incorporation mechanism at relevant growth temperatures, and to calculate the incorporation time. We find that the subsurface incorporation of N cannot be realized under typical growth conditions, contrary to their incorporation at the topmost surface layer which is easily achievable.

In Chapter 7 we discuss the interplay between surface kinetics and thermodynamics in the growth of dilute nitride alloys. First, by analyzing the interaction between In and N atoms at the surface and in the bulk, combined with our results of incorporation kinetics, we are able for the first time to achieve a full understanding of the driving mechanisms for the compositional correlations and anticorrelation between In and N in the quaternary $In_yGa_{1-y}As_{1-x}N_x$ alloys. This insight allows for a better understanding of the growth of highly mismatched multicomponent alloys. Second, we discuss the role of kinetics in determining the solubility limits of N in GaAs and InAs, and identify the mechanism that can lead to enhancing N solubility. These results allow us to explain and understand the experimentally achieved N concentrations. Finally, concluding remarks of this work are provided.

Chapter 2

Ab-initio Methods

2.1 Introduction

A main objective in modern physics is the development of methodological and numerical approaches that allow to describe the physical properties of interacting many-atomic systems, such as bulks, surfaces, interfaces, defects, dislocations, etc. This, in principle, requires solving the $3N_e$ dimensional quantum mechanical Schrödinger equation of the N_e interacting electrons in the system. A realistic condensed matter system involves order of $\sim 10^{23}$ atoms,[1] which makes the exact analytical or numerical solutions for such a problem impossible.

The capability of density functional theory (DFT) to elegantly transform the complex many-body problem into a single-body one [52, 53] has revolutionized computational material science. Its success in describing many physical properties together with the continuous technological development of computational resources have sparked a continuously increasing interest in this field over the last decades, with broadening areas of applications [54]. This makes computational material physics simulations tempting to serve as computer "experiments". In addition to describing and predicting material properties, computational material physics simulations have additional benefits in revealing impracticable/inaccessible experimental regimes and offering an understanding of the basic physical mechanisms at the atomic level.

It is important to note that the class of methods that employ DFT are called *ab-initio* methods, a Latin phrase meaning "from the beginning" or often referred to as "first principles". In practice, it means that the DFT theory is derived from the basic physical quantum mechanical laws and there are no experimentally derived/adjustable parameters used in the theory and its approximations.

In this chapter we shall therefore briefly introduce the basic concepts of density functional theory and the involved approximations, and the practical methodology for solving the Schrödinger equation for extended systems.

[1] Avogadro's number $= 6.022141 \times 10^{23}$ mol^{-1}, which is the number of atoms in 12 g of ^{12}C.

2.2 The many-body problem

The full information about any material system is included in its time-dependent Schrödinger equation. If one is concerned with only time-independent properties, the time-independent Schrödinger equation can be used instead, which can be written as:

$$H(\{\mathbf{x}_i\},\{\mathbf{R}_I\})\Psi_\nu(\{\mathbf{x}_i\},\{\mathbf{R}_I\}) = E_\nu\Psi_\nu(\{\mathbf{x}_i\},\{\mathbf{R}_I\}). \quad (2.1)$$

This is an eigenvalue problem with H is the Hamiltonian operator and E_ν is the eigenvalue corresponding to the many-body wavefunction Ψ_ν (eigenfunction). The set $\{\mathbf{x}_i\}$ denotes the generalized coordinates of the N_e electrons (containing both spacial $\{\mathbf{r}_i\}$ and spin coordinates) and $\{R_I\}$ are the spacial coordinates of the N_nucl nuclei. In case of no external fields present and negligible relativistic effects, the Hamiltonian in Eq. (2.1) can be simplified to:

$$\begin{aligned} H(\{\mathbf{x}_i\},\{\mathbf{R}_I\}) &= T^\mathrm{e}(\{\mathbf{x}_i\}) + T^\mathrm{nucl}(\{\mathbf{R}_I\}) + V^\mathrm{e-e}(\{\mathbf{x}_i\}) \\ &\quad + V^\mathrm{nucl-nucl}(\{\mathbf{R}_I\}) + V^\mathrm{e-nucl}(\{\mathbf{x}_i\},\{\mathbf{R}_I\}). \end{aligned} \quad (2.2)$$

This Hamiltonian contains the kinetic energy operator for electrons T^e and nuclei T^nuc, and the Coulomb electrostatic interaction operators between electrons ($V^\mathrm{e-e}$), nuclei ($V^\mathrm{nuc-nuc}$), and electrons-nuclei ($V^\mathrm{e-nuc}$). These terms are[2]

$$T^\mathrm{e}(\{\mathbf{x}_i\}) = -\frac{1}{2}\sum_{i=1}^{N_\mathrm{e}}\nabla_i^2\,, \quad (2.3)$$

$$T^\mathrm{nuc}(\{\mathbf{R_I}\}) = -\frac{1}{2}\sum_{I=1}^{N_\mathrm{nuc}}\frac{\nabla_I^2}{m_I}\,, \quad (2.4)$$

$$V^\mathrm{e-e}(\{\mathbf{x}_i\}) = \sum_{i=1}^{N_\mathrm{e}}\sum_{j>i}^{N_\mathrm{e}}\frac{1}{|\mathbf{r}_i-\mathbf{r}_j|}\,, \quad (2.5)$$

$$V^\mathrm{nuc-nuc}(\{\mathbf{R}_I\}) = \sum_{I=1}^{N_\mathrm{nuc}}\sum_{J>I}^{N_\mathrm{nuc}}\frac{Z_I Z_J}{|\mathbf{R}_I-\mathbf{R}_J|}\,, \quad (2.6)$$

and finally

$$V^\mathrm{e-nuc}(\{\mathbf{x}_i\},\{\mathbf{R}_I\}) = -\sum_{i=1}^{N_\mathrm{e}}\sum_{I=1}^{N_\mathrm{nuc}}\frac{Z_I}{|\mathbf{r}_i-\mathbf{R}_I|}\,. \quad (2.7)$$

The symbols m_α and Z_α denote the mass and the charge (atomic number) of nucleus α, respectively.

Because of the number of degrees of freedom in a realistic material system, a solution for the above Schrödinger equation is impossible without further simplifications and approximations. Within DFT there are a number of well-justified approximations that we shall present in the following.

[2]Hereafter, atomic units are used ($m_e = 1$, $\hbar = 1$, $e = 1$, and $4\pi\epsilon_0 = 1$).

2.3 The Born-Oppenheimer approximation

The approximation proposed by Born and Oppenheimer [55] in 1927 is the first involved approximation in the DFT formalism. It is based on the fact that nuclei are much heavier than electrons, with a mass ratio ranging from ≈ 2000 for m_p/m_e (proton to electron mass) and up to $\sim 10^5$ for heavier elements. Nuclei hence move orders of magnitude slower than electrons and therefore the electrons-nuclei motions can be decoupled. In other words, electrons are assumed to follow "instantaneously" the motion of the nuclei. This approximation provides a major simplification to the many body Schrödinger equation.

To formulate the Born-Oppenheimer approximation, let us define H^{s} as the Hamiltonian of the *static* system (for a certain fixed nuclei configuration) as

$$H^{\mathrm{s}}(\{\mathbf{x}_i\};\{\mathbf{R}_I\}) \quad = \quad T^{\mathrm{e}} + V^{\mathrm{e-e}} + V^{\mathrm{e-nucl}} + V^{\mathrm{nucl-nucl}}. \tag{2.8}$$

Note that this Hamiltonian depends parametrically on the nuclear coordinates $\{\mathbf{R}_I\}$ (as the semicolon indicates), i.e., it corresponds to a set of fixed nuclear positions. The Schrödinger equation for this system then reads

$$H^{\mathrm{s}}(\{\mathbf{x}_i\};\{\mathbf{R}_I\})\Phi_\nu(\{\mathbf{x}_i\};\{\mathbf{R}_I\}) \quad = \quad E^{\mathrm{s}}_\nu(\{\mathbf{R}_I\})\Phi_\nu(\{\mathbf{x}_i\};\{\mathbf{R}_I\}), \tag{2.9}$$

with E^{s}_ν and Φ_ν are the corresponding electronic eigenvalues and eigenfunctions, respectively. Since the Φ_ν form a complete basis set, the many body wavefunction Ψ_η can be expanded as

$$\Psi_\eta(\{\mathbf{x}_i\};\{\mathbf{R}_I\}) \quad = \quad \sum_\nu \Lambda_{\nu\eta}\Phi_\nu. \tag{2.10}$$

Substituting the above equation in Eq. (2.1), and after manipulation, multiplying from left with Φ_μ, integrating, and making use of the orthonormality relation $\langle \Phi_\mu \mid \Phi_\nu \rangle = \delta_{\mu\nu}$ (for more details See Ref. [56]), we get

$$\left[T^{\mathrm{nuc}} + E^{\mathrm{s}}_\mu - \left\{\sum_\nu \sum_{I=1}^{N_{\mathrm{nuc}}} \frac{1}{2m_I}(\langle\Phi_\mu|\nabla_I^2 \mid \Phi_\nu\rangle + 2\langle\Phi_\mu|\nabla_I|\Phi_\nu\rangle\nabla_I)\right\}\right] \quad \Lambda_{\mu\eta} = \quad E_\eta\Lambda_{\mu\eta}. \tag{2.11}$$

The last term in the Hamiltonian (between the curly brackets) is due to electron-phonon coupling. Due to the large nuclear masses in the denominator these matrix elements are small. Therefore, for the cases in which the inclusion of electron-phonon coupling is not essential (e.g. relatively slow movements of the nuclei) these matrix elements can be ignored. This applies to our systems. In that case, Eq. (2.11) simplifies to

$$\left[T^{\mathrm{nuc}} + E^{\mathrm{s}}_\mu\right] \quad \Lambda_{\mu\eta} = \quad E_\eta\Lambda_{\mu\eta}. \tag{2.12}$$

The above equation shows that the nuclei are experiencing the potential energy surface of the static system. This energy surface is usually referred to as the Born-Oppenheimer surface (BOS). Based on the above derivation, the –adiabatic– motion of nuclei can be derived from the solution

of the Schrödinger equations of the static systems.

For atomic motion, quantum effects are often neglected and hence the classical equations of motions are solved

$$m_I \frac{\partial^2}{\partial t^2}\mathbf{R}_I = -\frac{\partial}{\partial \mathbf{R}_I} E^{\mathrm{s}}(\{\mathbf{R}_I\}). \tag{2.13}$$

The forces acting on the atoms can be calculated using the Hellmann-Feynmann theorem[3] [57]:

$$\frac{\partial}{\partial \mathbf{R}_I} E^{\mathrm{s}}(\{\mathbf{R}_I\}) = \langle \Phi_\nu(\{\mathbf{x}_i\};\{\mathbf{R}_I\}) | \frac{\partial}{\partial \mathbf{R}_I} H^{\mathrm{s}}(\{\mathbf{x}_i\};\{\mathbf{R}_I\}) | \Phi_\nu(\{\mathbf{x}_i\};\{\mathbf{R}_I\}) \rangle, \tag{2.14}$$

which can be used e.g. for structure optimization (relaxation).

2.4 The basics of Density-Functional Theory (DFT)

The basic concept of density functional theory is to promote the *electron density* $n(\mathbf{r})$ from just an observable to the *basic variable*, instead of the many-body wave-function. Since the charge density is a function of only three spacial coordinates,[4] this approach remarkably simplifies the problem and reliefs it from the $3N$ dimensionality of the many-body wave-function. Although the first attempt to use this concept was in 1927 by Thomas and Fermi [58, 59], the core formulation of DFT was in 1964 by Hohenberg and Kohn [52]. After that, in 1965, Kohn and Sham [53] provided a practical approach to apply the DFT.

2.4.1 The Hohenberg-Kohn theorems

Let us consider N interacting electrons moving in an external potential V^{ext}, which can be, e.g., the electrons-nuclei interaction potential. The Hamiltonian for this electronic system can be written as (see Eq. (2.2)):

$$H = T^{\mathrm{e}} + V^{\mathrm{ee}} + V^{\mathrm{ext}}. \tag{2.15}$$

The first theorem of Hohenberg and Kohn states that the external potential V^{ext} is a unique functional, within an additive constant, of the ground-state electron density $n_0(\mathbf{r})$, i.e., $V^{\mathrm{ext}} = V^{\mathrm{ext}}[n_0(\mathbf{r})]$. Since the wavefunction is a functional of the external potential ($\Psi[V^{\mathrm{ext}}]$) then it is also a functional of the ground-state electron density. As a consequence, the expectation value of any observable is also a functional of the ground state density

$$O = O[n_0] = \langle \Psi[n_0] | O | \Psi[n_0] \rangle. \tag{2.16}$$

Now to arrive to the second Hohenberg-Kohn theorem let us first mention the variational principle, which states that the expectation value of the Hamiltonian is always grater than or

[3] This applies when expanding the wavefunction using a basis set that does not depend on atomic positions (see Sec. 2.7.2 on page 25), otherwise additional contributions should be included, called Pulay forces.

[4] Since throughout this work we will be dealing with spin compensated systems, the spin degrees of freedom will be suppressed hereafter.

equal to the ground state energy E_0:

$$E[\Psi] \quad = \quad \frac{\langle\Psi|H|\Psi\rangle}{\langle\Psi|\Psi\rangle} \geq E_0. \tag{2.17}$$

Combining the variational principle and the first theorem, the second Hohenberg-Kohn theorem can be stated as: The total energy of the system assumes its minimum value (that is, the ground state energy) at the true ground state electron density:

$$E_0 = E[n_0] \quad \leq \quad E[n]. \tag{2.18}$$

Hohenberg and Kohn have also defined a *universal* functional $F[n]$ such that

$$F[n] \quad = \quad T^{\mathrm{e}}[n] + V^{\mathrm{ee}}[n], \tag{2.19}$$

where $F[n]$ is universal in the sense that it does not depend on the external potential V^{ext}. The total energy functional hence can be written as

$$E[n] \quad = \quad F[n] + \int n(\mathbf{r}) V^{\mathrm{ext}}(\mathbf{r})\, d\mathbf{r}. \tag{2.20}$$

Based on the above theorems, the ground-state energy can be –exactly– found by minimizing Eq. (2.20) with respect to electron density, under the constraint that the total number of electrons in the system is conserved $\int n(\mathbf{r}) d\mathbf{r} = N$. This can be expressed as

$$\begin{aligned} \frac{\delta}{\delta n(\mathbf{r})} \left\{ E[n] - \mu \left[\int n(\mathbf{r})\, d\mathbf{r} - N \right] \right\} \quad &= \\ \frac{\delta F[n]}{\delta n(\mathbf{r})} + V^{\mathrm{ext}}(\mathbf{r}) - \mu \quad &= \quad 0, \end{aligned} \tag{2.21}$$

where the μ is the Lagrange multiplier associated with the requirement of constant number of electrons. Unfortunately, the explicit form of the universal functional $F[n]$ is *unknown*, and hence a practical approach to treat $F[n]$ is required. This approach was first introduced by Kohn and Sham in 1965, as will be introduced directly in the following section.

2.4.2 The Kohn-Sham formalism

Kohn and Sham [53] have proposed a practical approach to treat the universal functional $F[n]$ by mapping the many-body interacting system into a fictitious auxiliary *non-interacting* one. They introduced the following separation of the universal functional

$$F[n] \quad = \quad T^s[n] + E^H[n] + E^{\mathrm{XC}}[n], \tag{2.22}$$

where $T^s[n]$ is the kinetic energy functional of the *non-interacting* electrons, $E^H[n]$ is the classical electron-electron interaction term (Hartree term), and $E^{\mathrm{XC}}[n]$ is known as the exchange-correlation term and it contains all the remaining unknown contributions, i.e., the difference in

contributions between the real interacting system and the fictitious non-interacting system,

$$E^{\mathrm{XC}}[n] \equiv (T^{\mathrm{e}}[n] + V^{\mathrm{e-e}}[n]) - (T^{s}[n] + V^{H}[n]). \tag{2.23}$$

Provided that the single particle wavefunctions (of the non-interacting system) ψ_i are orthonormal, the terms $T^s[n]$ and $E^H[n]$ have the form

$$T^{s}[n] = -\frac{1}{2}\sum_{i=1}^{occ}\int \psi_i^*(\mathbf{r})\nabla^2\psi_i(\mathbf{r})\,d\mathbf{r} \tag{2.24}$$

and

$$E^{H}[n] = \frac{1}{2}\int\int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,d\mathbf{r}\,d\mathbf{r}', \tag{2.25}$$

and they can be calculated *exactly*. This is important since these terms contain the major contributions to the universal functional $F[n]$. The remaining unknowns are now merged to the exchange-correlation term E^{XC} solely. If its form were known, the Kohn-Sham formalism would provide an exact solution. Unfortunately its form is unknown and therefore, at this stage, approximations are required for this part. While it is still a challenging problem to find a good approximation, few fairly reliable approximations exist. In the next section we shall present the main approaches to approximate the exchange and correlation contribution.

To see how the Kohn-Sham formalism works in practice let us come back to the proposed separation in Eq. (2.22). Hence, the total energy functional can be written as

$$E[n] = T^{s}[n] + E^{H}[n] + E^{\mathrm{XC}}[n] + \int n(\mathbf{r})V^{\mathrm{ext}}(\mathbf{r})\,d\mathbf{r}. \tag{2.26}$$

Applying the variational principle (as in Eq. (2.21)) leads to

$$\begin{aligned}\frac{\delta T^s[n]}{\delta n(\mathbf{r})} + \frac{\delta E^H[n]}{\delta n(\mathbf{r})} + \frac{\delta E^{\mathrm{XC}}[n]}{\delta n(\mathbf{r})} + V^{\mathrm{ext}}(\mathbf{r}) - \mu &= \\ \frac{\delta T^s[n]}{\delta n(\mathbf{r})} + V^H[n] + V^{\mathrm{XC}}[n] + V^{\mathrm{ext}}(\mathbf{r}) - \mu &= 0,\end{aligned} \tag{2.27}$$

where the terms $V^H[n]$, $V^{\mathrm{XC}}[n]$, and $V^{\mathrm{ext}}(\mathbf{r})$ are defined as the Hartree, exchange-correlation, and external potentials, respectively. Consider now a non-interacting system of electrons moving in an external (effective) potential $V^{\mathrm{eff}}(\mathbf{r})$, the minimization condition is

$$\frac{\delta T^s}{\delta n(\mathbf{r})} + V^{\mathrm{eff}}[n] - \mu = 0. \tag{2.28}$$

It is clear now that Eq. (2.27) is mathematically identical to Eq. (2.28) if one defines

$$V^{\mathrm{eff}}[n] = V^{\mathrm{ext}}(\mathbf{r}) + V^{H}[n] + V^{\mathrm{XC}}[n]. \tag{2.29}$$

That is, one can calculate the density of the interacting many-body electrons moving in the

external potential $V^{\text{ext}}(\mathbf{r})$ by solving the Schrödinger equations of the non-interacting single-body electrons moving in the potential $V^{\text{eff}}[n]$ defined by Eq. (2.29). The Schrödinger equation for this auxiliary system is

$$\left[-\frac{1}{2}\nabla^2 + V^{\text{eff}}[n]\right]\psi_i(\mathbf{r}) \quad = \quad \varepsilon_i\psi_i(\mathbf{r}), \tag{2.30}$$

which yields the single particle wavefunctions ϕ_i with an electron density

$$n(\mathbf{r}) \quad = \quad \sum_i^{\text{occ}} |\psi_i(\mathbf{r})|^2. \tag{2.31}$$

Equations (2.29)-(2.31) are the so-called Kohn-Sham equations. Note that the eigenvalues of the auxiliary single-body system $\{\varepsilon_i\}$ (or the Kohn-Sham eigenfunctions) enter just as Lagrange multipliers and they have no formal interpretation except for the highest occupied state. It is only the ground state charge density and energy that have a strict physical meaning. Nevertheless, they show a good performance in describing the actual energy levels of extended systems, apart from underestimating the band gaps.

The Kohn-Sham equations are solved *self-consistently*. This is done by starting with an initial guess for the charge density (e.g. using atomic orbitals), calculating the corresponding effective potential, and then solving the differential equation Eq. (2.30) and calculating the charge density. This is then repeated until convergence is achieved, i.e, the input charge density is equivalent to the output one.

2.4.3 Exchange-correlation functionals

Since there exists no exact form of the exchange-correlation potential, it needs to be approximated. The most common (and the simplest) approximation is the local-density approximation (LDA). In LDA, the exchange-correlation energy E^{XC} of the real *non-homogeneous* density $n(\mathbf{r})$ is computed as if it *locally* equals that of a *homogeneous* electron gas of the *same local* density. This can be written as:

$$E^{\text{XC}}_{\text{LDA}}[n] \quad = \quad \int n(\mathbf{r})\epsilon^{\text{xc}}_{\text{hom}}(n(\mathbf{r}))\, d\mathbf{r}, \tag{2.32}$$

where $\epsilon^{\text{xc}}_{\text{hom}}(n(\mathbf{r}))$ is the exchange-correlation energy (per electron) of a homogeneous electron gas with electron density n. In this form the exchange-correlation energy can be decomposed into an exchange, ϵ^{x}, and correlation, ϵ^{c}, part:

$$\epsilon^{\text{xc}}_{\text{hom}}(n) \quad = \quad \epsilon^{\text{x}}_{\text{hom}}(n) + \epsilon^{\text{c}}_{\text{hom}}(n). \tag{2.33}$$

An exact analytical form of the exchange energy of a homogeneous electron gas can be derived using the Hartree-Fock approximation and it takes the following form:

$$\epsilon^{\text{x}}_{\text{hom}}(n) \quad = \quad -\frac{3}{4}\sqrt[3]{\frac{3}{\pi}n(\mathbf{r})}. \tag{2.34}$$

The correlation energy part $\epsilon^{\mathrm{c}}_{\mathrm{hom}}(n)$ was calculated using Quantum Monte Carlo calculations for a homogeneous electron gas by Ceperley and Alder [60], and was later parametrized by Perdew and Zunger [61].

LDA is exact only for a homogeneous system. Nevertheless, it has been surprisingly successful in providing good results for a wide range of material science problems [62]. This is probably due to good error cancellation between exchange and correlation parts. However, LDA becomes less accurate for systems of increased inhomogeneity, such as atoms and molecules. In addition, LDA tends to show an *over-binding*, i.e., overestimating binding and cohesive energies which leads to underestimated equilibrium bond lengths and lattice constants.

In order to reduce the deficiencies of LDA, the generalized gradient approximation (GGA) has been proposed. In GGA, additional information about the gradient of the electron density are included. This functional is expressed as:

$$E^{\mathrm{XC}}_{\mathrm{GGA}}[n] \quad = \quad \int n(\mathbf{r})\epsilon^{\mathrm{xc}}(n(\mathbf{r}), |\nabla n(\mathbf{r})|)\, d\mathbf{r}. \tag{2.35}$$

Several schemes have been proposed for the parametrization of the GGA functionals, such as those by Becke [63] and Perdew [64], Lee, Yang, and Parr [65], Perdew and Wang [66] (PW91), Perdew, Burke, and Ernzerhof [67] (PBE), and the revised PBE (RPBE) by Hammer et al. [68]. GGA is found to significantly improve binding and cohesive energies of many solids, and activation barrier energies of chemical reactions. On the other hand, GGA tends to overestimate bond lengths and lattice constants with a subsequent reduction of bulk moduli, in addition to the underestimation of band gaps (the so called the band-gap problem) which is a common problem in both GGA and LDA.

The problem in exchange-correlation is to find more reliable representations of the non-locality of this functional. For instance, trends to do that include (i) including higher-order powers of the gradient of the density, known as meta-GGA functionals, (ii) including an exact-exchange (EXX) in addition to a standard functional, or (iii) the use of Green functions method (the GW approximation). These functionals however result in a significant increase in the computational efforts. The development of more accurate exchange-correlation approximations is still a very active research field and it is beyond the aims of this work. Therefore, in this work we restrict ourselves to the accuracy provided by GGA functional of Perdew, Burke, and Ernzerhof [67] (PBE) as one of the most famous existing parametrizations of the GGA exchange-correlation functionals.

2.5 The pseudopotential approximation

In crystals and molecules, chemical binding happens mainly through valence electrons. Core electrons, on the other hand, are inert or not significantly affected by the chemical environment. Hence, as the wavefunctions of the core electrons of neighboring atoms do not overlap, they can be treated as "frozen". The main idea of the pseudopotential approach is therefore to eliminate the core states and treat only the valence states by replacing the actual all-electron potential by an effective potential that acts on the valence electrons, the so-called pseudopotential.

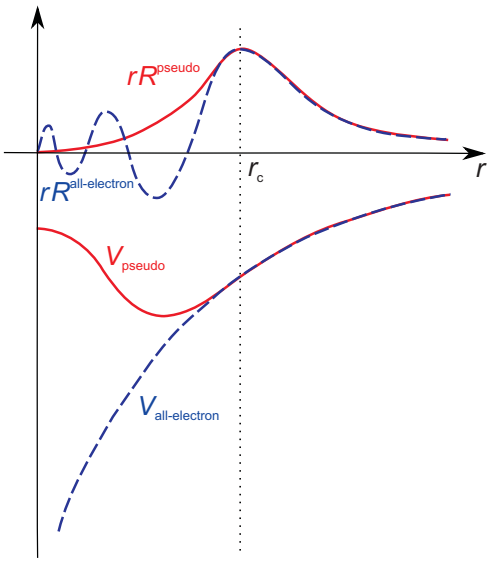


Figure 2.1: Schematic illustration of the pseudopotential approach. The figure shows the all-electron (blue dashed lines) and the pseudo (red solid lines) potentials and the corresponding radial parts of their wavefunctions (see text).

One might therefore replace the Coulomb potential of the nucleus with an effective ionic potential which affects valence electrons and which combines the Hartree potential of the core electrons and the Coulomb potential of the nucleus. However, while the wavefunctions of valence electrons are smooth outside the core region, they are rapidly oscillating inside the core region because of their orthogonality to the wavefunctions of the core electrons. The rapidly oscillating wavefunctions are numerically challenging because they require a huge basis set when using plane-waves basis (see Sec. 2.7.1) to describe them accurately. Therefore, a more efficient approach is to replace this ionic potential by a pseudopotential that leads to a *nodless* valence wavefunction that matches the true all-electron one beyond a certain core radius (r_c). The basic idea of the pseudopotential approach is schematically illustrated in Fig. 2.1.

The main advantage of using pseudopotentials is therefore significantly reducing the size of the basis set used to expand the wavefunctions by (i) reducing the number of treated electrons and by (ii) avoiding to represent the rapidly oscillating valence wavefunctions in the core region by producing nodeless valence wavefunctions while keeping their long tails which are mainly responsible for the chemical bonding. Consequently, this results in a significant decrease in the computational efforts without losing important physical or chemical information, when carefully applied.

To show the validity of the pseudopotential approach let us expand the one-particle valence electronic wavefunction ψ_v as[5]

$$|\psi_v\rangle \quad = \quad |\psi_{\rm ps}\rangle + \sum_i b_{c_i} |\psi_{c_i}\rangle, \tag{2.36}$$

where $\psi_{\rm ps}$ is a smooth wavefunction, ψ_{c_i} are core wavefunctions, and b_{c_i} are the expansion coefficients. These coefficients are determined from the condition that the valence and core

[5]In fact, the pseudopotential method originates from the orthogonalized plane-wave method (OPW).

states are orthogonal, i.e., $\langle\psi_v|\psi_{c_i}\rangle = 0$, which leads to

$$b_{c_i} = -\langle\psi_{c_i}|\psi_{\rm ps}\rangle. \tag{2.37}$$

Note that $\psi_{\rm ps}$ is not uniquely defined by Eq. (2.36) as the coefficients can be freely chosen. Since ψ_v and ψ_{c_i} are solutions of the Schrödinger equation with eigenvalues ε_v and ε_{c_i}, respectively, using Kohn-Sham equation Eq. (2.30) one can write

$$\left[-\frac{1}{2}\nabla^2 + V^{\rm eff}\right]|\psi_v\rangle = \varepsilon_v|\psi_v\rangle. \tag{2.38}$$

We now substitute the expanded valence wavefunction given by Eqs. (2.36) and (2.37) into Eq. (2.38). After some manipulation we get

$$\left[-\frac{1}{2}\nabla^2 + V^{\rm ps}\right]|\psi_{\rm ps}\rangle = \varepsilon_v|\psi_{\rm ps}\rangle, \tag{2.39}$$

where $V^{\rm ps}$ is the pseudopotential and it is defined as:

$$V^{\rm ps} = V^{\rm eff} + \sum_i (\varepsilon_v - \varepsilon_{c_i})|\psi_{c_i}\rangle\langle\psi_{c_i}|. \tag{2.40}$$

As $\varepsilon_v > \varepsilon_{c_i}$ and the core states are localized, the summation term in Eq (2.40) represents a short-range repulsive potential. Note that the pseudo-wavefunction has the same eigenvalue as the true valence wavefunction and it is not orthogonal to the core states, which leads to more efficient expansion of the wavefunction in terms of basis functions.

2.5.1 Norm-conserving pseudopotentials

Many methods and recipes have been proposed over the last decades to generate *ab-initio* pseudopotentials for practical use [69, 70, 71, 72, 73, 74, 75, 76]. The main objectives to achieve in theses different schemes are *transferability* and *efficiency*. Transferability of the pseudopotentials indicate their ability to accurately describe the valence electrons in different chemical environments (atomic, molecular, and solid states), while efficiency refers to minimizing the computational efforts as possible. In this work we use norm-conserving pseudopotentials [72, 69], first introduced by Hamann et al. [69]. These pseudopotentials ensure a good transferability property by the constraint that the total number of electrons inside the core region are conserved.

Norm-conserving pseudopotentials are constructed starting from *ab-initio* self-consistent all-electron calculations for the specific atom in a reference electronic configuration, typically the ground state configuration. This is done by solving the radial Schödinger equation for the valence wavefunctions $R_{nl}(r)$ and their corresponding eigenvalues ε_{nl}:

$$\left[-\frac{1}{2}\frac{d^2}{dr^2} + \frac{l(l+1)}{2r^2} + V^{\rm eff}[n;r]\right] rR_{nl}(r) = \varepsilon_{nl}\, rR_{nl}(r), \tag{2.41}$$

where

$$V^{\text{eff}}[n;r] \;=\; -\frac{Z}{r} + V^{H}[n] + V^{\text{XC}}[n]. \tag{2.42}$$

The pseudopotential is then constructed such that:

1. It yields the same valence eigenvalues as the real atom (hereafter we suppress the principal quantum number n):

$$\varepsilon_l^{\text{ps}} \;=\; \varepsilon_l. \tag{2.43}$$

2. The radial wavefunctions are nodeless and exactly match the all-electron radial wavefunctions beyond a chosen core radius for each angular momentum (r_{cl}), that is:

$$R_l^{\text{ps}}(r) \;=\; R_l(r) \qquad \text{for } r > r_{cl}. \tag{2.44}$$

3. Inside the core region ($r < r_{cl}$), a certain parametrized form for the radial wavefunction is assumed. The parameters are obtained by imposing certain boundary conditions, including e.g. norm-conservation. In this work we use the form of Troullier and Martins [75], which will be briefly described below.

4. The so-called screened pseudopotential ($V_l^{\text{ps,scr}}$) is obtained by inverting the radial Schrödinger equation for the pseudo atom

$$\left[-\frac{1}{2}\frac{d^2}{dr^2} + \frac{l(l+1)}{2r^2} + V^{\text{ps,scr}}[n;r]\right] rR_l^{\text{ps}}(r) \;=\; \varepsilon_l^{\text{ps}} rR_l^{\text{ps}}(r). \tag{2.45}$$

   which gives

$$V_l^{\text{ps,scr}}(r) \;=\; \varepsilon_l^{\text{ps}} - \frac{l(l+1)}{2r^2} + \frac{1}{2rR_l^{\text{ps}}(r)}\frac{d^2}{dr^2}[rR_l^{\text{ps}}(r)]. \tag{2.46}$$

5. Finally, the ionic pseudopotential ($V_l^{\text{ps,ion}} \equiv V_l^{\text{ps}}$) is obtained by unscreening, i.e., subtracting the Hartree and the exchange-correlation potentials due to the valence electrons n^v

$$V_l^{\text{ps}}(r) \;=\; V_l^{\text{ps,scr}}(r) - V^{H}[n^v;r] - V^{\text{XC}}[n^v;r]. \tag{2.47}$$

Note that Eq. (2.47) assumes that the exchange-correlation energy is *linear* in the electron density. While the actual exchange-correlation is nonlinear, it is still common (and adequate) to use this form for many cases. However, when the overlap between core and valence electrons is considerable the upper linearization becomes inadequate, which introduces errors. In such cases, a *nonlinear core correction* (nlcc) [77] is needed. This means that the core density should be also considered in the calculation of exchange-correlation energy to perform the unscreening. However, it is usually sufficient to include a partial core density outside a certain radius r^{nlc}.

In our case, it is better to consider the d-electrons of the group-III elements (Ga and In) as valence electrons [78]. However, to keep our calculations less expensive we treat them as core electrons and apply the non-linear core correction.

Troullier-Martins scheme

Troullier and Martins [75] scheme for generating norm-conserving pseudopotentials is a modification to the earlier scheme of Kerker [71] to improve the computational efficiency. They proposed the following form for the radial pseudo-wavefunction inside the core region:

$$R_l^{\rm ps}(r) \;=\; r^l e^{p(r)} \qquad \text{for } r < r_{cl}, \tag{2.48}$$

with $p(r)$ is a polynomial of order six in r^2

$$p(r) \;=\; c_0 + c_2 r^2 + c_4 r^4 + c_6 r^6 + c_8 r^8 + c_{10} r^{10} + c_{12} r^{12}. \tag{2.49}$$

The 7 coefficients in the upper polynomial are found by imposing the following boundary conditions:

- The continuity of the pseudo-wavefunction and its first four derivatives at the core radius r_{cl}.
- Zero curvature of the screened pseudopotential at the origin, $\frac{d^2}{dr^2} V_l^{\rm ps,scr}(r)|_{r=0} = 0$.
- Norm-conservation condition. Namely, the total charge inside the core radius of the pseudo-atom is identical to that of the real atom:

$$\int_0^{r_{cl}} |R_l^{\rm ps}(r)|^2 r^2 dr \;=\; \int_0^{r_{cl}} |R_l(r)|^2 r^2 dr. \tag{2.50}$$

  The norm conservation condition ensures to maintain the scattering properties of the real atom. We note that the logarithmic derivative of the radial wavefunction is related to the norm-conservation condition, and hence it is usually used as a measure to the transferability of the pseudopotentials.

Separable representation: Kleinman-Bylander

The ionic pseudopotential is not only r-dependent but also angular momentum dependent, i.e., it is a non-local operator. It can be hence written as

$$V^{\rm ps}(r) \;=\; \sum_{l=0}^{l_{\rm max}} V_l^{\rm ps}(r) \hat{P}_l, \tag{2.51}$$

where $\hat{P}_l$ is a projection operator, $l_{\rm max}$ is the maximum angular momentum component of the specific ionic pseudopotential. However, in order to reduce the computational efforts, it is useful

to decompose it into a purely local part and a semi-local part:

$$V^{\rm ps}(r) \;\; = \;\; V_{\rm loc}(r) + \sum_{l=0}^{l_{\rm max}} \delta V_l^{\rm ps}(r) \hat{P}_l, \tag{2.52}$$

where $\delta V_l^{\rm ps}(r) = V_l^{\rm ps}(r) - V_{\rm loc}(r)$ is the semi-local part, and $V_{\rm loc}(r)$ is the local part and it is typically chosen to be one of the l-components of the pseudopotential, i.e., $V_{\rm loc}(r) = V_{l=l_{\rm loc}}^{\rm ps}$. Kleinman and Bylander [79] have further reduced the computational efforts by using a fully separable form of the semi-local part. Further details of the pseudopotential generation and optimization can be found elsewhere [80].

Hard, soft, and ultrasoft

We note that the choice of the core radius, r_{cl}, is basically bound by two extreme limits. The lower limit is the position of the outermost node in the all-electron valence wavefunction, and the upper limit is the half of the atomic nearest neighbor distance in the actual chemical environment. While reducing r_{cl} improves the transferability, it also leads to "hard" pseudopotentials (i.e. larger plane-waves basis set will be required), and vise versa. Therefore, r_{cl} is usually optimized with respect to these bounds. However, when treating first-row transition metals or atoms with d or f electrons, the resulting pseudopotentials are usually very hard. A similar case results when treating semi-core states as valence which is sometimes necessary. To avoid this, the so-called *ultrasoft* pseudopotentials have been introduced by Vanderbilt [73]. In these pseudopotentials the norm-conservation condition is relaxed and augmentation charges are introduced in the core region to recover the correct charge density. This results in a remarkable reduction of the plane-wave basis set, with the price of more complicated formality than when using the normal pseudopotentials.

2.6 Periodic boundary conditions and Brillouin zone integrations

To apply the Kohn-Sham formalism for extended systems it is in general required to treat huge number of electrons. Nevertheless, due to the translation symmetry in periodic crystalline solids (i.e., the periodic boundary conditions) the treated system size is drastically reduced since it becomes sufficient to solve the Kohn-Sham equations in the unit cell.

Consider a periodic crystal with $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$ being the primitive translation vectors of the unit cell. The effective potential has the same periodicity property of the lattice, i.e.,

$$V^{\rm eff}(\mathbf{r}+\mathbf{R}) \;\; = \;\; V^{\rm eff}(\mathbf{r}), \tag{2.53}$$

where $\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3$ is a lattice translation vector, the $\{\mathbf{a}_i\}$ vectors are the unit vectors of the primitive lattice, and $\{n_i\}$ are integers. As a results, Bloch's theorem can be applied. It states that solutions of a Schrödinger equation with a periodic potential can have the

form [81, 82]:

$$\psi_{n\mathbf{k}}(\mathbf{r}) \quad = \quad u_{n\mathbf{k}}(\mathbf{r})\, e^{i\mathbf{k}\cdot\mathbf{r}}, \tag{2.54}$$

where $u_{n\mathbf{k}}(\mathbf{r})$ is a periodic function with the same periodicity of the effective potential, i.e.,

$$u_{n\mathbf{k}}(\mathbf{r}+\mathbf{R}) \quad = \quad u_{n\mathbf{k}}(\mathbf{r}), \tag{2.55}$$

n denotes a band index, and $\mathbf{k}$ is a reciprocal space vector which lies inside the first Brillouin-Zone (BZ) [81, 82] in the reciprocal lattice $\mathbf{G}$:

$$\mathbf{G} \quad = \quad m_1\mathbf{b}_1 + m_2\mathbf{b}_2 + m_3\mathbf{b}_3, \tag{2.56}$$

with $\{\mathbf{b}_i\}$ being the primitive reciprocal lattice vectors which are related to their real space counterparts through $\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij}$. An important consequence of applying Bloch's theorem is the transformation of the problem from calculating infinite number of electronic states of the infinite system to calculating a finite number of states for an infinite number of $\mathbf{k}$-points. For instance, the charge density (Eq. 2.31) is written as

$$n(r) \quad = \quad \frac{1}{\Omega_{\mathrm{BZ}}} \sum_{i}^{\mathrm{occ}} \int_{\mathrm{BZ}} |\psi_{i\mathbf{k}}(\mathbf{r})|^2 \, d\mathbf{k}. \tag{2.57}$$

Where Ω_{BZ} is the volume of the first Brillouin zone (BZ). Because the wavefunctions are very similar at nearby $\mathbf{k}$-points, this representation becomes further much more useful. Hence, the summation over infinite $\mathbf{k}$-points (integration over the first BZ) can be replaced with a summation over a *finite* set of $\mathbf{k}$-points, say $N_{\mathbf{k}}$, called *special $\boldsymbol{k}$-points*:

$$\int_{\mathrm{BZ}} d\mathbf{k} \quad \longrightarrow \quad \sum_{j}^{N_{\mathbf{k}}} w_j, \tag{2.58}$$

This represents a weighted sum, with $\{w_j\}$ being the weights. Note that by increasing the number of $\mathbf{k}$-points the error can be systematically reduced. Several schemes to construct special $\mathbf{k}$-points exist [83, 84, 85]. In this work we use the scheme by Monkhorst and Pack [85]. It turns out that for insulators and semiconductors only very few $\mathbf{k}$-points are usually required to achieve well converged total energy. However, for metals, a larger number is needed to describe the variable ($\mathbf{k}$-dependent) occupation of bands. Finally, we note that usually one is interested in converged energy differences rather than converged total energies. Energy differences converge better with respect to $\mathbf{k}$-points sampling when *equivalent* $\mathbf{k}$-points are used. Later in this work (chapter 4) we will present some convergence tests for our systems with respect to $\mathbf{k}$-points sampling.

2.6.1 Periodic boundary conditions for surfaces: The supercell slab approach

While bulk crystals are periodic in the three dimensions, a vast variety of physically interesting systems are not. For instance, surfaces are periodic only in two dimensions (in-plane periodicity).

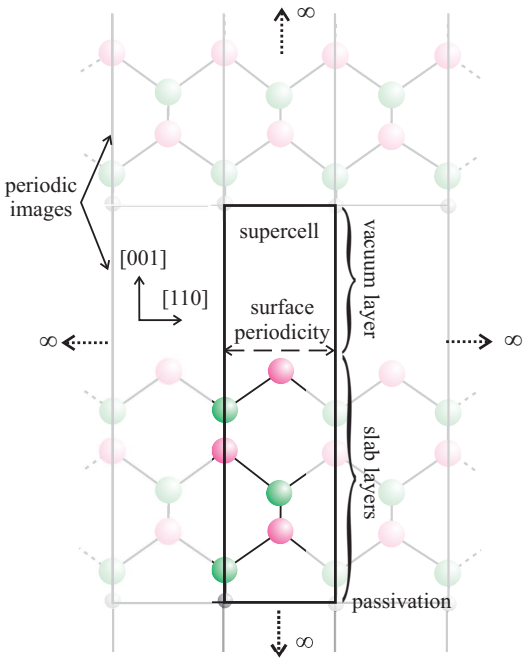


Figure 2.2: A two dimensional (side view) schematic representation of the supercell slab approach employed for calculating surfaces. The ideal (001) surface of the zinc-blend structure is used here. The atoms in the supercell are highlighted.

Furthermore, some structures are periodic in one dimension, such as long biomolecules and nanotubes, and others are non-periodic, such as molecules and point defects. Therefore, for such systems a special approach is required to allow treating them in methods that employ three dimensional periodicity. This is accomplished through the *supercell* approach.

Since in this work the study of surface systems is of particular interest, we briefly explain the supercell approach for surface calculations.[6] The supercell for surface calculations consists of a finite number of atomic layers, which together construct a "slab", and a vacuum layer. As the supercell is periodic in the three dimensions, the surface system is transformed to an infinite array of slabs separated by vacuum regions. The supercell slab approach is schematically explained in Fig. 2.2 for the ideal (001) surface of the zinc-blend crystal.

For this model system to be accurate enough two things need to be considered. First, the slab should be thick enough to reasonably describe the surface. Second, the vacuum layer should be thick enough in order to appropriately suppress the interaction between the slab and its periodic image in the perpendicular direction (the 3rd dimension). Therefore, convergence tests of the calculated quantity (e.g. surface energy) with respect to the thickness of the slab and the vacuum layers should be first performed. For our systems convergence tests are provided in Chapter 4. Finally we note that it is beneficial in many cases (e.g. when the surfaces on the upper side and the backside side of the slab are not identical (further discussion will follow in Section 5.3)) to passivate the backside of the slab by an additional layer of pseudohydrogens (partially-charged hydrogen atoms) [86]. This ensures its charge neutrality and also allows to reduce the thickness of the slab and vacuum layers.

[6] An extension of the supercell approach for other systems with lower dimensional periodicity is straightforward.

2.7 Solving the Kohn-Sham equations

In practice, the numerical solution of the Kohn-Sham equations proceeds by expanding the Kohn-Sham orbitals in a suitable *basis set*. The choice of a certain basis set is usually governed by performance versus accuracy. There is a vast variety of approaches available. In the following we briefly introduce few methods that fall into two main categories: (i) The plane-wave pseudopotential method and (ii) all-electron methods. For a more extended description of the various methods the reader may refer to Ref. [54]. We note that in this work two approaches have been used: The pseudopotential plane-wave (PS-PW) method within the S/PHI/nX *ab-initio* simulation package [87], and the projector-augmented waves (PAW) method (an all-electron method) within the VASP *ab-initio* simulation package [88]. Further computational details will follow later (Chapter 4).

2.7.1 The pseudopotential plane-wave method (PS-PW)

For extended periodic systems, one of the the most natural choices in solving the Kohn-Sham equations is the use of *plane-waves* as *basis* functions. Combined with the pseudopotential approximation, the use of a plane-waves basis results in a formally simple and effective method.

Let us expand the periodic function $u_{n\mathbf{k}}(\mathbf{r})$ (Eq. 2.55) in terms of Fourier series

$$u_{n\mathbf{k}}(\mathbf{r}) \quad = \quad \sum_{\mathbf{G}} C_{n\mathbf{k}}(\mathbf{G})\, e^{i\mathbf{G}\cdot\mathbf{r}}, \tag{2.59}$$

where the sum is over reciprocal space vectors $\mathbf{G}$. The $C_{n\mathbf{k}}$ are Fourier expansion coefficients. Substituting Eq. (2.59) in Eq. (2.54) we get

$$\psi_{n\mathbf{k}}(\mathbf{r}) \quad = \quad \sum_{\mathbf{G}} C_{n\mathbf{k}}(\mathbf{G})\, e^{i(\mathbf{G}+\mathbf{k})\cdot\mathbf{r}}. \tag{2.60}$$

This represents a plane-waves expansion of the wavefunctions. The single-particle Schrödinger equation (the Kohn-Sham Eq. (2.30)) can be transformed, after some mathematical manipulation, to an eigenvalue matrix problem of the form

$$\sum_{\mathbf{G}'} \left[\frac{1}{2} \left|\mathbf{k}+\mathbf{G}'\right|^2 \delta_{\mathbf{G}\mathbf{G}'} + V^{\mathrm{eff}}(\mathbf{G},\mathbf{G}') \right] C_{n\mathbf{k}}(\mathbf{G}') \quad = \quad \varepsilon_n(\mathbf{k}) C_{n\mathbf{k}}(\mathbf{G}), \tag{2.61}$$

where

$$V^{\mathrm{eff}}(\mathbf{G},\mathbf{G}') \quad = \quad V^{\mathrm{ps}}(\mathbf{G},\mathbf{G}') + V^{\mathrm{H}}(\mathbf{G},\mathbf{G}') + V^{\mathrm{XC}}(\mathbf{G},\mathbf{G}'). \tag{2.62}$$

These contributions to the effective potential are in practice described in terms of their Fourier transforms.

The practical solution of Eq. (2.61) is achieved by using a finite basis set. Therefore the plane-wave expansion is truncated at a certain value of the kinetic energy, called the *cutoff*

energy E_{cut}:

$$\frac{1}{2}\,|\mathbf{G}+\mathbf{k}|^2_{\text{max}} \quad \leq \quad E_{\text{cut}}. \tag{2.63}$$

The cutoff energy is a controllable parameter which should be carefully optimized depending on the specific pseudopotentials. The solution then proceeds iteratively for the required eigenstates, typically using conjugate gradient method [89], until self-consistent charge density, potential, and total energy are achieved.

2.7.2 All-electron methods

As the name indicates, in all-electron methods all of the electrons are included in the calculations, contrary to the pseudopotential method where only the valence electrons are considered. All-electron calculations become valuable in cases when there is a significant interaction between core and valence electrons. In addition, they provide information about the wavefunction close to the nucleus (i.e., not accessed with pseudopotentials) which is needed for many applications such as for the calculation of the hyperfine parameters [90].

The basic idea to make all-electron calculations feasible is to expand the electronic wavefunction in the core region (often called muffin-tin (MT)) in a different basis set than in the interstitial region. The muffin-tin spheres are chosen such that they do not overlap and also allow for structural relaxations. In the augmented plane waves method (APW), the potential is set to a constant in the interstitial region and a spherically symmetric in the muffin-tin regions. The augmented plane waves are then defined as

$$\varphi_{\mathbf{k},\varepsilon}(\mathbf{r}) \quad = \quad \begin{cases} e^{i\mathbf{k}\cdot\mathbf{r}} & \text{interstetial region} \\ \sum_{lm} a_{lm}(\mathbf{k},\varepsilon)\, u_l(r_i,\varepsilon)\, Y_{lm}(\theta,\phi) & \text{muffin-tin (MT) region} \end{cases}. \tag{2.64}$$

Here $Y_{lm}(\theta,\phi)$ are spherical harmonics, and $r_i = |\mathbf{r}-\mathbf{R}_i|$ where $\mathbf{R}_i$ is the position of the ith MT sphere (measured to its center). The a_{lm} coefficients are determined from the requirement that the augmented plane waves are continuous at the boundary of the muffin-tin spheres. The $u_l(r_i,\varepsilon)$ functions are solutions of the radial Schrödinger equation, with the muffin-tin potential. The electronic wavefunction is then written as a superposition of augmented plane waves:

$$\psi_{n\mathbf{k}}(\mathbf{r}) \quad = \quad \sum_{\mathbf{G}} C_{n\mathbf{k}}(\mathbf{G})\, \varphi_{\mathbf{G}+\mathbf{k},\varepsilon}(\mathbf{r}). \tag{2.65}$$

Because the augmented plane waves are energy dependent, the basis set can not be used for the whole energy spectrum and the solution of the secular equation becomes a nonlinear problem. This makes the APW method computationally rather inefficient.

The energy dependence problem can be avoided in the linearized augmented plane-wave (LAPW) method [91, 92], where the radial functions $u_l(r_i,\varepsilon)$ are expanded in a Taylor series around a fixed energy ε_l:

$$u_l(r,\varepsilon) \quad = \quad u_l(r,\varepsilon_l) + (\varepsilon-\varepsilon_l)\,\dot{u}_l(r_i,\varepsilon_l) + \cdots, \tag{2.66}$$

with $\dot{u}_l(r_i, \varepsilon_l)$ being the energy derivative at $\varepsilon = \varepsilon_l$. The expansion is truncated after the linear term. The LAPW basis functions are then given by

$$\varphi_{\mathbf{k}}(\mathbf{r}) = \begin{cases} e^{i\mathbf{k}\cdot\mathbf{r}} & \text{interstetial region} \\ \sum_{lm} \left[a_{lm}(\mathbf{k})\, u_l(r_i, \varepsilon_l) + b_{lm}(\mathbf{k})\, \dot{u}_l(r_i, \varepsilon_l)\right] Y_{lm}(\theta, \phi) & \text{MT region} \end{cases} . \quad (2.67)$$

The a_{lm} and b_{lm} coefficients are determined from the continuity requirement of both the wavefunction and its derivative across the muffin-tin boundary. Note that the LAPW basis depends on atomic positions and therefore the interatomic forces are not simply given by the Hellmann-Feynman theorem [57] (Sec. 2.3). To overcome this basis function corrections have to be taken into account (the so-called Pulay forces).

The full-potential LAPW method (FP-LAPW) applies the LAPW method without any shape restrictions on the potential, either in the muffin-tins or in the interstitial region. Since the full potential is taken into account, FP-LAPW calculations are considered to be the most accurate apart from exchange-correlation related errors, yet computationally more demanding.

The projector augmented-waves method (PAW) [93] combines the pseudopotential method and the LAPW in a natural way. In this method, the augmentation procedure is generalized such that the partial-waves expansions are not determined through the continuity requirement of the wavefunction and its derivative across the muffin-tins but through the overlap with localized projector functions. In the heart of the PAW method lies a linear transformation between the true all-electron wavefunction (ψ) and a a soft pseudo (auxiliary) wavefunction ($\tilde{\psi}$):

$$|\psi\rangle = \hat{\tau}|\tilde{\psi}\rangle. \quad (2.68)$$

The wavefunctions $\tilde{\psi}$ and ψ are expanded in terms of partial waves $\tilde{\phi}_i$ and ϕ_i, respectively. The transformation $\tilde{\tau}$ then takes the form:

$$\hat{\tau} = 1 + \sum_i (|\phi_i\rangle - |\tilde{\phi}_i\rangle)\langle\tilde{p}_i|, \quad (2.69)$$

where $\tilde{p}_i$ are localized projector functions chosen such that they obey the relation

$$\langle\tilde{p}_i|\tilde{\phi}_j\rangle = \delta_{ij}. \quad (2.70)$$

The method then proceeds by transforming also the operators to extract observables in terms of the pseudo operators [93, 94]. The main advantage of the PAW is that it provides reduced computational efforts within all-electron calculations accuracy.

2.8 Summary

In this chapter we have briefly introduced the *ab-initio* approach to material science and the involved approximations, which is the most accurate approach for electronic structure calculations. In the framework of this study we have employed two methods among the several available

methods for solving the Kohn-Sham equations for our calculations. These are the pseudopotential plane-waves method (PS-PW) and the projector augmented-waves method (PAW). Details about these choices and further computational details used for our systems are presented in Chapter 4. Keeping in mind the *ab-initio* approach, in the following chapter we briefly introduce and test the density functional-based tight-binding method, which is not *ab-initio* but rather based on *ab-initio* calculations, as an alternative promising approach to improve the computational efficiency.

Chapter 3

Density Functional-Based Tight-Binding: Methodology, Implementation, and Examination

3.1 Introduction

In the previous chapter we have introduced *ab-initio* methods for calculating the electronic structure of materials. With the currently available computational power, typical *ab-initio* methods allow to calculate up to few hundreds of atoms in a single unit cell. While the size limit is not critical for the systems to be considered in this work, it is appealing to increase the computational speed efficiency of our calculations without losing accuracy. An approach to achieve this is by initializing the *ab-initio* calculations with good pre-relaxed structures rather than starting from completely unrelaxed ones, which allows to save many structural optimization steps. However, to achieve the desired reduction in the computational time, an alternative more efficient but less accurate method than *ab-initio* methods has to be used for the per-relaxation calculations.

Various approaches can be employed for this purpose. For instance, empirical potentials allow for calculations that are orders of magnitudes faster than standard *ab-initio* methods. In this approach the quantum mechanical interactions between atoms are described based on a fitting to empirical data. Therefore they suffer from their very limited transferability, i.e, they are not able to correctly describe the atoms in different chemical environments.

The tight-binding (TB) method lies between the expensive more accurate *ab-initio* methods and the fast but limited accuracy empirical methods [95]. It is a *semi-empirical* method as it retains a part of the quantum mechanical description of atomic bonding. The TB approach is based on the the early work of Slater and Koster [96]. In the standard TB method, the eigenstates of the effective one-particle Hamiltonian are expanded in an orthogonal atomic-like basis set, with the many-body Hamiltonian replaced by parametrized Hamiltonian matrix elements. These matrix elements are fitted to the band structure of a suitable reference system which affects the transferability of the method.

One approach that has been proposed to overcome the problems of the standard TB method

is the density functional based tight-binding method (DFTB) [97] and its self-consistent-charge extension (SCC-DFTB) [98, 99]. The key feature of the DFTB method is the determination of the matrix elements via a new fitting procedure which is based on *ab-initio* DFT calculations, which promises to achieve better transferability. Therefore, we have here adapted the DFTB approach.

When included in an *ab-initio* code, the DFTB method is useful for several purposes. Besides pre-relaxing atomic structures, it can be used e.g. to perform inexpensive molecular dynamics calculations for testing purposes. Furthermore, some applications go beyond of the reach of *ab-initio* calculations, or simply they become extremely costly. For example, these applications include calculating amorphous materials, dislocations, large biological systems, quantum dots, surfaces and interfaces, as well as large time-scale simulations (e.g. molecular dynamics). For such cases it becomes worthwhile to consider the DFTB approach that can overcome some size/time limitations.

Therefore, during the early stages of the present work, the DFTB method (including the SCC extension) has been successfully implemented in the object-oriented modular *ab-initio* simulation package S/PHI/nX [87]. Few test calculations have been performed to test its performance for relevant physical properties in this work. Unfortunately, these tests have shown that the accuracy and the transferability of the DFTB method are not sufficient to applied here. Nonetheless, the S/PHI/nX version of DFTB has found some useful applications among the members in the department, specially for biological systems [100, 101].

In this chapter we restrict ourselves to a very brief introduction to the SCC-DFTB methodology and its implementation, and our accuracy performance test calculations. The SCC-DFTB methodology has been a subject of several PhD theses [102, 103, 104] and other publications [99, 98]. Further details on the method can be hence found in these references.

3.2 The DFTB Formalism

3.2.1 Zeroth-order approximation: Standard DFTB

In DFTB, the total energy is expanded around a reference (or input) density (n_0) up to second order in density fluctuations [105, 99]. To derive the total energy formula we write the true self-consistent charge density $n(\mathbf{r})$ as a superposition of a reference density n_0 plus a small density fluctuation δn, i.e. $n = n_0 + \delta n$, in the Kohn-Sham total energy functional. Expanding the energy functional using a Taylor series and truncating after second order terms one obtains [105]

$$\begin{aligned} E[n(\mathbf{r})] &= \sum_{i}^{occ} n_i \langle \psi_i | H_0 | \psi_i \rangle - \frac{1}{2} \int\int \frac{n_0(\mathbf{r}')n_0(\mathbf{r})}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r} d\mathbf{r}' \\ &\quad + E^{\mathrm{XC}}[n_0(\mathbf{r})] - \int V^{\mathrm{XC}}[n_0(\mathbf{r})]\, n_0(\mathbf{r}) d\mathbf{r} \\ &\quad + \frac{1}{2} \int\int \left[\frac{1}{|\mathbf{r}-\mathbf{r}'|} + \left. \frac{\delta^2 E^{\mathrm{XC}}[n(\mathbf{r})]}{\delta n(\mathbf{r}) \delta n(\mathbf{r}')} \right|_{n_0} \right] \delta n(\mathbf{r})\, \delta n(\mathbf{r}')\, d\mathbf{r} d\mathbf{r}'. \end{aligned} \tag{3.1}$$

Note that the first and second lines give the zero-th order terms in density fluctuations δn, the third line gives the second order term in δn. All first order contributions cancel. The first term in the upper equation is the band-structure energy, with H_0 being the Hamiltonian operator that depends only on the reference density n_0. We introduced here the occupation numbers n_i (where $N_\mathrm{e} = \sum_i^{occ} n_i$) for convenience.[1] By neglecting second order terms, one arrives to the total energy functional of the zero-th order DFTB, or the standard DFTB. This total energy functional is further simplified and expressed in the form:

$$E_0^{\mathrm{DFTB}} = \sum_i^{occ} n_i \langle \psi_i | H_0 | \psi_i \rangle + E_{\mathrm{rep}}. \tag{3.2}$$

Here, the second term (E_{rep}), is a repulsive pairwise short-range energy.

Let us expand the Kohn-Sham wave functions ψ_i in terms of localized atomic-like orbitals

$$|\psi_i(\mathbf{r})\rangle = \sum_\nu^M c_{\nu i} |\varphi_\nu(\mathbf{r} - \mathbf{R}_I)\rangle, \tag{3.3}$$

where $c_{\nu i}$ are the expansion coefficients, I labels the atom, ν labels the type of atomic orbital (s, p, d, etc). Inserting Eq. (3.3) in the total energy expression and applying the variational principle with respect to the expansion coefficients under the constraint of constant number of electrons leads to the DFTB secular equations:

$$\sum_\nu c_{\nu i} (H^0_{\mu\nu} - \varepsilon_i S_{\mu\nu}) = 0, \quad \forall\, \mu, i. \tag{3.4}$$

Here, $H^0_{\mu\nu}$ and $S_{\mu\nu}$ are the Hamiltonian and overlap matrix elements, respectively, which are defined as:

$$\begin{aligned} H^0_{\mu\nu} &= \langle \varphi_\mu | H_0 | \varphi_\nu \rangle; \\ S_{\mu\nu} &= \langle \varphi_\mu | \varphi_\nu \rangle, \qquad \forall\ \mu \in I,\ \nu \in J, \end{aligned} \tag{3.5}$$

and they contain only two-center terms. The Hamiltonian matrix elements are calculated according to

$$H^0_{\mu\nu} = \begin{cases} \varepsilon_\mu^{\text{free atom}} & \mu = \nu \\ \langle \varphi_\mu | T + V_I + V_J | \varphi_\nu \rangle & \mu \in I, \nu \in J. \end{cases} \tag{3.6}$$

Note that the eigenvalues of the free atom in the diagonal Hamiltonian matrix elements guarantee the correct limit for an isolated atom. Due to symmetry, only a few (non-zero) types of integrals enter in the matrix elements. For instance, for up to two angular momenta $l = 2$ (i.e. s-, p-, and d-bonded system) there are only 10 types of integrals: $dd\sigma$, $dd\pi$, $dd\delta$, $pd\sigma$, $pd\pi$, $pp\sigma$, $pp\pi$, $sd\sigma$, $sp\sigma$, and $ss\sigma$ (represented in the standard molecular orbital bonding notation [106]). These integrals (for the Hamiltonian and the overlap) are calculated for each *pair* of atomic species

[1] For spin compensated systems the maximum occupation number is 2.

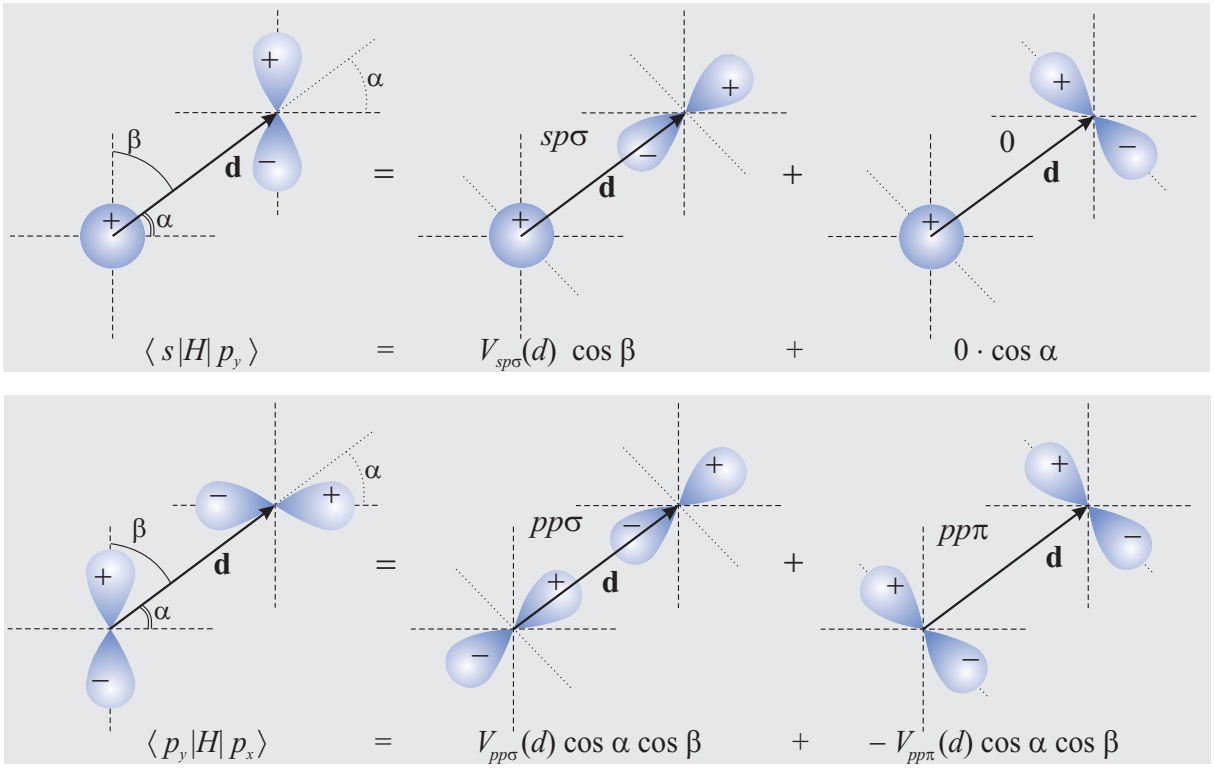


Figure 3.1: Schematic illustration for calculating the Hamiltonian interatomic matrix elements (for $\langle s|H|p_y\rangle$ and $\langle p_y|H|p_x\rangle$) using the direction cosines and the tabulated integrals.

of interest using self-consistent *ab-initio* calculations and are tabulated as function of distance. Confined pseudo-atoms are usually used for the calculation of these integrals as they better resemble the pseudo-atoms in the molecular/condensed matter environments, resulting in an improved description. For the system of interest, the Hamiltonian and overlap matrix elements are then constructed using the tabulated integrals and the direction cosines[2]. This is explained in the simple schematic illustration in Fig. 3.1 for s and p orbitals. Extended transformations, including the d orbitals, can be found somewhere else [96, 107, 106]. The use of the tabulated integrals hence results in a dramatic increase in the computational speed.

The repulsive energy E_{rep} is obtained as a function of distance by calculating the difference between self-consistent *ab-initio* calculations and the TB band-structure energy, in a suitable *reference structure*:

$$E_{\text{rep}}(r) \quad = \quad \left\{ E_{\text{tot}}^{\text{DFT}}(r) - \sum_{i}^{occ} n_i \varepsilon_i(r) \right\} \Bigg|_{\text{ref. str.}} . \tag{3.7}$$

Similar to the matrix elements, E_{rep} is calculated as function of the distance between each pair of atomic species within a certain cutoff radius and tabulated. The repulsive energy contribution for the system of interest is then calculated using the tabulated values. Hence, a critical issue for E_{rep} is its transferability, i.e., whether the choice of the reference structure changes the curve

[2]The direction cosines of the vector connecting from one atom to another ($\mathbf{d}$) are the cosines of the angles that the vector makes with the three axes: $\cos\alpha = \frac{\mathbf{d}\cdot\hat{i}}{d}$, $\cos\beta = \frac{\mathbf{d}\cdot\hat{j}}{d}$, and $\cos\gamma = \frac{\mathbf{d}\cdot\hat{k}}{d}$

describing $E_{\rm rep}$ or not. It has been found that atoms in different *molecules* produce repulsive energies that are *close* to each other, which does not hold for highly-coordinated crystalline phases [103]. Therefore, $E_{\rm rep}$ is only approximately transferable, and this leads to an additional inaccuracy.

The eigenvalues are then determined by solving the secular equations (Eq. (3.4)). The total energy is calculated using (see Eq. (3.2)):

$$E_0^{\rm DFTB} = \sum_i^{occ} n_i \varepsilon_i + E_{\rm rep}. \tag{3.8}$$

The interatomic forces can be calculated by taking the derivative of the DFTB energy (Eq. (3.8)) with respect to nuclear coordinates. This reads

$$\begin{aligned} m_I \frac{\partial^2}{\partial t^2}\mathbf{R}_I &= -\frac{\partial E_0^{\rm DFTB}}{\partial \mathbf{R}_I} \\ &= -\sum_i^{occ} n_i \sum_\mu \sum_\nu c_{\mu_i}^* c_{\nu i} \left[\frac{\partial H_{\mu\nu}^0}{\partial \mathbf{R}_I} - \varepsilon_i \frac{\partial S_{\mu\nu}}{\partial \mathbf{R}_I}\right] - \frac{\partial E_{\rm rep}}{\partial \mathbf{R}_I}. \end{aligned} \tag{3.9}$$

Here, use of Eq. (3.3) has been made. Atomic charges are evaluated as Mulliken charges:

$$q_I = \frac{1}{2}\sum_i^{occ} n_i \sum_{\mu\in I}\sum_\nu \left(c_{\mu i}^* c_{\nu i} + c_{\nu i}^* c_{\mu i}\right). \tag{3.10}$$

These charges will be useful for the SCC-DFTB schema described in the following.

3.2.2 Second-order approximation: Self-consistent charge DFTB

For systems that show considerable charge transfer the standard DFTB method may not give sufficient accuracy. Therefore, in order to improve total energies, forces, and transferability, the second order terms in the density fluctuations are included in the SCC-DFTB. By including the second order term in Eq. (3.1), the total energy in DFTB can be written as

$$E_2^{\rm DFTB} = \sum_i^{occ} n_i \langle\psi_i|H_0|\psi_i\rangle + E_{\rm 2nd} + E_{\rm rep}. \tag{3.11}$$

The second order term $E_{\rm 2nd}$ is further approximated to take the form

$$E_{\rm 2nd} = \frac{1}{2}\sum_{I,J}^{N_{\rm nuc}} \Delta q_I \Delta q_J \gamma_{IJ}. \tag{3.12}$$

Here, $\Delta q_I = q_I - q_I^0$. The term γ_{IJ}, after some approximations and lengthy mathematical manipulations [102], is expressed in the following form

$$\gamma_{IJ}(\tau_I, \tau_J, R) = \frac{1}{R} - \mathcal{F}(\tau_I, \tau_J, R), \tag{3.13}$$

with

$$\begin{aligned}\mathcal{F}(\tau_I,\tau_J,R) &= e^{-\tau_I R}\left(\frac{\tau_J^4\tau_I}{2(\tau_I^2-\tau_J^2)^2}-\frac{\tau_J^6-3\tau_J^4\tau_I^2}{(\tau_I^2-\tau_J^2)^3R}\right)\\ &+ e^{-\tau_J R}\left(\frac{\tau_I^4\tau_J}{2(\tau_J^2-\tau_I^2)^2}-\frac{\tau_I^6-3\tau_I^4\tau_J^2}{(\tau_J^2-\tau_I^2)^3R}\right); \quad \text{for } \tau_I\neq\tau_J, \end{aligned} \tag{3.14}$$

and

$$\mathcal{F}(\tau,R) = e^{-\tau R}\left(\frac{48+33\tau R+9(\tau R)^2+(\tau R)^3}{48R}\right); \quad \text{for } \tau_I=\tau_J=\tau. \tag{3.15}$$

Here $R=|\mathbf{R}_I-\mathbf{R}_J|$. The determination of the τ parameters will follow shortly. Note that in the limit of long atomic distances $\gamma_{IJ}\to 1/R$, thus representing a Coulomb interaction between the two charges Δq_I and Δq_J. On the other hand, the on-site contributions, i.e., in the limit as $R\to 0$ (with $J\equiv I$) read

$$\gamma_{II}=\lim_{R\to 0}\gamma_{IJ}=\frac{5}{16}\tau_I. \tag{3.16}$$

It has been suggested that γ_{II} can be approximated by the difference of the atomic ionization potential I_I and the electron affinity A_I [108], as widely applied in semi-empirical quantum chemistry methods. This difference is related to the chemical hardness η_I or the Hubbard parameter U_I [109]:

$$\gamma_{II}\approx I_I-A_I\approx 2\eta_I\approx U_I. \tag{3.17}$$

Therefore, using Eqs. (3.16) and (3.17)), the τ parameters can be determined in terms of Hubbard parameters:

$$\tau_I = \frac{16}{5}U_I. \tag{3.18}$$

For periodic systems, it is sufficient to explicitly calculate the short-range part of γ_{IJ} over a small number of cells. However, the long range Madelung-like part is calculated employing the standard Ewald technique [89, 81].

Starting from the DFTB energy up to second order in energy fluctuations

$$E_2^{\mathrm{DFTB}} = \sum_i^{occ} n_i\langle\psi_i|H_0|\psi_i\rangle+\frac{1}{2}\sum_{I,J}^{N_{\mathrm{nuc}}}\Delta q_I\Delta q_J\gamma_{IJ}+E_{\mathrm{rep}}, \tag{3.19}$$

and by applying the variational principle with respect to the expansion coefficients under the constraint of constant number of electrons leads us to the secular equations:

$$\sum_\nu c_{\nu i}(H_{\mu\nu}-\widetilde{\varepsilon}_i S_{\mu\nu}) = 0, \quad \forall\,\mu,i. \tag{3.20}$$

The Hamiltonian and overlap matrix elements are:

$$\begin{aligned} H_{\mu\nu} &= \langle\varphi_\mu|H_0|\varphi_\nu\rangle + \frac{1}{2}S_{\mu\nu}\sum_{K}^{N_{\text{nuc}}}(\gamma_{IK}+\gamma_{JK})\Delta q_K \\ &= H^0_{\mu\nu} + H^1_{\mu\nu}; \\ S_{\mu\nu} &= \langle\varphi_\mu|\varphi_\nu\rangle, \qquad \forall\ \mu\in I,\ \nu\in J. \end{aligned} \tag{3.21}$$

The repulsive energy term is now determined from the tabulated values obtained by calculating the difference between *ab-initio* self-consistent calculations and the SCC-DFTB electronic energy (the first two terms in Eq. (3.19)), for a suitable reference structure.

Interatomic forces are derived by taking the derivative of the SCC-DFTB energy in Eq. (3.19) with respect to nuclear coordinates. Hence, the interatomic forces within the SCC-DFTB take the form

$$\begin{aligned} \mathbf{F}_I &= -\sum_{i}^{occ} n_i \sum_{\mu,\nu} c^*_{\mu i}c_{\nu i}\left[\frac{\partial H^0_{\mu\nu}}{\partial \mathbf{R}_I} - \left(\widetilde{\varepsilon}_i - \frac{H^1_{\mu\nu}}{S_{\mu\nu}}\right)\frac{\partial S_{\mu\nu}}{\partial \mathbf{R}_I}\right] \\ &\quad -\Delta q_I \sum_{L}^{N_{\text{nuc}}} \frac{\partial \gamma_{IL}}{\partial \mathbf{R}_I}\Delta q_L - \frac{\partial E_{\text{rep}}}{\partial \mathbf{R}_I}. \end{aligned} \tag{3.22}$$

For periodic systems, the derivative of γ_{IJ} is also calculated using the Ewald technique.

3.3 Implementation

The DFTB method has been implemented in the *ab-initio* simulation package S/PHI/nX [87]. The implementation within S/PHI/nX has the great advantage of accessing its computationally optimized algebra library, which makes implementing mathematical formulas including vector and matrix operations very easy. In addition, because of its modular structure, it is possible to make use of the various already existing routines such as structure optimization routines and charge mixers. This allows to focus on the methodology itself, which resulted in a relatively short implementation time.

The implemented code supports both periodic and non-periodic systems (e.g. clusters). In addition, the self-consistent charge extension (SCC) has been also included. Various careful checks have been made with the DFTB code [110, 98] (e.g., energy-volume curves, band structures) to ensure that our code reproduces exactly the same results when using the same potential files. Additionally, because of the direct access to *ab-initio* calculations within the S/PHI/nX code, primary steps towards generating and testing the parameter files (Slater-Koster files) have been made. This allows for the generation of any parameter files as required in future.

Benchmarks for the speed performance show that DFTB calculations are 2–3 orders of magnitude faster than typical *ab-initio* plane-waves pseudopotential calculations. The most expensive part of DFTB calculations is the solution of the generalized eigenvalue problem (Eq. (3.4) and (3.20)). Unlike *ab-initio* calculations with plane-waves basis, DFTB calculations scale with the number of atoms not with the volume of the cell, which makes it valuable

for systems that include large vacuum such as molecules, surfaces, etc.

3.4 DFTB accuracy performance

An important issue to consider before using the DFTB method is to examine its accuracy for the systems of interest and for the interesting physical properties/quantities to be investigated during this work. Therefore, we have considered three key properties. First, we have calculated the structural and cohesive properties for selected material systems, which include the bulk Ga, As, and GaAs; and the N_2 molecule. Second, we have calculated the phase diagram of the well-established GaAs(001) surface, which is the growth surface for $GaAs_{1-x}N_x$ alloys and hence it is the starting point towards understanding the solubility of N in GaAs. Finally, in order to check the performance of DFTB in calculating adatom binding energies and diffusion barriers, we have investigated adatom-surface interaction. In specific, we have considered a Ga adatom above the $\beta 2(2\times 4)$ reconstructed GaAs(001) surface[3] since this allows to directly compare with the already-existing detailed *ab-initio* results by Kley et al. [111].

We note that the parameter files or the Slater-Koster files (containing the Hamiltonian and overlap matrix elements and the repulsive potential parameters) that we have used here were generated by the developers of the DFTB code [110, 98], within the LDA approximation for the exchange and correlation. All other computational details (e.g. **k**-points sampling, number of atomic layers for the slab calculations, etc) were taken to be identical to those used in our *ab-initio* calculations, which are provided in Chapter 4.

3.4.1 Structural and cohesive properties

In order to check the performance of DFTB in predicting the bulk properties we have calculated the structural and cohesive properties for several systems relevant for this work, which are Ga, As, and GaAs bulks, and the N_2 molecule. First, we have optimized the the volume of the unit cells[4] for the bulk systems (using Murnaghan's equation of state [112]), as well as the N_2 bond length. Note that the atomic structures and their internal parameters are presented in detail in the following chapter (Chapter 4). After that we have calculated the cohesive properties (see Sec. 4.3.2).

The results are summarized in Table 3.1. Previous experimental results are also shown for comparison. First, within the standard-DFTB, the results show that the unit cell volume is overestimated for bulk Ga and for bulk As, and it is slightly overestimated for bulk GaAs. This is in contrary to the expected underestimation when using the LDA exchange-correlation functional. Including the self-consistency charge cycle (SCC-DFTB) for GaAs resulted in a very slight insignificant improvement in the lattice constant. In overall all the structural parameters are still in acceptable agreement with experiment.

The calculated cohesive energies for Ga, As, and GaAs and the binding energy for the N_2 molecule are all overestimated, by up to 28 % for Ga bulk. Moreover, it is important to note that

[3] Surface reconstructions are introduced in Chapter 5.

[4] The internal structural parameters were taken from the *ab-initio* optimized values presented in Chapter 4.

		a	b	c	u	v	E_c
Ga	DFTB	4.76	8.22	4.91	0.156	0.078	3.59[a]
	Exp.[b]	4.51	7.63	4.50	0.153	0.079	2.81
As	DFTB	3.84	–	10.55	0.222	–	3.50[a]
	Exp.[c]	3.76	–	10.44	0.228	–	2.96
		a	E_c	ΔH_f			
GaAs	DFTB	5.66	7.95[a]	−0.86			
	SCC-DFTB	5.65	7.94[a]	−0.85			
	Exp.[d]	5.65	6.52	−0.74			
		d	E_b				
N_2	DFTB	1.08	11.33[a]				
	Exp.[e]	1.10	9.80				

[a] Spin correction included from Refs. [113, 114]
[b] Ref. [115] (a, b, c, u, v) and Ref. [81, p50] (E_c)
[c] Ref. [116] (a, c, u) and Ref. [81, p50] (E_c)
[d] Ref. [106] (a, E_c) and Ref. [117] (ΔH_f)
[e] Ref. [117]

Table 3.1: Calculated structural and cohesive properties for bulk Ga, As, and GaAs, and the N_2 molecule using DFTB, compared to experimental values. Here E_c, ΔH_f, and E_b denote the cohesive energy, the formation enthalpy, and the binding energy, respectively. The structural internal parameters are introduced in Chapter 4. All dimensions are in Å, and all energies are in eV.

including the charge self-consistency for GaAs did not lead to a considerable improvement over the standard-DFTB. For a better visualization of the performance of DFTB versus experiment these values are plotted and shown in Fig 3.2. Note that for convenience the figure includes additionally the corresponding *ab-initio* results obtained in this work using PS-PW and PAW basis (see Chapter 4) within the GGA approximation, taken from Tables 4.1, and 4.2.

3.4.2 GaAs(001) surface phase diagram

In order to check the performance of DFTB with respect to surface energies we have calculated the stability phase diagram of the GaAs(001) surface reconstructions as a function of growth conditions (partial pressures). The calculations were performed using both the standard DFTB [118] and the SCC-DFTB [98]. Note that information about surface energy and reconstructions are first provided in Chapter 5, where the atomic structures of various GaAs(001) surface reconstructions are provided.

The calculated phase diagram using the standard DFTB is shown in Fig. 3.3(a). The figure shows the poor performance of the standard DFTB with respect to surface energies. A major deficiency is that the $\beta 2(2\times4)$ reconstruction is predicted to be unstable in the thermodynamically allowed region, and it is almost degenerate with the $\beta(2\times4)$ reconstruction. This contradicts to what is known from both *ab-initio* calculations and experiment [120]. Similarly, the $\alpha 2(2\times4)$ and $\alpha(2\times4)$ reconstructions are almost degenerate. Moreover, the other reconstructions of mixed-dimer(2×4) (md(2×4)) and $\zeta(4\times2)$ are predicted to be unstable in the full thermodynamically

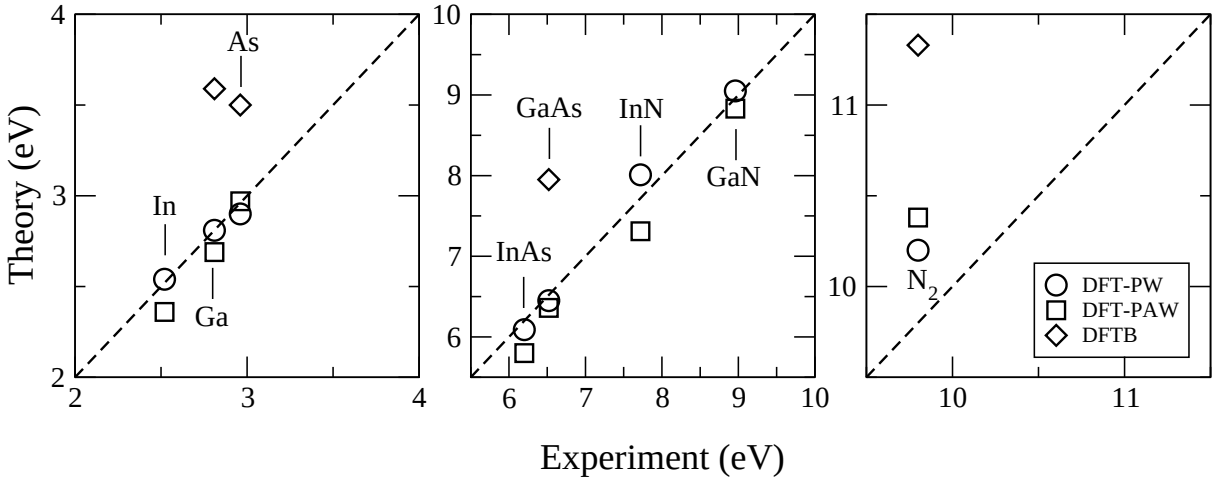
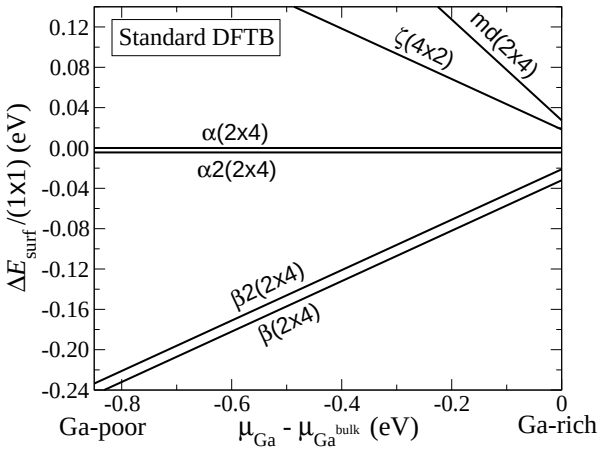


Figure 3.2: Experimental versus theoretical DFTB cohesive energies of unary bulk systems (left), binary bulk systems (middle), and the binding energy of the N_2 molecule (right); split to allow for a better resolution due to the scale difference. The values are taken from Table 3.1. Additional *ab-initio* results employing GGA and using PS-PW (denoted DFT-PW) or PAW basis are included for convenience (taken from Tables 4.1 and 4.2).

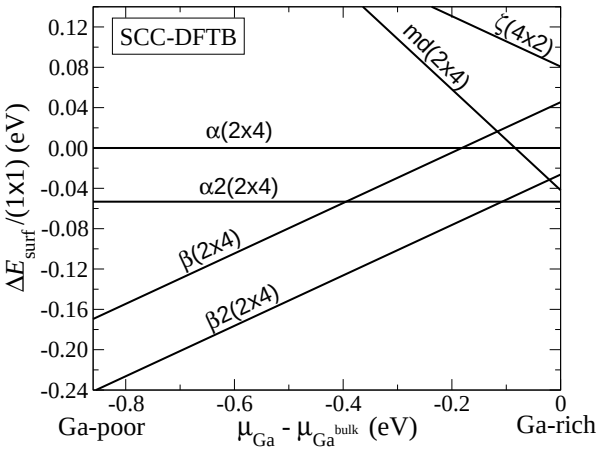
allowed region. These results are in large discrepancy with *ab-initio* (DFT-LDA) calculations that are shown in Fig. 3.3(c) for comparison (taken from Ref. [119]).

Let us now see the effect of including the self-consistent charge cycle within DFTB (SCC-DFTB). The resulting phase diagram is presented in Fig. 3.3(b). It shows that SCC indeed leads to a substantial improvement over the standard DFTB results with respect to surface energies. Specifically, the improvement is mainly on the relative stabilities of $\alpha 2(2\times 4)$ versus $\alpha(2\times 4)$, and $\beta 2(2\times 4)$ versus $\beta(2\times 4)$ reconstructions. This is a consequence of a better description of the electrostatic interactions within SCC-DFTB, being an important mechanism behind the stability of $\alpha 2(2\times 4)$ and $\beta 2(2\times 4)$ over $\alpha(2\times 4)$ and $\beta(2\times 4)$, respectively [121, 119]. The better performance of SCC-DFTB over the standard DFTB for the case of $\beta 2(2\times 4)$ was already reported earlier [118, 98]. On the other hand, the SCC-DFTB phase diagram in Fig. 3.3(b) shows that both the $\mathrm{md}(2\times 4)$ and the $\zeta(4\times 2)$ reconstructions are still predicted to be highly unstable, which in clear disagreement with the DFT-LDA calculations shown Fig. 3.3(c).

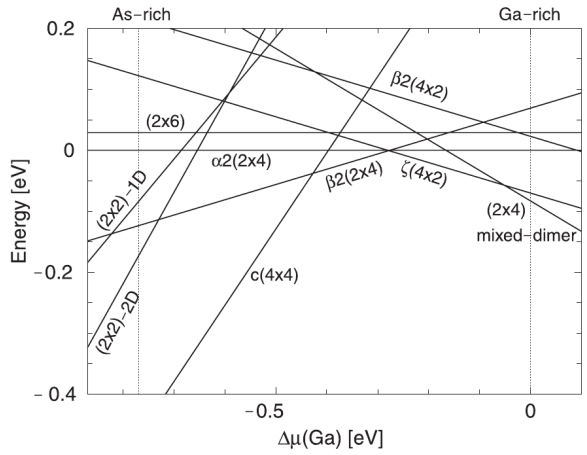
Finally, it is interesting to compare our SCC-DFTB GaAs(001) phase diagram with that of Elstner et al. [98] calculated previously, which is shown in Fig. 3.3(d). There are only three common reconstructions between Figs. 3.3(b) and (d) that can be compared. These are the $\beta 2(2\times 4)$, $\beta(2\times 4)$, and $\alpha(2\times 4)$ reconstructions. First, in Ref. [98] the authors reported that the $\beta 2(2\times 4)$ reconstruction is energetically more favorable than the $\beta(2\times 4)$ reconstruction by $3.7\,\mathrm{meV/\AA^2}$. From our phase diagram (Fig. 3.3(b)) the corresponding energy difference reads $71.7\,\mathrm{meV}/(1\times 1) = 4.5\,\mathrm{meV/\AA^2}$, i.e., the agreement between the two calculations is very good ($0.8\,\mathrm{meV}$). Second, in the phase diagram in Fig. 3.3(d) the $\beta 2(2\times 4)$ and $\alpha(2\times 4)$ reconstructions become equally stable (i.e. have the same surface energy) at $\mu - \mu_{\mathrm{Ga^{bulk}}} \approx -0.25\,\mathrm{eV}$. Beyond this value in the direction of more Ga-rich conditions the $\alpha(2\times 4)$ becomes more stable. This can be understood since the more stable $\alpha 2(2\times 4)$ reconstruction was not taken into account. On the



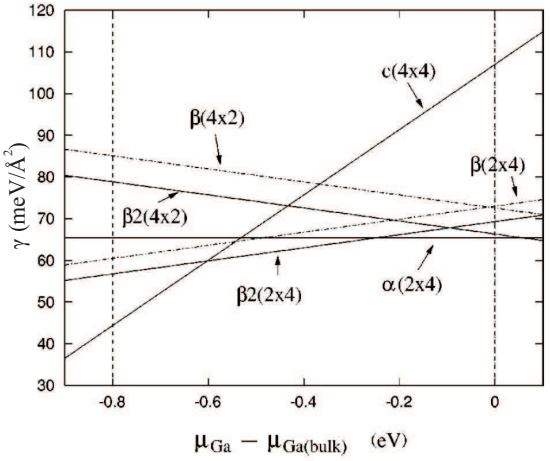
(a) Standard DFTB



(b) SCC-DFTB



(c) DFT-LDA, from Ref. [119]



(d) SCC-DFTB, from Ref. [98]

Figure 3.3: Phase diagrams of the GaAs(001) surface: (a) using the standard DFTB method, (b) using SCC-DFTB method, (c) using DFT-LDA; taken from Schmidt [119], and (d) using SCC-DFTB; taken from Elstner et al. [98]. In (c) the energy is per the (1×1) surface area. The (1×1) area equals $\frac{a^2}{2} = 15.96\,\text{Å}^2$ for (a) and (b), and $15.51\,\text{Å}^2$ for (c).

other hand, in our case the surface energies of the $\beta 2(2 \times 4)$ and $\alpha(2 \times 4)$ reconstructions become equal only outside the thermodynamically allowed region, at a chemical potential μ_{Ga} that is different by $\sim 0.35\,\text{eV}$ compared to the case in the phase diagram in Fig. 3.3(d). Apart from surface energies, one further quantity can be compared in the two calculations, which is the the formation enthalpy of GaAs bulk. From Fig. 3.3(d) the GaAs formation enthalpy reads $-0.80\,\text{eV}$ (the difference between Ga-rich and Ga-poor chemical potential limits (see Eq. (5.10)), while in our case it reads $-0.85\,\text{eV}$ (from Table 3.1). Unfortunately there are no further quantities available to compare with (e.g., bulk properties).

It becomes important here to determine the source of these discrepancies. First, we note that the employed computational details employed in Ref. [98] are not available, such as **k**-points sampling, the number of slab layers and relaxed layers used in the slab calculations, and about the determination of the thermodynamic Ga-rich and Ga-poor limits (e.g. the reference

structures and the corresponding computational details). On the other hand, we note that there are no information available whether the parameter files that we have used here are identical to those used in Ref. [98]. Since it is unlikely that the large discrepancies we have found are due to differences in the computational details, the most probable reason for the discrepancies is the use of different parameter files (Slater-Koster files). In specific, as we have noted earlier, the repulsive potentials are sensitive to the choice of the reference structures (environments) used to fit the energy. This indicates that the DFTB parameter files suffer from a transferability problem.

3.4.3 Adatom-surface interaction: Ga adatom

For this work it is important to test the performance of DFTB in calculating binding energies and kinetic barriers of adatoms on surfaces. Therefore, as a final examination, we have studied the interaction of a Ga adatom above the $\beta 2(2 \times 4)$ reconstruction of GaAs(001) surface. This choice is motivated by the existence of the previous detailed *ab-initio* results by Kley et al. [111] which can be used for comparison.[5] Note that in their study GGA was used for the exchange and correlation, however, this is not critical as long as we are interested to check the general performance qualitatively.

Let us first consider the potential energy of a Ga-adatom as a function of its height above the center of an As dimer of the $\beta 2(2 \times 4)$ GaAs(001) surface (see Fig. 5.3 in Chapter 5). In Ref. [111] the authors reported two different minima associated with two different bonding configurations between the Ga-adatom and the As-dimer, as shown in Fig. 3.4(b). In the first bonding configuration $m1$ which occurs at adatom height of $z_{\text{adatom}} - z_{\text{dimer}} \approx 2\,\text{Å}$, the adatom interacts with the completely filled dangling bonds of the dimer; while in the second bonding configuration $m2$, at $z_{\text{adatom}} - z_{\text{dimer}} \approx 0.1\,\text{Å}$, the adatom breaks the dimer bond and forms directional bonds with the adjacent As atoms. The energy difference between the two minima is $\sim 1.2\,\text{eV}$, separated by an energy barrier of $\sim 1\,\text{eV}$ at $z_{\text{adatom}} - z_{\text{dimer}} \approx 1.2\,\text{Å}$.

In order to test whether the DFTB method is able to correctly describe these two different bonding situations we have performed the same mapping of the potential energy employing the standard-DFTB. For these calculations, the Ga adatom is kept fixed at a certain position above the center of the As dimer and the surface atoms are allowed to fully relax, for various adatom heights. The structure of the clean surface has been used as the input/initial structure for all adatom heights. The obtained height versus energy values are shown by the black dots in Fig. 3.4(a). Note that it was not possible to map the energy for the points that are closer than a certain value to the surface due to the strong repulsive energy which destroys the structure in the initial structural optimization steps. To overcome this problem we have employed an adiabatic structural update scheme. That is, the relaxed structure obtained for a certain adatom height is taken as the initial input structure for the next (nearby) point, with changing adatom's height. Two adiabatic update directions were considered: In the first path the initialization starts from the highest adatom position (above the surface), while in the second it starts from the lowest

[5] Similar to Ref. [111], the calculations were performed in a 4×4 surface cell to reduce fictitious adatom interactions from the periodic images.

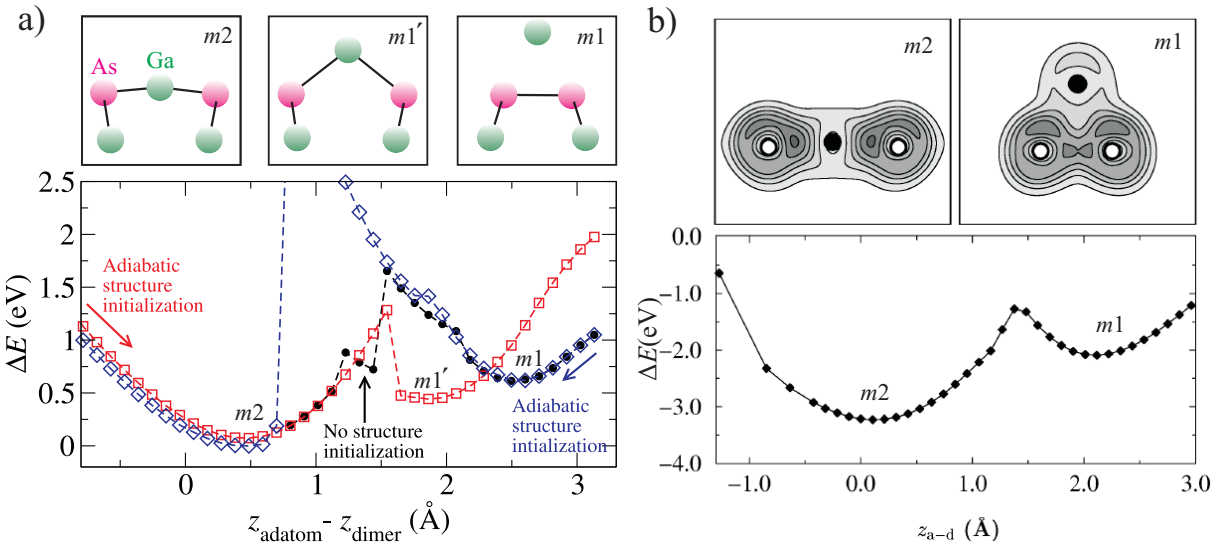


Figure 3.4: (a) The energy of a Ga adatom as a function of its height above the center of an As dimer on the $\beta 2(2\times 4)$ GaAs(001) surface using the standard DFTB method. Three curves were obtained as a consequence of different structural initialization before relaxation (see text). The adatom-dimer configuration at each of the minima is schematically shown in the upper panel. (b) The corresponding previous *ab-initio* results from Kley et al. [111] shown for comparison.

position (inside the surface). The obtained results are shown in Fig. 3.4(a) by the open blue diamonds and the open red squares, respectively.

From the results in Fig. 3.4(a) three minima can be identified. The choice of the used structural initialization schem affected the resulting minima. By investigating the atomic configurations at these minima (in the upper panel of the figure), we find that the minima $m1$ and $m2$ are the ones that correspond to those in Fig. 3.4(b). They occur at adatom height of 2.5 Å and 0.5 Å with an energy difference of 0.56 eV, i.e., significantly lower than the difference obtained from the *ab-initio* results. For the minimum $m1'$, which occurs at ~ 1.8 Å, the system prefers the open dimer configuration over the closed dimer configuration. Note that the kink in the first curve (black dots) that appears in between the two minima (see the black arrow in Fig. 3.4(a)) also indicates the existence of the third minimum $m1'$. In contrast, in the *ab-initio* calculations neither the minimum $m1'$ nor the kink in the curve between the two minima occurs. Finally, Fig. 3.4(a) shows that the results are very sensitive to the input structure. For instance, this can be seen in the difference between the blue and the red points for adatom heights < 0.5 Å, and in the difference between the black and the blue points for the heights between 1.5–2 Å. Therefore, the use of DFTB can lead to physically wrong binding configurations.

As a final test we have investigated its performance in mapping the potential energy surface (PES)[6] of a Ga-adatom above the the $\beta 2(2\times 4)$ surface employing the standard-DFTB method. We note that the PES mapping is usually performed on a lateral grid covering the surface. At each point the adatom is placed at a certain initial height above the surface and allowed to relax only in the direction perpendicular to the surface, with allowing surface relaxations. In our test employing DFTB we have mapped the potential energy along the straight line indicated in

[6]The concept of the potential energy surface is discussed in Chapter 6.

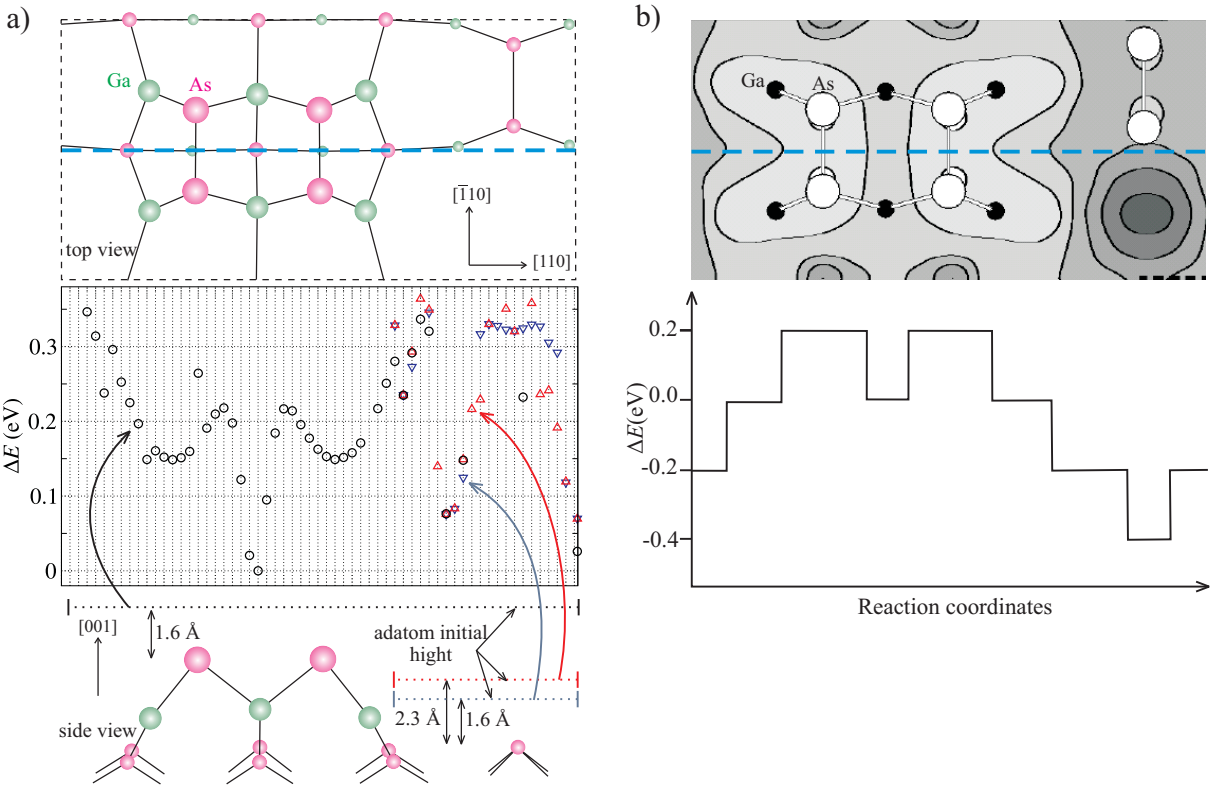


Figure 3.5: (a) One dimensional potential energy mapping of a Ga adatom above the GaAs(001) $\beta 2(2\times4)$ surface using the standard DFTB methodology. The mapping was performed along the dashed blue line indicated in the upper panel (showing a schematic top view of the surface). The middle panel shows the corresponding resulting energy values when the adatom is relaxed from three different initial heights above the surface, in the regions as indicated in the side view in the lower panel (see text). The energy zero was set at the minimum value. (b) Upper panel: the total potential energy surface for the adatom above the surface using DFT-GGA calculations, adapted from Ref. [111]. The contour lines spacing is 0.2 eV. Using the contour lines, the energy values along the dashed blue line (same as in (a)) were extracted and connected with straight lines, as shown in the lower panel. The energy zero is set at the central point between the two pairs of As dimers, identical to the position of the energy zero in (a).

the upper panel of Fig. 3.5(a). In the first trail the adatom was relaxed from 1.6 Å above the uppermost (first layer) As dimers and along the whole line, as schematically indicated in the lower panel in Fig. 3.5(a). The resulting energies are shown by the black circles in the middle panel of Fig. 3.5(a). First, we note that the calculations failed to converge properly when the vertical distance between the adatom and the underneath surface atoms becomes higher due to the shape of the $\beta 2(2\times4)$ surface which has missing first and second layer atoms. This can be seen in the region above the third layer As-dimer where most energy points are missing (unphysical or too high in energy). In order to overcome this we repeated the mapping in this region with lowering the starting height of the adatom, as indicated the lower panel in Fig. 3.5(a). In the first trial we relaxed the adatom from 2.3 Å above the third layer As-dimer. The resulting energies are shown by the red triangles in the middle panel of Fig. 3.5(a). Although the results have improved, there are still few unphysical points and the energy curve is not smooth enough

and does not reproduce the symmetry around the As-dimer correctly. By moving the adatom slightly closer to 1.6 Å above the third layer As dimer the results have further improved, as can be seen by the inverted blue triangles in the middle panel of Fig. 3.5(a). Surprisingly, this slight change has resulted in a strong effect; as can be seen in the figure most of the blue and red triangles do not overlap. The symmetry around the lower As-dimer is now preserved and the energy curve is smoother with very less unphysical points. Therefore, the choice of the initial adatom high is very significant for DFTB calculations and should be carefully checked, which can be a cumbersome task.

Let us now compare these results with the *ab-initio* PES presented in Ref. [111], shown in the upper panel of Fig. 3.5(b).[7] Using the contour lines we have extracted the energies along the same line for which we performed the mapping using DFTB (also indicated in the upper panel), and connected them with straight lines. The corresponding potential energy along the line is shown in the lower panel. By comparing the DFTB results with the ab-*initio* results we note that there is a significant discrepancy in the region above the third layer As-dimer. In this region the DFTB calculations predict energies that are significantly higher. This indicates that the DFTB will not be able to reproduce the correct diffusion barriers.

3.5 Summary

The DFTB method provides a scheme to bridge the gap between *ab-initio* and empirical tight binding approaches, resulting in a computationally very efficient method without any empirical fitting. As a promising approach to employ for pre-relaxing the expensive surface and adatom calculations we have successfully implemented the DFTB method in the S/PHI/nX simulation package. Several accuracy performance tests for physical properties of interest have been performed. These tests include bulk properties, surface energies of the GaAs(001) surface reconstructions, and energetics of a Ga adatom on the GaAs $\beta 2(2\times 4)$ surface. The tests show several critical drawbacks in using the DFTB, as summarized in the following. First, the DFTB did not reproduce correctly the *ab-initio* predicted stabilities of various reconstructions of the GaAs(001) surface. Second, the surface energies calculated with the DFTB were found to be sensitive to the used parameter files, which indicates a transferability problem. Moreover, the DFTB has led to incorrect Ga-adatom As-dimer binding configurations and energies. Furthermore, the DFTB has failed to describe the *ab-initio* potential energy of a Ga adatom accurately. Finally, the DFTB adatom calculations have shown strong sensitivity to the initial position/configuration of the adatom on the surface, which needs to be carefully checked.

These issues show that the DFTB is unfortunately not a suitable approach to employ for our calculations. Hence it is restricted to this chapter and was not used in the later stages of this work. Nevertheless, it is necessary to mention here that the S/PHI/nX DFTB implementation has found various useful applications, specially for biological systems.

[7] This PES was obtained when starting from the positions of the clean surface (intact dimer bonds) for the substrate (similar to the case in our DFTB calculations) and an adatom height of 3 Å above the surface.

Chapter 4

Computational Details and Material Properties

4.1 Introduction

The calculations within this work were performed employing the density functional theory using the generalized gradient approximation for the exchange-correlation functional (GGA(PBE)) [67]. For the sake of computational efficiency and based on the available computational resources we have used two *ab-initio* computer codes. These two codes differ in the form of the basis set used for solving the Kohn-Sham equations. For studying the thermodynamics of the clean and N substituted surfaces (presented in Chapter 5) we have used the S/PHI/nX code package [87], which employs pseudopotential plane-waves (PS-PW). The S/PHI/nX code delivers very good performance for relatively small systems with fully occupied bands, such as the semiconducting surfaces treated here. However, it becomes often required to simulate larger surface cells to reduce interactions from the periodic images, such as for the surfaces that include adatoms which additionally have partially occupied bands. Hence, to be accomplished in a reasonable time, the calculations for such systems require to be performed on parallel processors. Unfortunately this option has not been available in the S/PHI/nX code. Therefore, to study adatom kinetics (Chapter 6) and large surface cells (Chapter 7) we have used the VASP *ab-initio* code [88] which employs projector augmented waves (PAW).

In the following we introduce the common general computational details that were used throughout this work. We then present the calculated structural and cohesive properties of the material systems of interest compared to experimental results, which is essential to check the accuracy of our calculations. To guarantee that the two codes are consistent the presented structural and cohesive properties were calculated employing both codes. Further details will be provided at the corresponding positions in this work as necessary.

4.2 General computational details

For the PS-PW calculations, which were performed using S/PHI/nX *ab-initio* code [87], we used fully-separable [79] norm-conserving pseudopotentials of the Troullier-Martins type [75]. The 3*d*-electrons of Ga and the 4*d*-electrons of In were treated as core electrons employing the non-linear core correction (NLCC). The GGA(PBE) approximation were used for the exchange-correlation functional.

An important parameter that needs to be carefully determined is the cutoff energy for the plane-waves. Since nitrogen is the most demanding atomic species in our systems, we fix the cutoff energy in all calculations to that required for a system containing N to give a converged quantity interest within a certain accuracy. Let us first consider the convergence of the total energy of GaN in the bulk phase. The energy of GaN bulk were calculated at different cell volumes and for various energy cutoffs: 30, 40, 50, and 60 Ry. The volume versus energy data were fitted according to Murnaghan's equation of state [112] and the results are shown in Fig. 4.1(a). The results indicate that 60 Ry cutoff leads to converged total energies to a very good accuracy. Reducing the cutoff to 50 Ry leads to energy differences of ~ 90 meV per atom, while reducing it further to 40 Ry leads to considerable energy differences of ~ 480 meV per atom. Unfortunately, using a cutoff of 60 Ry makes the computations expensive for larger systems (slab calculations). Therefore we check the convergence of a more relevant quantity for our work which is the formation energy of substitutional incorporation of N at surfaces. Since this involves calculating energy differences, a better convergence with the cutoff energy is expected. For this purpose we have chosen the $\beta 2(2\times 4)$ reconstruction of the GaAs(001) surface.[1] The formation energy of N substitution at the topmost surface layer under rich As and N conditions[2] has been calculated using cutoffs of 40 and 60 Ry's. The results are shown in Fig. 4.1(b). The figure shows that reducing the cutoff energy from 60 Ry to 40 Ry results in a small difference in the formation energy of ~ 40 meV per the 2×4 surface cell. In addition, we note that the calculated GaN equilibrium lattice constants in Fig. 4.1(a) are almost identical for cutoffs > 40 Ry. Therefore, we have used a cutoff energy of 40 Ry (equivalent to 544 eV)[3] in all of our calculations, which allows to significantly reduce the computational efforts especially for calculating surface systems.

For the PAW calculations, which were performed using the VASP *ab-initio* code [88], the standard set of the PAW potentials that are distributed with the code [93, 122] have been used, within the GGA approximation for exchange and correlation. For the cutoff energy we have used the maximum recommended value (provided in the potential files) to accurately describe systems containing nitrogen, which is 29 Ry (or 400 eV). Using this value is feasible for our systems and hence there is no need to be further optimized as for the case of the PS-PW calculations.

For brillouin zone (BZ) integrations we have used the special **k**-points sampling scheme of Monkhorst-Pack (MP) [85]. For the calculations in the two codes we have employed identical **k**-points meshes. For the calculating the properties of the bulk systems presented here we have

[1] Surface reconstructions are introduced in Chapter 5.

[2] Substitutional incorporation of N atoms and the associated formation energies are discussed in Chapter 5. The substitutional site considered here is the $a4$ site indicated in Fig 5.5(a). The formation energies are calculated according to Eq. (5.12) at As and N rich conditions ($\mu_{\rm As} = \mu_{\rm As^{bulk}}$ and $\mu_{\rm N} = \mu_{\rm N^{N_2}}$).

[3] 1 Rydberg $= \frac{1}{2}$ Hartree ≈ 13.6057 eV.

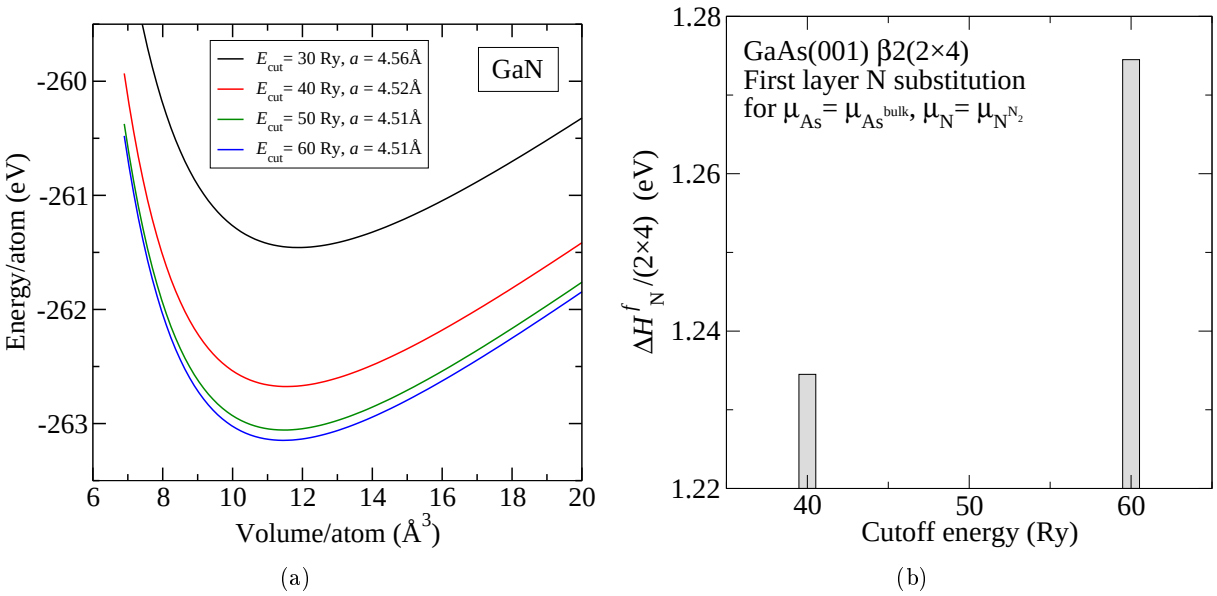


Figure 4.1: (a) Murnaghan's equation of state fits for GaN bulk (zinc blende) at various PW's cutoff energies. The hence obtained equilibrium lattice constants are shown in the legend. (b) The formation energy of N substitution in topmost surface layer of the $\beta 2(2\times4)$ reconstruction of the GaAs(001) surface for two cutoff energies.

used: $6\times6\times6$ **k**-points mesh for bulk Ga; $12\times12\times12$ **k**-points mesh for bulk As and In; and $4\times4\times4$ **k**-points mesh for GaAs, InAs, GaN, and InN. The generating **k**-point was always placed at $(\frac{1}{2},\frac{1}{2},\frac{1}{2})$. A Fermi broadening corresponding to an electronic temperature of 0.1 eV was used for bulk Ga, As, and In. The cell shape of the ZB structures were optimized using Murnaghan's equation of state [112]. For bulk Ga, In, and As the cell volume and shape were optimized using automatic cell optimization.[4]

Finally, for calculating the total energy of molecules (the N_2 molecule) and of single atoms within periodic boundary conditions the supercell approach was employed (see e.g. Sec. 2.6.1) to suppress fictitious interactions between the atom/molecule and its periodic images. We have used cubic supercells with side length of 10 Å and a single **k**-point. In addition, for calculating isolated atoms spin polarization has been included.

4.2.1 Surface calculations

For our slab calculations it is essential to find the required number of atomic layers and the thickness of the vacuum layer to be used. In Fig. 4.2(a) we show the convergence of the *total* energy (shifted) versus the number of the relaxed layers for a 10-layer thick GaAs $\beta 2(2\times4)$ slab.[5] The atoms in the other lower layers were kept fixed (unrelaxed) at their ideal bulk positions. The results show that it is sufficient to relax only 6 layers; relaxing additionally 2 layers more lowers the energy by $< 4\,\text{meV}/(1\times1)$. In a further check we have found that reducing the total

[4]Since this option has not been available in the S/PHI/nX code, cell optimizations were performed with the help of the ABINIT code [123, 124] using the same pseudopotential files, and employing the same computational details (energy cutoff, **k**-points sampling, etc).

[5]For surface reconstructions see Chapter 5, Sec. 5.4.

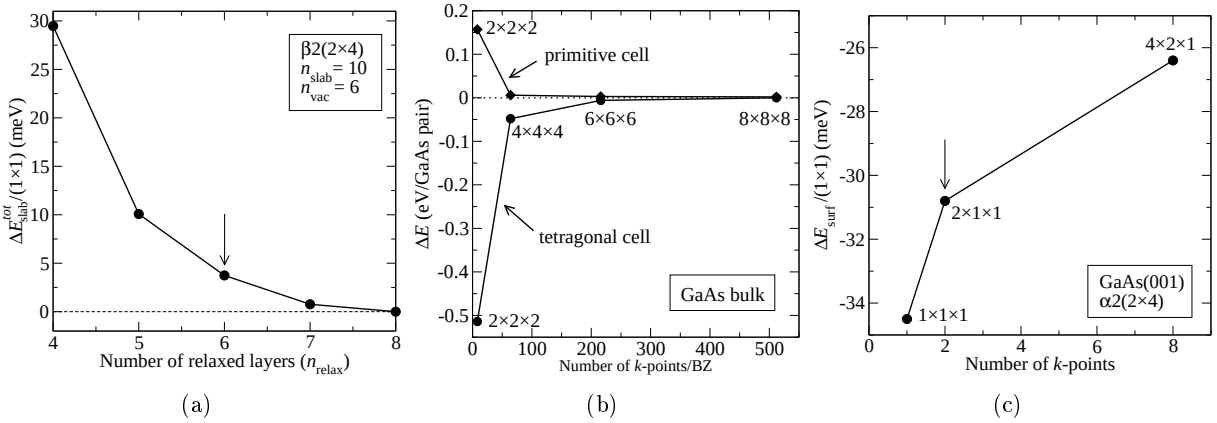
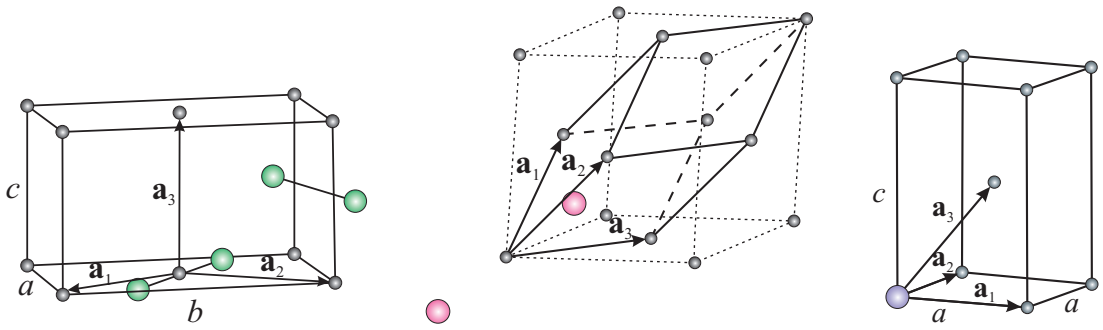


Figure 4.2: (a) Convergence check with respect to the number of relaxed atomic layers in a $\beta 2(2\times4)$ GaAs(001) surface slab. A vacuum of a thickness equivalent to 6 atomic layers (∼8.6 Å) has been used. The energy zero is set at the minimum. For (b) and (c): Convergence checks with respect to **k**-points sampling for (b) the total energy of GaAs bulk cells and (c) the surface energy of the $\alpha 2(2\times4)$ surface of GaAs(001), relative to the $\alpha(2\times4)$ surface. In (c) and the slab consists of 8 atomic layers and an equivalent vacuum thickness, and the indicated **k**-point meshes are per 2×4 surface.

number of layers in the slab from 10 to 8 (relaxing 6 layers) produces an energy difference of $< 3\,\mathrm{meV}/(1\times1)$. For the vacuum layer, we have found that reducing the vacuum from 11.5 Å (equivalent to the thickness of 8 atomic slab layers) to 8.6 Å (equivalent to the thickness of 6 atomic slab layers) results in an energy difference of $\sim 1\,\mathrm{meV}/(1\times1)$. Therefore, we have used in all of our calculations slabs that consist of 8 atomic layers; with the uppermost 6 layers allowed to relax, and a vacuum layer of thickness 11.5Å. Additionally, the dangling bonds on the backsides of the slabs were passivated with pseudohydrogens (fractionally charged) in order to ensure local charge neutrality [86].

It is necessary after that to find the optimal sampling for BZ integrations (**k**-points mesh) for our slab calculations. Therefore, in the first step we have examined the convergence of the total energy of the primitive unit cell and the tetragonal unit cell (see Sec. 4.3.1)[6] of bulk GaAs, for several MP meshes of **k**-points: $2\times2\times2$, $4\times4\times4$, $6\times6\times6$, and $8\times8\times8$; with the generating point placed at $(\frac{1}{2},\frac{1}{2},\frac{1}{2})$. The results, shown in Fig. 4.2(b), show that the total energy converges slower for the tetragonal cell. Taking an $8\times8\times8$ **k**-points mesh for the tetragonal cell, the similar sampling for our typical 2×4 slab systems is $4\times2\times1$. However, the case is different when energy differences are compared. This can be seen in Fig. 4.2(c), where we have calculated surface energy of the $\alpha 2(2\times4)$ surface of GaAs(001) relative to the $\alpha(2\times4)$ surface,[7] for several meshes of **k**-points ($1\times1\times1$, $2\times1\times1$, $4\times2\times1$ per 2×4 surface cell; with the generating point placed at $(\frac{1}{2},\frac{1}{2},0)$). As the results show, increasing the **k**-points mesh from $2\times1\times1$ to $4\times2\times1$ results in a small energy difference of $\sim 4\,\mathrm{meV}/(1\times1)$. That is, the convergence in energy differences is significantly faster than in the total energies. Therefore, in all of our surface calculations we have employed equivalent $2\times1\times1$ **k**-points per 2×4 surface, and an equivalent

[6] A tetragonal unit cell compiles a 4-layers-thick 1×1 (001) slab cell, see Sec. 4.3.1.
[7] For surface energies see Chapter 5, Sec. 5.3. Equivalent **k**-points sampling was used.



(a) α-Ga (b) α-As (c) In

Figure 4.3: (a) The base-centered orthorhombic lattice of the α-Ga structure. (b) The rhombohedral cell of the α-As atomic structure. (c) The body-centered tetragonal cell of bulk Indium. The primitive translation vectors and the basis atoms are also shown in all of the figures. Note that in (a) the basis atoms are located in one plane. Ga atoms are in green, As in pink, and In in violet. The small gray spheres represent lattice points.

sampling for bulk calculations, i.e., $4 \times 4 \times 4$ per the tetragonal unit cell.

Finally, since the uppersides and the backsides of our slabs are not equivalent, it might be necessary to include dipole correction [125]. Nevertheless, calculations for clean and for nitrogen substituted $\beta 2(2 \times 4)$ GaAs surface showed that the dipole correction to the total energy is very small ($< 1\,\mathrm{meV}$). Therefore, the dipole correction was ignored throughout this work.

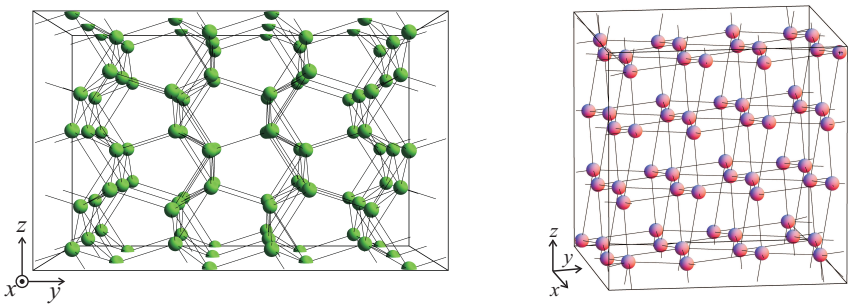
4.3 Structural and cohesive properties

4.3.1 Structural properties

Bulk Ga, As, and In

First, we present the structural properties of unary bulk systems: Ga, As, and In. Gallium crystallizes in the α-Ga (A11) structure. The underlying lattice is the base centered orthorhombic lattice. The primitive translation vectors for this lattice read: $\mathbf{a}_1 = \frac{1}{2}a\,\hat{\mathbf{x}} - \frac{1}{2}b\,\hat{\mathbf{y}}$, $\mathbf{a}_2 = \frac{1}{2}a\,\hat{\mathbf{x}} + \frac{1}{2}b\,\hat{\mathbf{y}}$, and $\mathbf{a}_3 = c\,\hat{\mathbf{z}}$. The basis of this structure consists of four atoms, located at $(\mp u, \pm u, \pm v)$ and $(-\frac{1}{2} \pm u, \frac{1}{2} \pm u, \frac{1}{2} \mp v)$ in terms of $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$. The conventional cell is shown in Fig. 4.3(a) together with the four basis atoms and the primitive translation vectors. In Fig. 4.4(a) a perspective view of a repeated α-Ga structure is shown.

Arsenic crystallizes in the α-As (A7) structure. The underlying lattice is the rhombohedral (trigonal) lattice. The primitive translation vectors can be written as: $\mathbf{a}_1 = s\,\hat{\mathbf{x}} + r\,\hat{\mathbf{y}} + r\,\hat{\mathbf{z}}$, $\mathbf{a}_2 = r\,\hat{\mathbf{x}} + s\,\hat{\mathbf{y}} + r\,\hat{\mathbf{z}}$, and $\mathbf{a}_3 = r\,\hat{\mathbf{x}} + r\,\hat{\mathbf{y}} + s\,\hat{\mathbf{z}}$. The structure has a two atomic basis, located at $(\pm u, \pm u, \pm u)$ in terms of $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$. The lattice constant in this description equals $a_{\mathrm{rh}} = \sqrt{2r^2 + s^2}$. The lattice together with the basis atoms and the primitive translation vectors are shown in Fig. 4.3(b). A more common way to describe the As structure [116] is by the following translation vectors: $\mathbf{a}'_1 = \frac{a}{\sqrt{3}}\,\hat{\mathbf{x}} + \frac{c}{3}\,\hat{\mathbf{z}}$, $\mathbf{a}'_2 = \frac{-a}{2\sqrt{3}}\,\hat{\mathbf{x}} + \frac{-a}{2}\,\hat{\mathbf{y}} + \frac{c}{3}\,\hat{\mathbf{z}}$, $\mathbf{a}'_3 = \frac{-a}{2\sqrt{3}}\,\hat{\mathbf{x}} + \frac{a}{2}\,\hat{\mathbf{y}} + \frac{c}{3}\,\hat{\mathbf{z}}$, with basis atoms at $\pm u(0, 0, c)$. These two definitions are related according to $a = 2a_{\mathrm{rh}} \sin(\alpha/2)$ and $c/a = (9[4\sin^2(\alpha/2)]^{-1} - 3)^{1/2}$ where α is the rhombohedral angle [126]. Equivalently $a = \sqrt{2}(r - s)$



(a) α-Ga repeated structure (b) α-As repeated structure

Figure 4.4: A perspective view of periodically repeated (a) α-Ga and (b) α-As structures.

	Method	a	b	c	u	v	E^c(eV)
Ga	PS-PW (GGA)	4.46	7.70	4.60	0.156	0.080	2.81
	PAW (GGA)	4.56	7.76	4.62	0.156	0.084	2.69
	Experiment[a]	4.51	7.63	4.50	0.153	0.079	2.81
As	PS-PW (GGA)	3.82	–	10.82	0.227	–	2.89
	PAW (GGA)	3.82	–	10.69	0.227	–	2.97
	Experiment[b]	3.76	–	10.44	0.228	–	2.96
In	PS-PW (GGA)	3.22	–	4.89	–	–	2.54
	PAW (GGA)	3.31	–	5.02	–	–	2.36
	Experiment[c]	3.25	–	4.94	–	–	2.52

[a]Ref. [115] (a, b, c, u, v) and Ref. [81, p50] (E_c)
[b]Ref. [116] (a, c, u) and Ref. [81, p50] (E_c)
[c]Ref. [117] (a, c) and Ref. [81, p50](E_c)

Table 4.1: Calculated and experimental structural parameters and cohesive properties for bulk Ga, As, and In. Note that the definition of the unit cell of α-Ga here is different from the one given in Ref. [115] for an 8 atoms cell and thus the values were adjusted accordingly. All dimensions are given in Å.

and $c = \sqrt{3}(2r+s)$, and hence $a_{\rm rh} = \frac{1}{3}\sqrt{3a^2+c^2}$. In Fig. 4.4(b) a perspective view of a repeated α-As structure is shown.

The underlying lattice of bulk Indium is the body-centered tetragonal (BCT) lattice (A6). The primitive translation vectors can be written as: $\mathbf{a}_1 = a\hat{\mathbf{x}}$, $\mathbf{a}_2 = a\hat{\mathbf{y}}$, and $\mathbf{a}_3 = \frac{1}{2}(a\hat{\mathbf{x}}+a\hat{\mathbf{y}}+c\hat{\mathbf{z}})$. The basis consists of a single atom located at $(0,0,0)$. The body-centered tetragonal cell is shown in Fig. 4.3(c) together with the primitive translation vectors and the basis In atom. Note that sometimes this structure is represented in terms of a face-centered tetragonal (FCT) lattice, a tetragonally distorted FCC lattice, such that $a_{\rm FCT} = \sqrt{2}\, a_{\rm BCT}$ and $c_{\rm FCT} = c_{\rm BCT}$.

In Table 4.1 the optimized structural parameters for the above mentioned bulk systems are summarized together with the corresponding experimental values. There is a general overestimation of the volume of the unit cell, as expected for the GGA functionals, except for bulk Indium within the PS-PW calculations where the unit cell volume is slightly underestimated by 2.8 %. The results are generally in very good agreement with experiment.

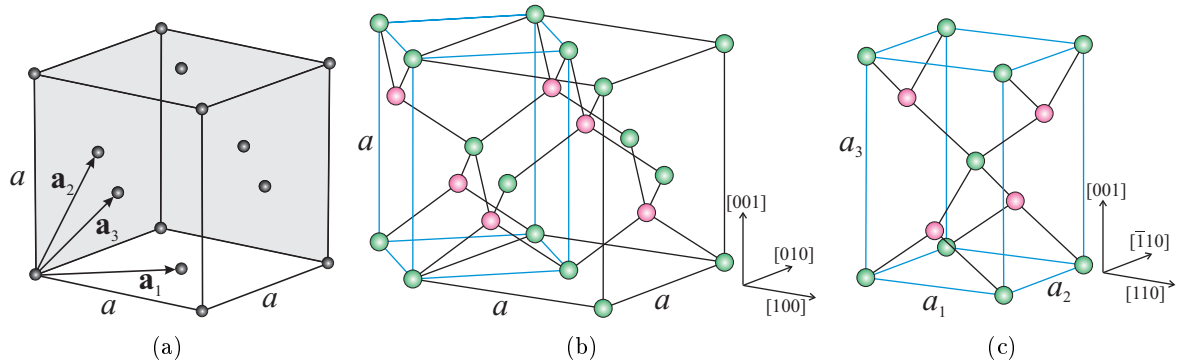


Figure 4.5: (a) The FCC conventional cell and the primitive translation vectors. (a) The zinc blende structure of GaAs/InAs showing the FCC conventional cell (black edges) together with the tetragonal cell (blue edges). (b) The tetragonal cell, with $a_1 = a_2 = \frac{a}{\sqrt{2}}$ and $a_3 = a$. The small gray spheres in (a) represent lattice points. Green spheres represent Ga/In and pink spheres represent As atoms.

Bulk GaAs, InAs, GaN, and InN

Now we present the structural properties for the binary bulk systems: GaAs, InAs, GaN, InN. GaAs and InAs crystallize in the zinc blende (ZB) structure. The underlying lattice is the face-centered cubic (FCC) shown in Fig. 4.5(a). The primitive translation vectors are given by: $\mathbf{a}_1 = \frac{1}{2}a(\hat{\mathbf{x}} + \hat{\mathbf{y}})$, $\mathbf{a}_2 = \frac{1}{2}a(\hat{\mathbf{y}} + \hat{\mathbf{z}})$, and $\mathbf{a}_2 = \frac{1}{2}a(\hat{\mathbf{x}} + \hat{\mathbf{z}})$. The primitive unit cell of the ZB structure consists of two basis atoms (one cation atom and one anion atom) at $(0, 0, 0)$ and $(0.25, 0.25, 0.25)$. The ZB structure is shown in Fig. 4.5(b) with the conventional cell shown in black edges.

The thermodynamically stable bulk GaN and InN have the wurtzite structure. However, since the energy difference between the wurtzite and the ZB phases is small (estimated to be $\lesssim 45$ meV/pair [127, 78]), we assume here the ZB structure.

For the purpose of studying the (001) surfaces of these material systems it is more convenient to consider the tetragonal unit cell, with one of its translation vectors along the [001] direction. The tetragonal cell is marked in Fig. 4.5(b) with the blue edges, and can be more clearly seen in Fig. 4.5(c). The translation vectors for this unit cell read $\mathbf{a}_1 = \frac{1}{2}a(\hat{\mathbf{x}} + \hat{\mathbf{y}})$, $\mathbf{a}_2 = \frac{1}{2}a(-\hat{\mathbf{x}} + \hat{\mathbf{y}})$, and $\mathbf{a}_3 = a\,\hat{\mathbf{z}}$. The basis of the tetragonal unit cell contains 4 atoms: 2 cation atoms at $(0, 0, 0)$ and $(0.5, 0.5, 0.5)$, and 2 anion atoms at $(0.5, 0, 0.25)$ and $(0, 0.5, 0.75)$; in terms of $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$.

In Table 4.2 the optimized lattice constants for GaAs, InAs, GaN, and InN using PS-PW and PAW are shown together with the corresponding experimental values. Both PS-PW and PAW approaches give results that are in good agreement with experiment. As expected from the GGA approximation the lattice constants are overestimated except for the PS-PW calculations of bulk GaN and InN. Here, for bulk GaN the lattice constant is found to be virtually identical to experiment, while for bulk InN it is slightly underestimated by 0.6 %. The overestimation of the lattice constant is larger in the case of the PAW calculations but the relative differences between the two approaches are small < 3.5 %.

	Method	a(Å)	ΔH^f(eV)	E^c(eV)
GaAs	PAW	5.77	−0.71	6.36
	PS-PW	5.73	−0.75	6.45
	Experiment	5.65[a]	−0.74[b]	6.52[a]
InAs	PAW	6.22	−0.47	5.80
	PS-PW	6.12	−0.63	6.09
	Experiment	6.03[a]	−0.61[b]	6.20[a]
GaN	PAW	4.60	−0.95	8.83
	PS-PW	4.52	−1.09	9.00
	PS-PW[c]	4.51	−1.11	9.21
	Experiment	4.52[d]	−1.15[b]	8.96[a]
InN	PAW	5.11	0.23	7.31
	PS-PW	4.95	−0.28	7.95
	PS-PW[c]	4.99	−0.31	7.53
	Experiment	4.98[d]	−0.30[e]	7.72[a]
		d(Å)		E^b(eV)
N_2	PAW	1.12		10.38
	PS-PW	1.10		10.20
	PS-PW[c]	1.095		10.69
	Experiment	1.10[b]		9.80[b]

[a] Ref. [106]
[b] Ref. [117]
[c] Ref. [78]
[d] Ref. [128]
[e] Ref. [129]

Table 4.2: The calculated structural and cohesive properties for bulk GaAs, InAs, GaN, and InN; and for the N_2 molecules, using both the PS-PW and the PAW approaches within the GGA approximation. The results are compared to previous theoretical and experimental data.

N_2 molecules

Finally, for the N_2 molecule, the optimized bond length for the N_2 molecules using both PW and PAW approaches are shown in Table 4.2. The PAW calculations show a small overestimation of the bond length of 1.8 %. However, both PS-PW and PAW are in very good agreement with experiment.

4.3.2 Cohesive properties

Additionally we have calculated the cohesive energies and heat of formations for the different material systems. The cohesive energy is the energy required to separate atoms in the crystal to their free form. That is

$$E^c_{\text{bulk}} = -E^{tot}_{\text{bulk}} + \sum_{\alpha} E^{tot}_{\alpha^{\text{atom}}}, \tag{4.1}$$

where $E^{tot}_{\rm bulk}$ is the total energy of the bulk and $E^{tot}_{\alpha^{\rm atom}}$ is the total energy of the atomic species α (in its atomic form) which include spin polarization. For instance, for a binary bulk material AB the cohesive energy can be written as: $E^{c}_{\rm AB^{bulk}} = -E^{tot}_{\rm AB^{bulk}} + E^{tot}_{\rm A^{atom}} + E^{tot}_{\rm B^{atom}}$. For molecular systems this is usually expressed by the binding energy. Hence, the binding energy of N_2 molecules is written as:

$$E^{b}_{\rm N_2} = -E^{tot}_{\rm N_2} + 2\,E^{tot}_{\rm N^{atom}}. \tag{4.2}$$

The heat of formation of a compound is defined as the heat absorbed or released when the compound is decomposed into its constituents in their standard state. For example, the heat of formation of bulk GaAs is

$$\Delta H^{f}_{\rm GaAs} = E^{tot}_{\rm GaAs^{bulk}} - E^{tot}_{\rm Ga^{bulk}} - E^{tot}_{\rm As^{bulk}}, \tag{4.3}$$

and similarly for InAs. For GaN the heat of formation is

$$\Delta H^{f}_{\rm GaN} = E^{tot}_{\rm GaN^{bulk}} - E^{tot}_{\rm Ga^{bulk}} - \frac{1}{2}E^{tot}_{\rm N_2}, \tag{4.4}$$

and similarly for InN. Note that if $\Delta H^{f} > 0$ for a compound material then it is thermodynamically unstable.

The calculated cohesive energies and heats of formation for the above systems are shown in Tables 4.1 and 4.2. The cohesive properties show generally good agreement with respect to experimental data. One exception is in the case of bulk InN within the PAW calculations where it is thermodynamically unstable. However, this was also found earlier within the GGA approximation [78] and it can be understood to be a result of the high overestimation of the binding energy of the N_2 molecules (with PAW it is over-bound by 0.58 eV). Overall, the PW calculations give results that are slightly in better agreement to experiment.

Chapter 5

Thermodynamics of N at GaAs and InAs (001) Surfaces

5.1 Introduction

The growth of diluted nitrides for optoelectronic applications has been challenged by the very low thermodynamic equilibrium solubility of N in bulk GaAs and InAs. The low N solubility is a direct consequence of the large strain energies resulting from the large lattice mismatch between GaAs and GaN and between InAs and InN of about 20 % and 17.4 %, respectively [4]. For instance, the equilibrium solubility of N in bulk GaAs was first estimated to be $\sim 10^{-7}\%$ at $T = 1000\,\mathrm{K}$ [23]; very low for practical applications. Nevertheless, employing MBE, the growth of $GaAs_{1-x}N_x$ alloys containing *few percents* of N is nowadays routinely realized.

The concept of *enhanced solubility* of impurities *near* crystal *surfaces*, first demonstrated by Tersoff [49], allows for a basic explanation of the achievable high experimental N solubilities. The first reason behind the enhanced solubility at surfaces is the more efficient atomic relaxations than in the bulk. This allows to partially relieve the strain caused by the impurity, thus lowering its energy and hence enhancing its solubility. The second reason is the existence of *reconstructions* at semiconductor surfaces, i.e., rearrangements of atoms to lower the surface energy. A surface reconstruction causes an intrinsic compressive/tensile strain in the surface/subsurface layers, which allows for a further considerable relief of the strain around the impurity. By employing the appropriate kinetic regime for the surface/subsurface incorporation, the high solubility at the surface can be frozen-in during crystal growth, allowing to overcome the limited solubility in the bulk. This concept and the effect of the various incorporation mechanisms have been discussed more thoroughly in the introduction chapter.

Surface enhanced solubility has so far attracted little attention, and very few theoretical studies attempted to explain the achieved experimental solubilities [50, 51]. For example, it has been proposed that the solubility of N can be dramatically enhanced by five orders of magnitude in the *subsurface* region of the GaAs(001) surface at $T = 1000\,\mathrm{K}$, which is a result of the compressive surface strain induced by the As surface dimers in the subsurface layers [50]. These studies, although succeeding in predicting an enhanced solubility at the surfaces compared to

bulk, failed however to explain the experimentally achievable and observed N concentrations. In addition, they do not provide a comprehensive understanding regarding the effect of various surface reconstructions and growth conditions. Even more importantly, they are unable to predict the existence of kinetic mechanisms that can allow e.g. for subsurface N incorporation and freezing-in during growth.

Therefore, the aim of this chapter is to analyze and understand the thermodynamics of N in GaAs and InAs bulk, and at surface layers. Specifically, methods and strategies will be developed to determine the thermodynamic limits of N solubility and identify the optimal growth conditions that can be utilized to achieve higher N solubilities. The kinetic mechanisms will be discussed in the next chapter (Chapter 6).

The chapter is organized as follows. In the first step we briefly introduce the basics of surface reconstructions. After that, we develop the basic concepts for calculating surface energies and the formation energies for N incorporation. Being the initial stage for the incorporation of dopant atoms during the epitaxial growth, we next focus on the clean GaAs and InAs (001) surfaces. In specific, we introduce the surface reconstructions and construct their thermodynamic stability phase diagrams. After that, considering a set of stable/metastable reconstructions, we calculate the formation energies for N substitutions at the various surface layers and sites. A special analysis is given for the effect of the surface reconstruction and the reconstruction-induced surface strain on N solubility. Then, we construct the stability phase diagrams of N at GaAs and InAs (001) surface and subsurface layers and in the bulk of as function of the relevant/accessible range of chemical potentials. After that, we develop the basic methodology for calculating N solubility, and based on the phase diagrams we calculate N concentrations as functions of growth conditions (partial pressures and temperatures). Finally we discuss and summarize our results.

We note that after having addressed the kinetic mechanisms for N incorporation in the next chapter, in Chapter 6 we discuss the interplay between surface kinetics and thermodynamics and the consequences on N solubility and provide our conclusions.

5.2 Semiconductor surface reconstructions

When creating a surface by cleaving a semiconducting material at least one bond per atom is cut. The cut bond is left with less than two spin paired electrons and it is called a "dangling" bond. Because this raises the free energy of the system, the surface atoms search for the new free energy minimum. As a result the surface atoms adjust their position by *relaxation* or by *reconstruction*. In surface relaxation the atoms are displaced from their equilibrium bulk positions without changing the surface periodicity, while in surface reconstruction the atomic displacements are large enough to change (lower) the surface periodicity. In addition, the reconstructed surface may have a different chemical composition (stoichiometry) than the bulk. The semi-infinite surface crystal has no longer the three dimensional periodicity of the solid. Atoms on the exposed surface will however have regular patterns with a two dimensional periodicity. This periodicity can be defined by one of the five possible Bravais lattices in two-dimensions and described by two lattice vectors ($\mathbf{a}_1$ and $\mathbf{a}_2$). The ideal unreconstructed surface is therefore referred to as the (1×1)

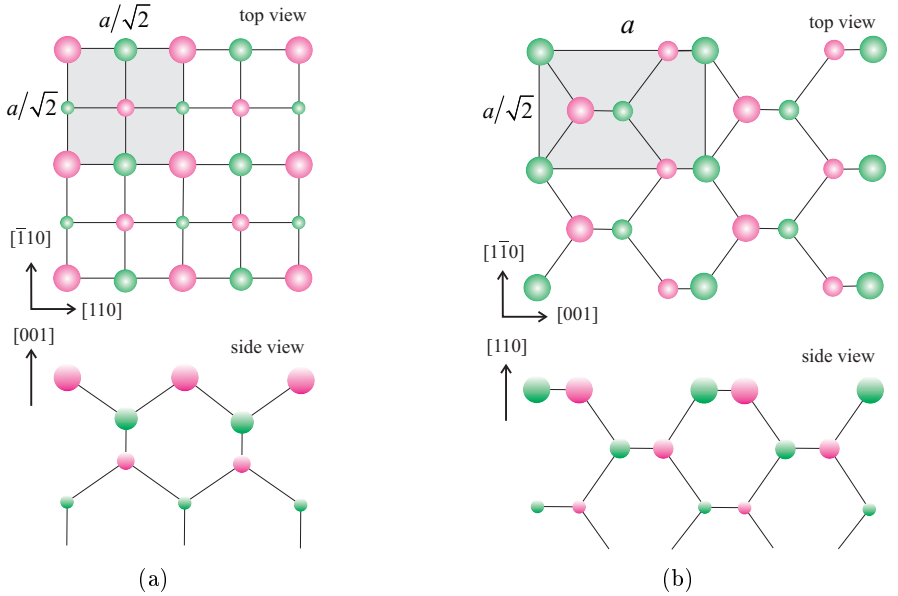


Figure 5.1: Structure of the ideal unrelaxed unreconstructed (001) (a) and (110) (b) surfaces of the zinc-blende crystal. The 1×1 primitive surface unit cells are shaded in gray. Pink (green) spheres represent anion (cation) atomic species. Spheres radii are scaled inversely with respect to their distance from the surface.

structure. Fig. 5.1 shows top and side views of schematic representation of the atomic structure of the ideal (001) and (110) surfaces of the zinc-blende crystal, showing the primitive (1×1) surface unit cells. When the surface reconstructs the symmetry is reduced and the size of the unit cell increases. The reconstructed surface is labeled with respect to the ideal termination of the corresponding surface plane. If the new unit cell of the (hkl) plane surface is spanned by $m\mathbf{a}_1$ and $n\mathbf{a}_2$ then the surface is labeled by an identifier (e.g. α, β, etc) and its periodicity $(m \times n)$.

The basic principle to explain the surface reconstruction is that atoms tend to create a semiconducting surface to lower their energy. One mechanism to achieve this is dimerization, that is the dangling bonds couple to create new bonds and thus reduce the surface energy. Another mechanism is that atoms donate their electrons to fully pair neighboring atoms dangling bonds. For compound semiconductors (such as III-V materials) the cation atoms donate their electrons to anion atoms (more electronegative) such that the dangling bonds of the cations are fully empty and the dangling bonds of the anions are fully occupied. This is known as the electron counting rule (ECR) [130]. This rule however can only provide an indication of which structures might be stable and which are not, and it cannot discriminate energetically between the different structures. More importantly, violations of the ECR have been reported. For instance, while for III-V (001) surfaces the smallest cell that fulfills the ECR is the 2×4 cell, 2×1 and 2×2 reconstructions were reported for P-rich InP surfaces [131]. Further examples for ECR violations include Sb-rich GaSb(001) surfaces [132] and Bi stabilized GaAs and InAs (001) surfaces [133]. The stabilization of such surfaces is attributed to significant stress relief due to considerable atomic relaxations [134, 133]. Nevertheless, the ECR has been shown to govern the

vast majority of III-V surface reconstructions [130]. For our case, the considered clean GaAs and InAs (001) reconstructions comply with the ECR. Moreover, we assume that the incorporation of N at surfaces does not impose new reconstructions due to the dilute concentrations studied, as experiments indicate [135, 34].

5.3 Surface energy and thermodynamics

The surface energy γ is defined as the energy needed to create the surface per unit area. The surface energy for a compound material system can be expressed by

$$\gamma A = G(T, p, \{n_i\}) - \sum_i n_i \mu_i, \tag{5.1}$$

where A is the surface area, G is the *Gibbs free energy* of the system, p is the pressure, T is the temperature, n_i is the number of atoms of species i, and μ_i is the *chemical potential* of species i:

$$\mu_i = \left(\frac{\partial G}{\partial n_i}\right)_{T,p,n_{j\neq i}}. \tag{5.2}$$

The Gibbs free energy is defined as:

$$G(T, p, \{n_i\}) = F(T, \{n_i\}) + pV, \tag{5.3}$$

where $F(T, \{n_i\})$ is the *Helmholtz free energy*. This term can be split into electronic, vibronic, and configurational parts: $F = F^{\text{el}} + F^{\text{vib}} + F^{\text{config}}$. Since the configuration entropy is zero for ordered surfaces, and since typically for free energy differences the finite temperature contributions tend to cancel, we take into account only the electronic contributions. Furthermore, for *bulk* systems and under the very low pressures in the MBE chamber, the pressure contributions to the Gibbs free energy are negligible. Therefore, we replace the Gibbs free energy by the total energy of the system:

$$G(T, p, \{n_i\}) \approx E^{tot}_{\text{system}}(\{n_i\}). \tag{5.4}$$

Due to the supercell slab approach for calculating surfaces (Sec. 2.6.1) the simulation slab has two surfaces; one on the upper side and one on the backside. In the simplest case the two surfaces are identical. An example is the (110) surface slab of the ZB structure (see Fig. 5.2(a)). In this case, and using Eqs. (5.1) and (5.4), the surface energy can be written as

$$\gamma A = \frac{1}{2}(E^{tot}_{\text{slab}} - \sum_i n_i \mu_i), \tag{5.5}$$

where E^{tot}_{slab} is the total energy from the slab calculations. The factor $\frac{1}{2}$ takes care of the fact that the slab has two surfaces. For more complex cases, when the slab cannot have two identical surfaces for geometrical reasons, calculating the *absolute* surface energy becomes less straight-

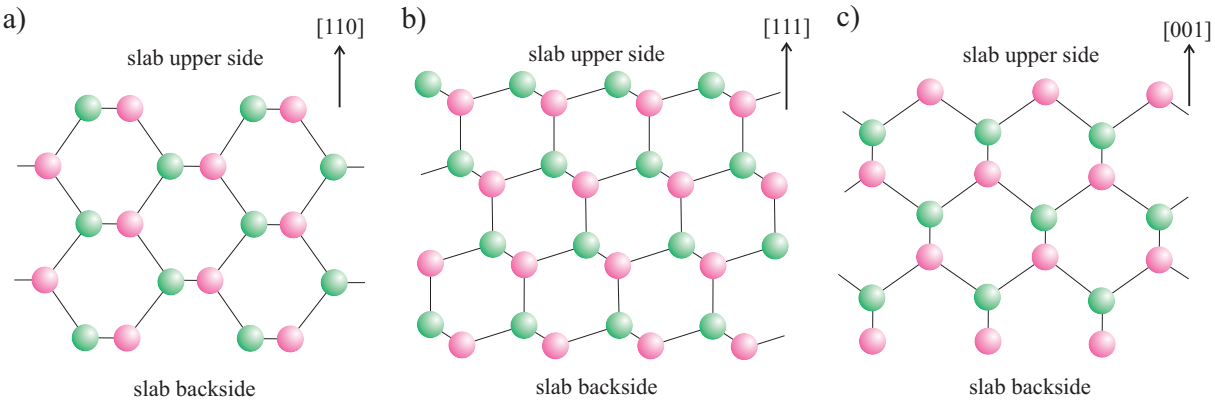


Figure 5.2: Upper- and backsides of (a) the (110), (b) the (111), and (c) the (001) surface slabs of the zinc-blende crystal.

forward. An example for such a case is the (111) surface slab in the ZB structure, where the surface on the backside of the slab ($\bar{1}\bar{1}\bar{1}$) is not equivalent to the (111) surface on the upper side (see Fig. 5.2(b)). Another example is the (001) surface slab in the ZB structure, where the surface on the backside is similar but it is rotated by 90 degrees with respect to the (001) surface. Only very few methods have been developed to calculate absolute surface energies for such cases. These methods include a local density approach [136, 137] and a wedge-shaped geometry approach [138]. The absolute surface energy is most useful to compare the energetics of surfaces of different orientations. However, for many applications the *relative* surface energy is sufficient to determine the quantity of interest. In that case, the surface energies are calculated relative to a certain reference surface. The relative surface energy ($\Delta E_{\rm surf}$) is therefore defined as

$$\begin{aligned} \Delta E_{\rm surf}\, A &= (E^{tot}_{\rm slab} - E^{tot}_{\rm slab^{ref}}) - \sum_i (n_i^{\rm slab} - n_i^{\rm slab^{ref}})\mu_i \\ &= \Delta E^{tot}_{\rm slab} - \sum_i \Delta n_i \mu_i, \end{aligned} \tag{5.6}$$

where $E^{tot}_{\rm slab^{ref}}$ is the calculated total energy of the reference slab, and $n_i^{\rm slab^{ref}}$ is the number of atoms of species i in the reference slab.

We note that the chemical potentials are not independent variables, but have to follow certain boundary conditions. For instance, based on Eq. (5.6), we can calculate the relative energies for the *clean* GaAs(001) surface reconstructions according to

$$\Delta E_{\rm surf} = \frac{1}{A}(\Delta E^{tot}_{\rm slab} - \Delta n_{\rm Ga}\mu_{\rm Ga} - \Delta n_{\rm As}\mu_{\rm As}). \tag{5.7}$$

For the bulk GaAs phase to be thermodynamically stable it is required to be in thermodynamic equilibrium with Ga and As atoms, i.e.,

$$\mu_{\rm GaAs^{bulk}} = \mu_{\rm Ga} + \mu_{\rm As}, \tag{5.8}$$

where $\mu_{\rm GaAs^{bulk}} = G_{\rm GaAs^{bulk}} = E^{tot}_{\rm GaAs^{bulk}}$/GaAs pair. Hence, the Ga and As chemical potentials

are *not* independent variables. Therefore, Eq. (5.7) can be rewritten as a function of only the As chemical potential:

$$\Delta E_{\rm surf} = \frac{1}{A}(\Delta E^{tot}_{\rm slab} - \Delta n_{\rm Ga}\mu_{\rm GaAs} - (\Delta n_{\rm As} - \Delta n_{\rm Ga})\mu_{\rm As}). \quad (5.9)$$

The difference $(\Delta n_{\rm As} - \Delta n_{\rm Ga})$ determines surface stoichiometry with respect to the reference surface, and hence the slop of the surface energy line. Usually the $\alpha(2 \times 4)$ reconstruction (see Fig. 5.3) is chosen to be stoichiometric, which is our choice here as the reference surface. In order to avoid the formation of other parasitic phase (e.g. As bulk or As droplets on the surface), another restriction on the As chemical potentials limit should be applied. Namely $\mu_{\rm As} \leq \mu_{\rm As^{bulk}}$. Therefore the allowed range of the As chemical potential becomes:

$$\mu_{\rm As^{bulk}} + \Delta H^f_{\rm GaAs} \leq \mu_{\rm As} \leq \mu_{\rm As^{bulk}}, \quad (5.10)$$

where $\Delta H^f_{\rm GaAs} = \mu_{\rm GaAs^{bulk}} - \mu_{\rm Ga^{bulk}} - \mu_{\rm As^{bulk}}$ is the formation energy of GaAs bulk (as in Eq. (4.3)), with $\mu_{\rm Ga^{bulk}} = E^{tot}_{\rm Ga^{bulk}}/{\rm atom}$ and $\mu_{\rm As^{bulk}} = E^{tot}_{\rm As^{bulk}}/{\rm atom}$. The right hand side of Eq. (5.13) represents As-rich (Ga-poor) conditions, while the left hand side of Eq. (5.13) represents As-poor (Ga-rich) conditions. Finally, for the clean InAs surfaces the same formalism applies, with replacing Ga by In in the above equations.

5.3.1 Thermodynamics and solubility of N

In this chapter we are mainly interested in calculating the *solubility* of N at surfaces and in bulks. In order to determine the solubility of an impurity in a given structure the knowledge of the energy required to incorporate/substitute it in that structure, i.e., its *formation energy*, is necessary in the first place. Therefore, the formation energy required to substitute an As atom by a N atom at a certain *surface* is the difference between the surface energies of the N incorporated surface and a clean reference surface. For N at GaAs(001) surface, and based on the above formalism (Eq. (5.6)), the formation energy can be written as:

$$\Delta H^f_{\rm N} = \frac{1}{A}(\Delta E^{tot} - \Delta n_{\rm Ga}\mu_{\rm Ga} - \Delta n_{\rm As}\mu_{\rm As} - \Delta n_{\rm N}\mu_{\rm N}), \quad (5.11)$$

where ΔE^{tot} is the total energy difference between the system slab and the reference slab, Δn_α is the difference in the number of atoms of species α between the system and the reference, and μ_α is the chemical potential of species α. Note that unless otherwise specified we consider the 2×4 area of the (001) surface as the unit area, and hence we drop A hereafter. By applying the condition in Eq. (5.8) in order for GaAs bulk to be thermodynamically stable, Eq. (5.11) can be reduced to a function of two variables ($\mu_{\rm As}$ and $\mu_{\rm N}$):

$$\Delta H^f_{\rm N} = \Delta E^{tot} - \Delta n_{\rm Ga}\mu_{\rm GaAs} - (\Delta n_{\rm As} - \Delta n_{\rm Ga})\mu_{\rm As} - \Delta n_{\rm N}\mu_{\rm N}. \quad (5.12)$$

In our case we consider only single N substitutions, i.e., $\Delta n_{\rm N} = 1$.

The thermodynamically allowed range of the As chemical potential is (similar to the case in

Eq. (5.10)):

$$\mu_{\mathrm{As}^{\mathrm{bulk}}} + \Delta H^{f}_{\mathrm{GaAs}} \leq \quad \mu_{\mathrm{As}} \quad \leq \mu_{\mathrm{As}^{\mathrm{bulk}}}. \tag{5.13}$$

For the N chemical potential we initially assume that the formation of N_2 molecules at the surface is the limiting conditions for its upper thermodynamic limit. Therefore, the values of the N chemical potential should follow the condition

$$\mu_{\mathrm{N}} \quad \leq \quad \mu_{\mathrm{N}^{\mathrm{N}_2}}. \tag{5.14}$$

The right hand sides of Eq. (5.13) and Eq. (5.14) represent As-rich (Ga-poor) and N-rich conditions, respectively, while the left hand side of Eq. (5.13) represents As-poor (Ga-rich) conditions. It is important to note here that while for solids the temperature and pressure effects on the Gibbs free energy are negligible, for gases they can be significant. For instance, in our case, at a typical temperature for the growth of diluted nitrides of 500 C° (which is assumed throughout this work), and assuming a typical pressure in MBE of $\sim 10^{-5}$ atm, the entropy correction to the total energy of the N_2 molecule is as large as 2.528 eV per molecule [139]. This energy correction includes translational, rotational, and vibrational entropy contributions to the N_2 total energy [140, 141]. The entropy corrections have thus the effect of *lowering* the actual chemical potential of N, and hence the total energy is an *upper limit* of the chemical potential. Nonetheless, we note that for the N_2 gas in MBE growth of dilute nitrides the actual upper limit of the chemical potential is affected by the growth procedure and factors like the *nitrogen source*. Further discussion on this issue will follow in Chapter 7, Sec. 7.3.1.

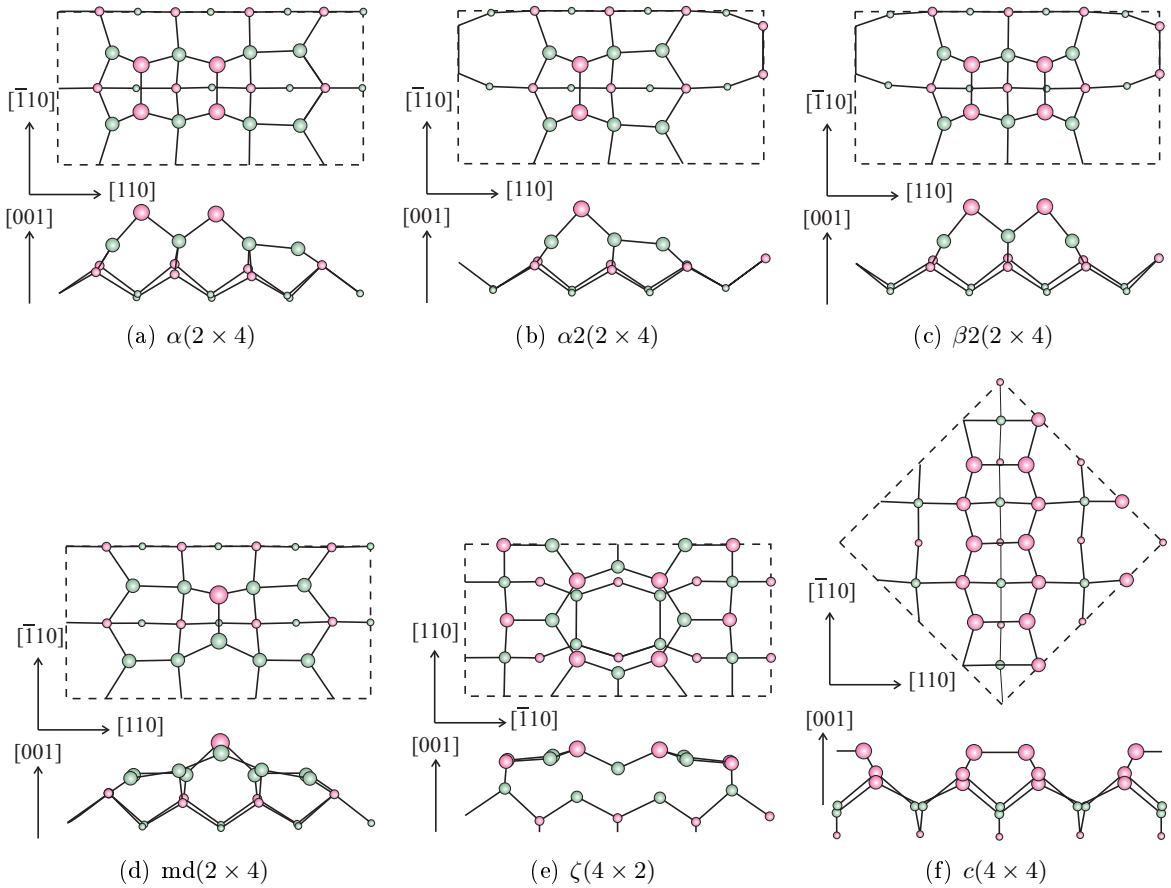
The same formalism applies for N incorporation at InAs surfaces, with replacing Ga by In in the above equations. Finally, to calculate the formation energy of substituting an As atom by a N atom in bulk GaAs and InAs the same formalism is applied, with using the *host* bulk (without N) as the reference system. The formation energy thus reads

$$\Delta H^{f}_{\mathrm{N}^{\mathrm{bulk}}} \quad = \quad \Delta E^{tot}_{\mathrm{N}^{\mathrm{bulk}}} + \mu_{\mathrm{As}} - \mu_{\mathrm{N}}. \tag{5.15}$$

5.4 The clean GaAs and InAs (001) surfaces

The (001) surface of the III-V semiconductors is of specific technological importance for fabricating and developing optoelectronic devices. Over the last years, extensive theoretical and experimental work has been employed to study the atomic and electronic properties of GaAs and InAs (001) surfaces. These surfaces show a large variety of constructions depending on growth conditions such as substrate temperature and stoichiometry. For the epitaxial growth of dilute nitride systems, including $GaAs_{1-x}N_x$, $InAs_{1-x}N_x$ and $In_xGa_{1-x}As_{1-y}N_y$, understanding the structures and the properties of the clean surfaces is important as a starting point to understand their growth, which assists to control and improve their physical properties. It is therefore essential to first examine the atomic structure of the clean GaAs and InAs (001) surfaces.

In this section we therefore introduce the various reconstructions of these surfaces, calculate



(a) $\alpha(2\times4)$ (b) $\alpha2(2\times4)$ (c) $\beta2(2\times4)$

(d) $\mathrm{md}(2\times4)$ (e) $\zeta(4\times2)$ (f) $c(4\times4)$

Figure 5.3: Schematic representation of few reconstructions of the GaAs and InAs (001) surfaces: (a) $\alpha(2\times4)$, (a) $\alpha2(2\times4)$, (b) $\beta2(2\times4)$, (c) $\mathrm{md}(2\times4)$, (d) $\zeta(4\times2)$, (e) $c(4\times4)$. The upper (lower) part of each subfigure represents top (side) view. Green spheres represent Ga(In) atoms and pink ones represent As atoms. Spheres radii are scaled inversely with respect to their distance from the surface (i.e., larger spheres are closer to the surface).

their energies, and construct their stability phase diagrams as functions of the growth conditions. This allows us to identify the most relevant surface reconstructions for N incorporation, and allows us to guarantee the accuracy of our calculations by comparing to previously reported results.

5.4.1 The GaAs(001) surface

As-moderate to As-rich reconstructions

For the GaAs(001) surface, the (2×4) surface reconstructions were extensively studied for their importance in the MBE growth for optoelectronic devices applications. Using reflection high-energy electron-diffraction (RHEED) Farrell and Palmstrom [142] identified three 2×4 phases (called α, β, and γ) under the different growth conditions.

The α phase occurs at the highest substrate temperatures ($T > 550\,\mathrm{C}^\circ$) and less As-rich conditions. The $\alpha(2\times4)$ structure was suggested to be consisting of two As-dimers in the top layer and Ga-Ga bonds in the second layer (5.3(a)). Schmidt et al. [143] suggested an

energetically more favorable structure of the same stoichiometry, called the $\alpha 2(2 \times 4)$, which is shown in 5.3(b). Experimentally, the atomic structure of the α phase has been controversial. For instance, while Laukkanen et al. [144] observed the $\alpha 2(2 \times 4)$ structure, it is argued by other groups that this structure can only be stabilized only through appropriate kinetic pathways (Ohtake et al. [145], Ohtake [146, 147]). Their work supports that the α phase consists mainly of $\beta 2(2 \times 4)$ with fractions of the $\alpha 2(2 \times 4)$ structure $(20\% - 30\%)$, in agreement with the work of Takahasi et al. [148] in which they suggested that the $\alpha 2(2 \times 4)$ structure can form and coexist with the $\beta 2(2 \times 4)$ when the surface is under conditions close to the $(2 \times 4) \rightarrow (4 \times 2)$ transition.

The atomic structure of the β phase was first theoretically discussed by Chadi [149]. He proposed the atomic structures of a $\beta(2 \times 4)$ and the $\beta 2(2 \times 4)$ structures. The structural model of the $\beta 2(2 \times 4)$ surface reconstruction consists of two As-dimers in the topmost layer and one As-dimer in the third atomic layer, as shown in Fig. 5.3(c). While the $\beta(2 \times 4)$ consists of three As-dimers in the topmost surface layer. Theoretical calculations by Northrup and Froyen [121] showed that $\beta 2$ structure is the energetically favorable one because of the electrostatic energy. The $\beta 2(2\times4)$ reconstruction was confirmed using highly resolved STM and first-principles calculations by LaBella et al. [120].

The γ phase forms under more As-rich conditions and at lower temperatures ($T < 510\,\mathrm{C}^\circ$). It was later found that this phase is a mixture of the $\beta 2(2 \times 4)$ and the $c(4 \times 4)$ structures [150].

Under the most As-rich MBE conditions, the $c(4 \times 4)$ is usually observed. The most common structure model of this surface reconstruction consists of three As dimers per (4×4) cell, as shown in Fig. 5.3(f). This atomic structure is supported by STM observations [151] and found to be stable by first principles calculations [152, 121, 137]. Recently, a new structure model which consists of three As-Ga heterodimers instead of the three As dimers has been proposed [145]. Furthermore, recent first principles calculations showed that the $c(4\times4)$ structure features diverse combinations of As-As dimers and Ga-As heterodimers [153].

Ga-rich reconstructions

Compared to the As-rich reconstructions, the Ga-rich (As-poor) GaAs(001) surface has been less extensively studied. A variety of reconstructions have been reported, however their atomic structure is still not completely understood. A possible reason for this difficulty is the coexistence of more than one phase. The reported reconstructions include (4×2), (6×6), and (4×6) cells. The ζ structure of the (4×2) cell phase was first proposed by Lee et al. [154] (shown in Fig. 5.3(e)). This structure was also experimentally supported [155]. The (6×6) reconstruction occurs at lower temperatures, and the (4×6) reconstruction is the most Ga-rich phase. A mixed-dimer reconstruction ($\mathrm{md}(2 \times 4)$) was suggested by Schmidt [119] and predicted to be stable In the Ga-rich limit using *ab-initio* calculations. Its atomic structure is shown in Fig. 5.3(d). However, this reconstructions has not been observed experimentally. Combining *ab-initio* calculations and experiment, Ohtake et al. [156] proposed an atomic structure model for the (4×6) phase, featuring surface Ga-As dimers and subsurface Ga-Ga dimers. Further detailed *ab-initio* studies of the various Ga-rich reconstructions is available in Ref. [157]. An extensive review of the progress and of the proposed atomic structure models of these reconstructions is available in Ref. [147].

5.4.2 The InAs(001) surface

Because of similar electrostatics, the GaAs(001) surface was always considered as a prototype for other III-V(001) surfaces. The InAs(001) surface has been less intensively studied than the GaAs(001), however they show many similarities. A (2×4) phase is observed over a large range of growth conditions. This phase was interpreted in terms of the $\beta 2(2 \times 4)$ reconstruction, from both experiment [158, 159, 160] and *ab-initio* calculations [158, 159, 119, 161]. The $\beta 2(2 \times 4)$ transforms into a $\alpha 2(2 \times 4)$ upon annealing as also supported by *ab-initio* [158, 159, 160]. On the other hand, going to the more As-rich limit, *ab-initio* calculations predict the $c(4 \times 4)$ reconstruction to be stable, while there exist only very little experimental support [162, 151].

For In-rich conditions, a (4×2) phase is observed. It is interpreted in terms of the $\zeta(4 \times 2)$ reconstructions, supported by *ab-initio* calculations [163, 119]. In addition, in Ref. [119] the $\mathrm{md}(2 \times 4)$ reconstruction is suggested to be stable in the more In-rich limit. However, such a transition was not observed experimentally.

5.4.3 The *ab-initio* phase diagrams

Based on the formalism derived in Sec. 5.3 and according to the computational details presented in Chapter 4 we have calculated the *ab-initio* phase diagrams for the GaAs and InAs (001) surface reconstructions. Due to the similarities in the surface structures of GaAs and InAs (001) surfaces, we have considered the same set of surface reconstructions for both materials. This set includes the $\beta 2(2 \times 4)$, the $\beta(2 \times 4)$, the $\alpha 2(2 \times 4)$, and the $\alpha(2 \times 4)$, for the rich to moderate As conditions; and the $\zeta(4 \times 2)$ and the $\mathrm{md}(2 \times 4)$ for the As-poor limit. We did not consider other larger cells reconstructions for the extreme As-poor conditions in order to keep the computational efforts feasible, specially for N incorporation. In addition, the $c(4 \times 4)$ reconstruction was not considered throughout this work because it is stable only under As-rich conditions, which are irrelevant conditions for the practical incorporation of N.

The hence calculated surface stability phase diagrams are shown in Fig. 5.4(a) for GaAs(001) and in Fig. 5.4(b) for InAs(001). First, for the case of GaAs(001) surface the results show the expected sequence of stability of the different reconstructions, in agreement with previous theoretical calculations and experiment: Under moderate As conditions the $\beta 2(2 \times 4)$ is predicted to be stable for a large window in the phase diagram, while under As-poor conditions the $\zeta(4 \times 2)$ is predicted to be stable. Furthermore, the $\alpha 2(2 \times 4)$ is predicted to be unstable in the whole thermodynamically allowed region, but it becomes lowest in energy close to the $\beta 2(2 \times 4) \rightarrow \zeta(4 \times 2)$ transition. The $\mathrm{md}(2 \times 4)$ is predicted to be unstable. For the InAs(001) surface, few differences compared to the GaAs(001) surface can be seen. First, the $\alpha 2(2 \times 4)$ becomes stable in a small window close to the $\beta 2(2 \times 4) \rightarrow \zeta(4 \times 2)$ transition. Second, the $\mathrm{md}(2 \times 4)$ becomes stable over the $\zeta(4 \times 2)$ under As-poor conditions. These calculations show that the The phase diagrams are in a good agreement with previous *ab-initio* calculations [119, 154].

Using on the calculated surface phase diagrams we can now identify the reconstructions that are relevant for N incorporation. For the GaAs(001) surface these are the $\beta 2(2 \times 4)$, the $\alpha 2(2 \times 4)$, and the $\zeta(4 \times 4)$ reconstructions, which are stable/metastable under As-moderate to As-poor

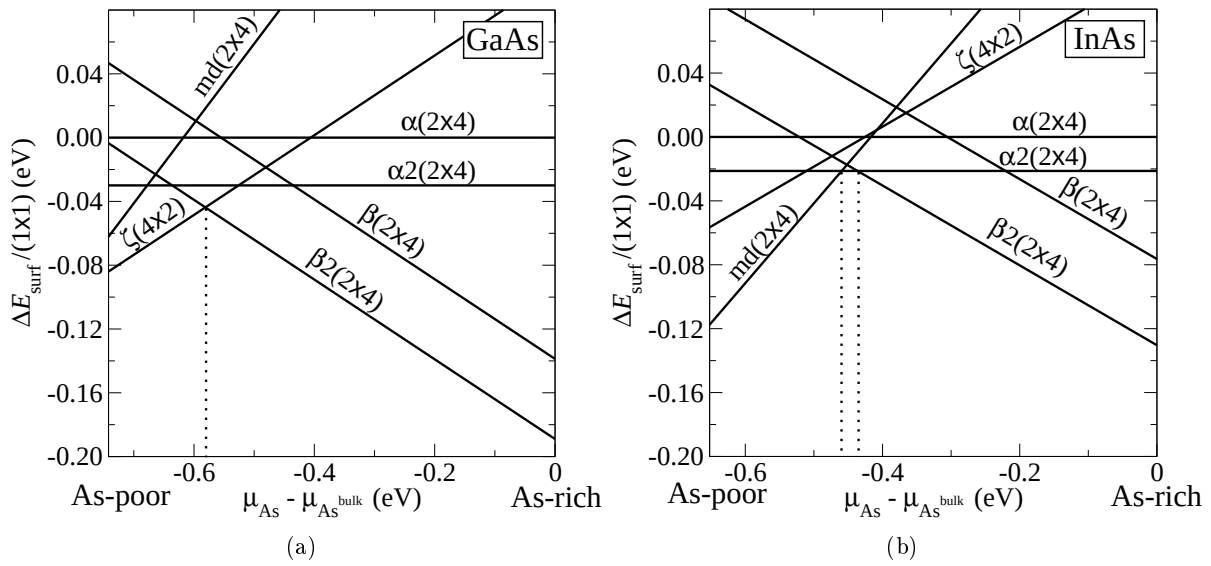


Figure 5.4: Calculated surface phase diagrams of the (001) surface of (a) GaAs and (b) InAs. Atomic structures corresponding to the various reconstructions are shown in Fig. 5.3.

conditions. For the InAs(001) we additionally consider the md(2×4) reconstruction since it is predicted to be stable in the As-poor limit. Based on these information we can proceed for calculating N thermodynamic solubilities at surfaces, as will be discussed in the following.

5.5 The effect of local strain on the solubility of N

In the introduction we have pointed out that the strain resulting from the reconstruction of the surface has a significant influence on the formation energy of N substitution and hence its solubility. Therefore, it is interesting in the first step to understand the interplay between surface reconstruction and its local structure on the solubility. This allows to determine the surface reconstructions that show maximal N solubility, and hence to employ them in practice. Therefore, in this section we analyze the formation energy of N substitution at various surface reconstructions and substitutional sites, and determine its relation to the induced local strain at these sites.

Based on the calculated phase diagram of the GaAs(001) surface shown in Fig. 5.4(a) we have chosen three most stable/metastable surface reconstructions. These are the $\beta 2(2 \times 4)$, $\alpha 2(2 \times 4)$, and $\zeta(4 \times 4)$ reconstructions.[1] A schematic representation of the atomic structure of these reconstructions is shown in Fig 5.5. After that we have replaced an As atom at the surface with a N atom, and allowed the structure to relax. All symmetry nonequivalent sites in the first and second anion layer have been considered. The surface sites are labeled as shown in Fig. 5.5 for later reference; the a-sites are located at the first anion layer (a-layer) and the c-sites are in the second anion layer (c-layer).

[1] In the dilute concentration limit, it is unlikely that N changes the surface reconstruction, as experiments indicate [135, 34]. Moreover, it is crucial to avoid N-induced surface reconstructions in the growth of GaAsN and InAsN because these act as precursors to the formation of parasitic phases such as GaN and InN, thus resulting

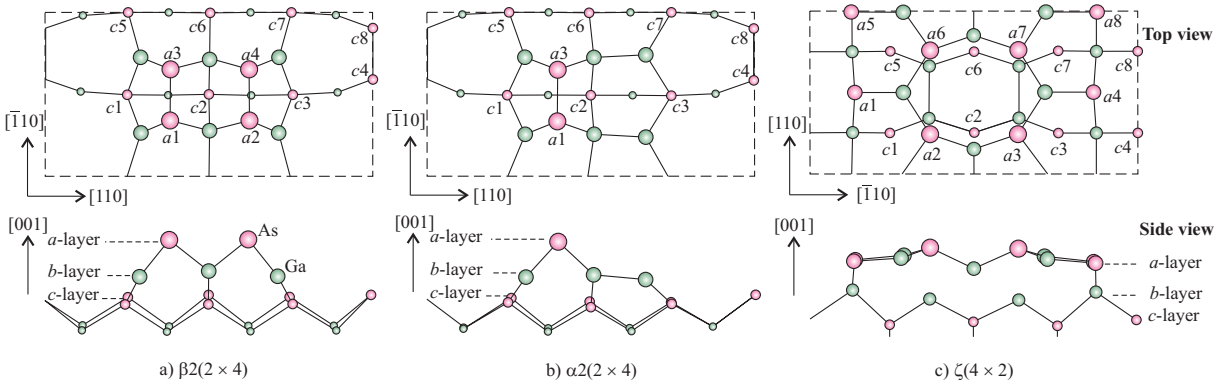


Figure 5.5: Schematic representation of the (a) $\beta 2(2 \times 4)$, (b) $\alpha 2(2 \times 4)$, and (c) $\zeta(4 \times 4)$ reconstructions of the GaAs(001) surface showing top (upper parts) and side views (lower parts). The labels give the notation for the N substitutional sites as used in the text.

A useful quantity to identify the surface enhanced solubility is the *segregation energy*, rather than the formation energy. The segregation energy is defined as the difference between N substitution formation energy in the surface and in the bulk, i.e.,

$$E_{\text{seg}} = \Delta H^f_{\text{N}^{\text{surf}}} - \Delta H^f_{\text{N}^{\text{bulk}}}. \tag{5.16}$$

Here, $\Delta H^f_{\text{N}^{\text{surf}}}$ and $\Delta H^f_{\text{N}^{\text{bulk}}}$ are the formation energies of N substitution at a surface site and in the bulk, respectively. Using Eq. (5.12), with the reference systems being the equivalent clean systems, the dependence on the chemical potentials cancels out and we get:

$$E_{\text{seg}} = \Delta E^{tot}_{\text{N}^{\text{surf}}} - \Delta E^{tot}_{\text{N}^{\text{bulk}}}. \tag{5.17}$$

According to this definition, a negative segregation energy indicates that N is more stable at that surface site than in the bulk, i.e. surface segregation will occur, while a positive segregation energy implies that surface segregation is suppressed.

In the following calculations, the formation energies of N incorporation were calculated using single (unrepeated) surface unit cells, which here have areas of 2×4 or 4×2 (001) unit cells. For the bulk calculations $2 \times 4 \times 1$ tetragonal supercells were used. Convergence checks using repeated cells (of double size, i.e., (4×4) cells) for the $\beta 2(2 \times 4)$ reconstruction showed that the error due to the supercell size is 0.01–0.23 eV (depending on the substitutional site) for surface calculations and $\sim$0.1 eV for bulk calculations. A detailed analysis showed that the large error values (> 0.05 eV) apply only for energetically highly unfavorable sites. For the relevant low energy structures the error is always smaller than 0.05 eV and will therefore not affect the calculation of N solubility. Other computational details have been discussed in Chapter 4.

in a deterioration of the material quality.

Anion layer	N site	Coordination	E_{seg}	$\overline{E_{\text{seg}}}$
First	$a1$, $a2$	3	−0.12	
	$a3$, $a4$	3	−0.13	
Second	$c1$, $c3$	4	−0.49	
	$c5$, $c7$	4	0.70	0.11
Second	$c2$	4	−0.68	
	$c6$	4	1.00	0.16
Second	$c4$	3	0.13	
	$c8$	3	0.25	

Table 5.1: Segregation energies (in eV) for all N substitutions in the first and second anion layers of the $\beta 2(2\times 4)$ reconstructed GaAs(001) surface. The last column shows the averaged segregation energies of each two neighboring N substitutional sites in the $[\bar{1}10]$ direction. The labels of N sites are indicated in Fig. 5.5(a).

The $\beta 2(2\times 4)$ reconstruction

Let us first discuss the results for the $\beta 2(2\times 4)$ reconstruction. This reconstruction is of specific importance because of its stability over a wide range of chemical potentials (e.g. see Fig. 5.4(a) and Ref. [154]). The calculated segregation energies are summarized in Table 5.1 (forth column). The results clearly show that the N segregation energy strongly depends on the specific surface site. The lowest energy configuration for N is in the second anion layer (c-layer) at the central position below two As-dimers (i.e. at position $c2$). The strong energetical preference of this site agrees with results reported in previous studies [164, 51]. A similar site for a (2×1) reconstruction was also reported to be energetically preferable [50]. In the first anion layer, the energetically preferred sites are $a3/a4$. However, compared to the most favorable $c2$ site their energy is higher by 0.55 eV. The segregation energies at the other first anion layer sites $a1/a2$ are similar because they are almost equivalent to the $a3/a4$ sites (see Fig. 5.5(a)).

An interesting observation is that the segregation energy for a substitution in the second anion layer is strongly site dependent. For example, going from the $c2$ site along the $[\bar{1}10]$ direction to the $c6$ site the segregation energy increases by almost 1.7 eV, despite the fact that the N atom is fully coordinated in both sites. This clearly indicates that local strain, caused by the surface reconstructions in the top surface layer, plays a major role. For example, the dimer contraction [150, 165, 166, 167] of an As dimer induces a compressive strain in the subsurface layer bonds which are directly below it and a tensile strain in the neighboring bonds in the $[\bar{1}10]$ direction, i.e., the $c2$ site is under compressive strain while the $c6$ site is under tensile strain (Fig. 5.5(a)). Incorporating a N atom at the $c2$ site partially releases the compressive strain, while incorporating a N at the neighboring site $c6$ further increases the tensile strain. In general, the As dimers in the first layer induce compressive strain on the atomic sites directly below them ($c1$, $c2$, and $c3$), and tensile strain on the neighboring atomic sites ($c5$, $c6$, and $c7$). In order to verify that the local compressive/tensile strain induced is the main origin for the strong variations in the segregation energies, we list in Table 5.1 (last column) the average segregation energy for each pair of sites along the $[\bar{1}10]$ axis (i.e. parallel to the dimer bonds). To

Anion layer	N site	Coordination	E_{seg}	$\overline{E_{\mathrm{seg}}}$
First	$a1$	3	0.13	
	$a3$	3	0.12	
Second	$c1$	4	−0.16	
	$c5$	4	0.68	0.26
Second	$c2$	4	−1.01	
	$c6$	4	0.05	−0.48
Second	$c3$	4	−0.01	
	$c7$	4	0.34	0.16
Second	$c4$	3	0.24	
	$c8$	3	0.30	

Table 5.2: As for Table 5.1 but for the $\alpha 2(2 \times 4)$ reconstruction. The labels of N sites are indicated in Fig. 5.5(b).

avoid chemical contributions only fully coordinated c-layer sites are considered. We find that the average segregation energies are energetically almost degenerate within 0.05 eV, implying that indeed local strain effects cause the large energy differences between the individual sites.

The $\alpha 2(2 \times 4)$ reconstruction

Now let us consider the case of the $\alpha 2(2 \times 4)$ reconstruction. This structure is similar to the $\beta 2(2 \times 4)$; the main difference is a missing As dimer at the top surface layer (see Figs. 5.5(a) and (b)), making this structure less symmetric. The resulting segregation energies are listed in Table 5.2. Also, for this surface reconstruction the N segregation energy strongly depends on the specific surface site. The lowest N segregation energy is again for the $c2$ site being even lower in energy than for the $c2$ site in the $\beta 2(2 \times 4)$ reconstruction. We note that the N segregation energy for the $c3$ site is higher than that for the $c1$ site because of the missing As dimer above $c3$. The averaged segregation energy is also shown in the last column for each pair of sites along the $[\bar{1}10]$ direction for the fully coordinated c-layer sites. The situation for the averaged segregation energies is more complicated than in the case of the $\beta 2(2 \times 4)$, and cannot be explained solely in terms of a simple elastic picture. The reason is that the surrounding chemical environment of the various sites is more different than on the $\beta 2(2 \times 4)$. Nevertheless, the effect of the local strain is still visible; the difference between the average segregation energies is significantly lower than the difference between the individual segregation energies.

The $\zeta(4 \times 2)$ reconstruction

Finally we show the results for the $\zeta(4 \times 2)$ reconstruction. This structure is geometrically rather complex and structurally very different from the former (2×4) reconstructions (see Fig. 5.5). The resulting segregation energies are summarized in Table 5.3. The averaged segregation energy is also shown in the last column for each two neighboring sites in the c-layer as discussed previously. As the results show, the elastic picture no longer holds since the bonding between the top surface

Anion layer	N site	Coordination	E_{seg}	$\overline{E_{\text{seg}}}$
First	$a1$, $a4$	3	0.01	
	$a5$, $a8$	3	−0.23	
	$a2$, $a3$, $a6$, $a7$	4(3)	−0.34	
Second	$c1$, $c3$, $c5$, $c7$	4	−0.39	−0.39
	$c2$, $c6$	4	0.24	0.24
	$c4$, $c8$	4	0.45	0.45

Table 5.3: Segregation energies (eV) of all N substitutions in the first and second anion layers of the $\zeta(4\times2)$ reconstructed GaAs(001) surface. The last column shows the averaged formation energies of each two neighboring N substitutional sites in the [110] direction. The coordination number between parentheses is the original coordination number of the As atom at that site. The labels of N sites are indicated in Fig. 5.5(c).

layer and the lower lying layers is extremely inhomogeneous (see Fig. 5.5(c), bottom).

5.6 Stability phase diagrams of N at GaAs(001) surfaces

In order to determine the N solubility for the various surface reconstructions we first construct the stability phase diagram of N at the GaAs(001) surface as function of the specific growth conditions (chemical potentials). The calculation of the phase diagrams is done according to the following procedure:

1. First, using the calculated surface phase diagram of the *clean* surface (Fig. 5.4(a)) we identify the stable reconstructions and their region of thermodynamic stability.

2. Then, following the procedure described in the previous section, we identify for each reconstruction and in each layer the energetically most preferable site for N substitution.

3. Using the *most favorable* N substitution and Eq. (5.12), we calculate for each reconstruction and in each layer the formation energy of the N substitution as a function of growth conditions, relative to the thermodynamically stable clean surface under the corresponding As chemical potential.

4. Finally we determine the thermodynamically stable regions for N substitution (defined by μ_{As}, μ_{N}), i.e. the growth window where the N substituted surfaces become more stable than the clean surfaces (i.e. $\Delta H^f_{\text{N}} \leq 0$). This is conventionally represented in a two dimensional plot with μ_{As} and μ_{N} as independent variables.

5.6.1 Results

The complete N stability phase diagrams are shown in Fig. 5.6 for N incorporation (a) in bulk, (b) in the second anion surface layer (denoted with the adjunct $\text{N}^{2\text{nd}}$), and (c) in the first anion surface layer (denoted with the adjunct $\text{N}^{1\text{st}}$). The region where the corresponding N-rich surface becomes stable is marked by a solid line. A point which may be a bit unusual is that the shown

surface diagrams include also regions outside the thermodynamically allowed region. Specifically, structures are included which are thermodynamically unstable against the decomposition into N_2 molecules. As will be later shown this is highly useful to discuss/derive surface solubility and to take into account kinetic growth aspects.

Based on the computed phase diagrams the following conclusions can be drawn:

(*i*) The energy required to substitute N in GaAs is dramatically reduced at surfaces compared to bulk.

(*ii*) The second layer is energetically more attractive for N atoms than the first.

(*iii*) In a small thermodynamic window the formation of a GaAsN surface alloy becomes exothermic (formation of the $\alpha 2(2 \times 4)$ reconstruction with N in the second anion layer).

(*iv*) N is very reactive and may change the relative stabilities of surface reconstructions for fixed Ga/As-ratio (fixed μ_{As}).

The last issue (*iv*) becomes evident when inspecting e.g. Fig. 5.6(b) where under modest As-rich conditions the $\alpha 2(2 \times 4)$ reconstruction becomes stable which is *not* a stable one in the bare GaAs(001) surface phase diagram (Fig. 5.4(a)). Another example is shown in Fig. 5.6(c) where the $\zeta(4\times2)$ reconstruction becomes stable in a region where in the absence of N only the $\beta 2(2\times4)$ reconstruction is observed.

An interesting observation from the phase diagram (Fig. 5.6) is that the incorporation of N is strongly enhanced when going from As- to Ga-rich conditions. This by itself is not unexpected since to replace an As with a N atom the former has to be lowest in its chemical potential where the energy gain is largest for Ga-rich conditions. What is more interesting is that at the surfaces this behavior is not linear like in bulk, but shows different slopes and even a non-monotonous behavior. An important conclusion we can therefore draw that that the maximal N solubility is not necessarily achieved under extreme As-poor conditions as one might naively expect from the pure bulk behavior.

5.7 Solubility of N at GaAs(001) surface

5.7.1 Calculating N solubility: Methodology

Based on the calculated phase diagrams and employing Boltzmann statistics, the concentration of N in the dilute limit can be determined as a function of the specific growth conditions (μ_{As}, μ_{N}, T). As Fig. 5.6 shows, the N substitution energy depends strongly on the surface layer as well as on the specific site. The maximal solubility which can be achieved thus depends on surface kinetics, which determines the kinetically accessible layers (see Fig. 1.2). In order to gain insight how different kinetics might affect solubility we consider the *layer resolved* N solubility, i.e. the N solubility at each individual surface layer. The equilibrium concentration of N in the

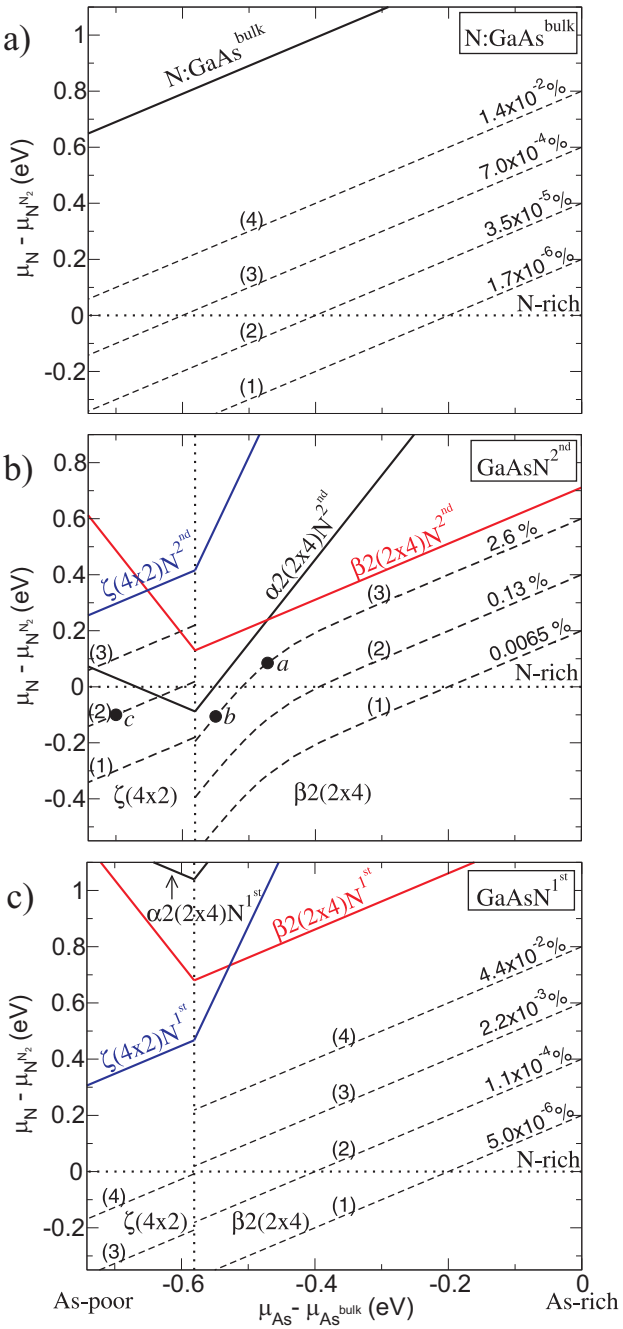


Figure 5.6: Solid lines: Stability phase diagram of N in (a) bulk GaAs, (b) second anion layer, and (c) first anion layer of the GaAs(001) surface reconstructions. Dashed lines: contour lines of constant N concentrations calculated at $T = 500$ C°. The dotted vertical line indicates the phase boundary between the $\beta 2(2 \times 4)$ and the $\zeta(4 \times 2)$ reconstructions. The three large dots in (b) marked as *a b c* indicate the growth conditions that were selected to calculate N concentration using a Metropolis Monte Carlo simulation (see Fig. 5.7)

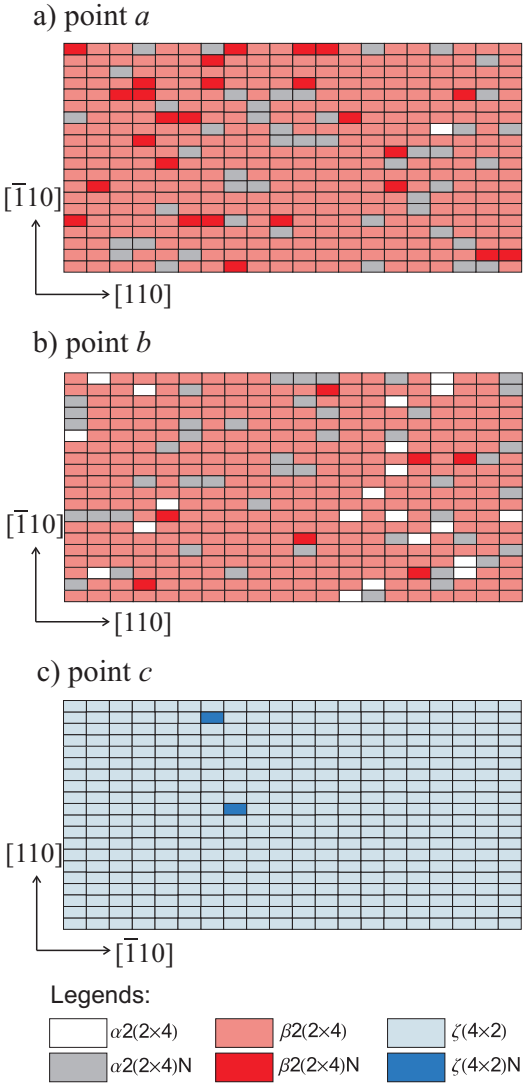


Figure 5.7: Schematic representation of surface configurations obtained by Metropolis Monte Carlo simulation for the growth conditions specified by (a) point *a*, (b) point *b*, and (c) point *c* in the phase diagram in Fig. 5.6(b). Each of the small rectangles represents a $(2 \times 4)/(4 \times 2)$ surface unit cell. The corresponding calculated average N concentrations are: (a) 2 %, (b) 2 %, and (c) 0.15 %.

l^{th} surface layer is given by:

$$c_l^\sigma = \sum_i c_{l,i}^{0,\sigma} e^{-\Delta H_{l,i}^f / k_B T}. \tag{5.18}$$

Here, the superscript σ specifies the surface reconstruction. The index i runs over the nonequivalent substitutional sites in that layer, $c_{l,i}^{0,\sigma} = \frac{n_i^\sigma}{n_{\text{bulk}}}$ is the reference concentration of that site, where n_i^σ is the number of equivalent substitutional sites in that layer and n_{bulk} is the number of all substitutional sites in a bulk (001) layer of an equivalent area, and ΔH_i^f is calculated according to Eq. (5.12) for given chemical potentials (μ_{As} and μ_{N}). For the present systems it turned out that in most layers a single set of symmetry equivalent sites dominates since these sites have a much lower energy than all others (see tables 5.1, 5.2, and 5.3). In this case, to a very good approximation, Eq. (5.18) can be simplified to:

$$c_l^\sigma = c_{l,i_{min}}^{0,\sigma} e^{-\Delta H_{l,i_{min}}^f / k_B T}. \tag{5.19}$$

Here the index i_{min} gives the energetically preferred structure at each layer. For bulk, the equilibrium N concentration (Eq. (5.18)) simplifies to:

$$c_{\text{bulk}} = e^{-\Delta H_{\text{N}^{\text{bulk}}}^f / k_B T}. \tag{5.20}$$

Here $\Delta H_{\text{N}^{\text{bulk}}}^f$ is the formation energy of N incorporation in the bulk relative to the clean bulk system (Eq. (5.15)).

An important issue which needs specific attention is that substituting N in the surface may destabilize the original surface reconstruction and stabilize a new one, i.e. incorporating N may induce a local phase transition. A prominent example is the stabilization of the $\alpha 2(2\times 4)$ reconstruction under modest to poor As chemical potential conditions as clear from Fig. 5.6(b). This reconstruction is in the absence of N unstable against the formation of the $\beta 2(2\times 4)$ surface at modest As chemical potential and against the formation of a $\zeta(4\times 2)$ reconstruction at poor As chemical potential. Also, the presence of N significantly affects the surface phase diagram, and e.g. shifts the transition from $\zeta(4\times 2)$ to $\beta 2(2\times 4)$ towards more As-rich conditions (Fig. 5.6(c)). The fact that N can be incorporated both in the original surface reconstruction and in a new N induced structure is very different from the conventional impurity picture where a fixed host (and thus reference system) is assumed. Calculating the net N solubility at the surface is therefore no longer possible with the simple Boltzmann relation in Eq. (5.19). Various phases that can form at the surface are induced by N itself, which thus affects the net N concentration. It is therefore essential to know what are the phases that contribute to the net N concentration under certain conditions and what are their contributions. This rather unusual behavior will be shown in the following to have subtle consequences in calculating N concentration and solubility.

Monte-Carlo simulations

Because of the complex effect of N on the surface reconstruction and therefore on its concentration, as described above, the direct calculation of N concentration using a conventional impurity

picture (e.g. using Eq. (5.19)) is not possible. An alternative option is to employ Monte Carlo (MC) simulations (see Appendix A). In order to do that the system is described by the following simplified Hamiltonian:

$$H_l(\{\delta_i\}, \mu_{\mathrm{As}}, \mu_{\mathrm{N}}) = \sum_{i,\sigma,n_l} \delta_i^{\sigma,n_l} \Delta H_i^{\sigma,n_l}(\mu_{\mathrm{As}}, \mu_{\mathrm{N}}) + \frac{1}{2} \sum_{\substack{i,j \\ i \neq j}} \sum_{\substack{\sigma\sigma' \\ n_l n_l'}} J_{n_l n_l'}^{\sigma\sigma'} \delta_i^{\sigma,n_l} \delta_j^{\sigma',n_l'}. \quad (5.21)$$

Here, l denotes the surface layer, the indices i and j run over the surface unit cells, σ denotes the surface reconstruction in which the N atom is substituted, n_l specifies the number of N atoms per surface unit cell in layer l, the parameter δ_i^{σ,n_l} is set to one if there are n N atoms in cell i and layer l of reconstruction σ and zero otherwise, and $J_{n_l n_l'}^{\sigma\sigma'}$ is the interaction energy between the two surface reconstructions σ and σ' with n and n' N atoms in the l layer, respectively. For simplicity, we restrict ourselves here to pair nearest neighbor interactions. An extension to higher order interactions is straightforward but was found to be not necessary for the present system.

In the Hamiltonian two sets of parameters enter: the formation energies $\Delta H_i^{\sigma,n_l}$ (which can be directly taken from Fig. 5.4(a) and Fig. 5.6) and the pair interaction parameters $J_{n_l n_l'}^{\sigma\sigma'}$. While the latter are principally straightforward to compute *ab-initio*, a quick analysis showed that some major simplifications are possible for the present system. First, the N-N interaction between N atoms in neighboring cells having the same reconstruction σ is small and can be neglected, i.e., $J_{n_l n_l'}^{\sigma\sigma} \approx 0$. Second, some surface reconstructions are very similar in structure: An example is the $\beta 2(2 \times 4)$ and the $\alpha 2(2 \times 4)$ reconstructions which can be transformed into each other by simply adding/removing a single As-dimer (see Fig. 5.5). *Ab-initio* calculations for a (4×4) cell containing adjacent $\beta 2(2 \times 4)$ and $\alpha 2(2 \times 4)$ cells result in an excess energy of 21 meV (per the 4×4 cell) compared to the sum of the energies of the two single (2×4) cells calculations. Ignoring the interaction energy from the second nearest neighboring cells (because of the periodic boundary conditions) results in an interaction energy of < 11 meV. This value is the approximate side-by-side interaction energy between the two reconstructions on the long boundary side. The short side interaction is expected to be much lower. Hence in this case we specify that $J_{nn'}^{\sigma\sigma'} = 0$. Therefore, if e.g. only these two surface reconstructions occur the Hamiltonian in Eq. (5.21) simplifies to a sum of the on-site terms, i.e.,

$$H_l(\{\delta_i\}, \mu_{\mathrm{As}}, \mu_{\mathrm{N}}) = \sum_{i,\sigma,n_l} \delta_i^{\sigma,n_l} \Delta H_i^{\sigma,n_l}(\mu_{\mathrm{As}}, \mu_{\mathrm{N}}). \quad (5.22)$$

Finally, for the reconstructions that are structurally substantially different the pair interaction energy has a non-zero value ($J_{n_l n_l'}^{\sigma\sigma'} > 0$) and it is expected to be $\gg k_B T$. An example is the case of the $\beta 2(2 \times 4)$ (or the $\alpha 2(2 \times 4)$) and the $\zeta(4 \times 2)$ reconstructions. Since the value of the pair interaction energy is significantly larger than $k_B T$, its exact value has virtually no effect on the MC results. Here, it is taken to be 0.1 eV per boundary as a lower limit; using a larger value does not affect our results.

Using the Hamiltonian in Eq. (5.21) and employing a Metropolis Monte Carlo simulation [168]

the net N concentration under any growth conditions can be calculated. Using this approach we have considered three different growth conditions (defined by μ_{As} and μ_{N}), which are represented by three black dots (marked a, b, c) in the phase diagram in Fig. 5.6(b), and a temperature of 500 °C. These points represent most of the possible situations in the phase diagram and in the phase diagrams for the other layer/systems. The MC calculations have been performed for a periodically repeated sample surface consisting of 20×20 surface unit cells. Convergence checks using a sample consisting of 100×100 surface unit cells have shown that the reduced sample size (20×20) is sufficient to obtain converged results, within a variance of $< 5 \times 10^{-3}$ % in the obtained N concentrations. The input configuration is set to the clean stable surface under the corresponding As chemical potential conditions. The configuration is updated randomly according to the Metropolis Monte Carlo algorithm (see Appendix A) for a sufficiently high number of steps. We note that the formation energies under certain conditions defined by μ_{As} and μ_{N} can be directly deduced from Fig. 5.4 for the clean surfaces (per 1×1 surface area) and from Fig. 5.6 for the surfaces with N substitutions (per 2×4 cell). Moreover, as previously noted, since there is for each reconstruction a set of symmetry equivalent sites that dominate over other sites we have included only single N substitutions in the single unit cell at one of these most favorable energy sites.

Fig. 5.7 shows characteristic configurations from the Metropolis Monte Carlo simulation for each of the three selected growth conditions. First, at point a, the formation energies of N in the $\beta 2(2 \times 4)$ and $\alpha 2(2 \times 4)$ are equal, while the formation energy of N in the $\zeta(4 \times 2)$ is high. Consequently, the average N concentration of 2% obtained by MC is realized by a similar number of N atoms incorporated in both $\alpha 2(2 \times 4)$ and $\beta 2(2 \times 4)$ reconstructions (Fig. 5.7(a)) randomly distributed in the stable clean $\beta 2(2 \times 4)$ surface.

Second, at growth conditions given by point b, the formation energy of substituting N in the $\alpha 2(2 \times 4)$ reconstruction is lower than that in the $\beta 2(2 \times 4)$ reconstruction by about 1.6 eV while it is still much higher in the $\zeta(4 \times 2)$ reconstruction. It is therefore expected that most N atoms are incorporated in the $\alpha 2(2 \times 4)$ reconstruction, as indeed our results show in Fig. 5.7(b). The overall average N concentration is also found to be about 2%. Note that under these growth conditions also some bare $\alpha 2(2 \times 4)$ cells appear at the surface. This is a result of the reduced formation energy of this reconstruction at this point (see Fig. 5.4(a)).

In the last case, at point c, the formation energy of N in $\alpha 2(2 \times 4)$ is much lower than in the other phases implying principally a high concentration. However, since the stable clean surface under these conditions is the $\zeta(4 \times 2)$ surface, the formation of the $\alpha 2(2 \times 4)$ reconstruction is prevented due to the energetically highly unfavorable boundary between $\beta 2(2 \times 4)$ and $\zeta(4 \times 2)$ reconstruction. Note that a local phase transition can occur only if the N atoms are incorporated in neighboring cells and for a cluster large enough to form a stable nucleus. Such agglomeration of N atoms is highly prohibited due to the configuration entropy and hence it is excluded in the considered dilute N concentration limit. Therefore, N is incorporated only into the $\zeta(4 \times 2)$ surface, where the incorporation is rather unfavorable (see Fig. 5.7(c)), which results in a rather low N concentration of 0.15 %.

Calculating N concentrations

The MC calculations allow to determine the net N concentration for each point in the phase diagram, i.e., for each set of μ_{As} and μ_{N} pairs. This would be however rather time consuming. We therefore performed some further simplifications in calculating the N concentration and used the MC calculations to verify the accuracy of these approximations. Based on the analysis of the MC calculations we approximate the total concentration of N as a sum of the concentrations from the independent phases:

$$c_l = \sum_{\sigma' \in \sigma} c_l^{\sigma'}. \tag{5.23}$$

The index σ' goes only over the *relevant* reconstructions, i.e., only the reconstructions which have a negligible interaction energy with the stable clean reconstruction for the given As chemical potential are considered. The independent concentrations c_l^{σ} are calculated according to Eq. (5.18), with the formation energies ΔH_i^f are calculated according to Eq. (5.12) and taking the stable clean reconstruction at the specific μ_{As} as the reference surface. Note that Eq. (5.23) principally allows N atoms to occupy the same site more than once. Therefore, this approximation is valid only in the dilute concentration limit. Using this approach, lines with constant N concentration have been calculated and plotted in Fig. 5.6. A growth temperature of 500 °C has been assumed. Fig. 5.6(b) shows that these approximations work well and provide results in good agreement with the MC simulations.

5.7.2 Results

Let us start by discussing the solubility limit in GaAs bulk. From the concentration lines in Fig. 5.6(a) it is clear that the maximum solubility limit which can be achieved in bulk GaAs under thermodynamic equilibrium is very low. It is of the order of $\sim 10^{-3}$ % under extreme N and As poor conditions.

At the second anion surface layer (c-layer), a set of three selected N-concentration lines (obtained using Eq. (5.23)) are shown in the phase diagram in Fig. 5.6(b). The concentration lines were chosen so that they partially coincide (at the As-rich limit) with the bulk concentration lines to allow for a direct comparison. We note the close relation between the concentration lines and the stability lines (the solid lines) of the corresponding phases. The concentration lines indicate a general increase of N solubility with decreasing As chemical potential, except at the discontinuity associated with the phase transition between the $\beta 2(2 \times 4)$ and the $\zeta(4 \times 2)$ reconstruction. This discontinuity causes an abrupt decrease in the N solubility. These results highlight again the large effect the specific surface reconstruction has on the N solubility. This insight allows us to determine the optimal As chemical potential (μ_{As}) at which N concentration becomes maximal: It is just before the phase transition from $\beta 2(2 \times 4)$ to $\zeta(4 \times 2)$, i.e., at $\mu_{\mathrm{As}} - \mu_{\mathrm{As}^{\mathrm{bulk}}} = -0.58\,\mathrm{eV}$ (Fig. 5.6(b)). We note that under these conditions it is possible to achieve N concentrations as high as 12.5 %. Comparing this solubility to that in the bulk, an increase by about 3–5 orders of magnitude can be realized.

In the next step we consider the case of N incorporation at the top surface layer (i.e., in the first anion layer). The corresponding concentration lines are included in Fig. 5.6(c). The concentration lines are again chosen to coincide with those in Fig. 5.6 (a and b) at the As-rich region to allow for a direct comparison between the three cases. Since the $\zeta(4\times2)$ reconstruction is usually not suitable for growing GaAs, the optimal As chemical potential (μ_{As}) to achieve high N concentrations is equivalent to that in the case of the second anion layer substitution, i.e., at $\mu_{\mathrm{As}} - \mu_{\mathrm{As}^{\mathrm{bulk}}} = -0.58\,\mathrm{eV}$.

Comparing the case of incorporation into the first anion layer with that into the second anion layer, we note that the N solubility in the first anion layer is lower by up to 4 orders of magnitude (under moderate As conditions). This finding is a direct consequence of the reconstruction induced strain in the subsurface layers as discussed in Sec. 5.5. Comparing to the bulk solubility, the maximal N concentration for the first anion layer incorporation is still higher by $\lesssim$1–2 orders of magnitude.

In order to summarize the main results, we show in Fig. 5.9(a) the maximal N solubility as function of the N chemical potential for optimal As chemical potential conditions ($\mu_{\mathrm{As}}-\mu_{\mathrm{As}^{\mathrm{bulk}}} = -0.58\,\mathrm{eV}$) and for the three different N incorporation mechanisms: surface, subsurface, and bulk. Experimentally relevant N concentrations are indicated in the shaded region. The figure shows that under the condition $\mu_{\mathrm{N}} \leq \mu_{\mathrm{N}^{\mathrm{N_2}}}$ (Eq. (5.14)) only the subsurface incorporation predicts concentrations in accordance with experimental values. However, we leave the discussion about the incorporation mechanisms and experiments for Chapter 7 after discussing the kinetic barriers in the following chapter (Chapter 6) to reveal the interplay between thermodynamics and surface kinetics.

5.8 Solubility of N at InAs(001) surfaces

We now determine the solubility of N at the InAs(001) surface by employing the same procedure as used for the GaAs(001) surface. In the first step the clean surface phase diagram has been calculated (Fig. 5.4(b)). The phase diagram provides a set of low energy reconstructions which we consider for N substitutions. These are the $\beta2(2\times4)$, $\alpha2(2\times4)$, md(2×4) and the $\zeta(4\times2)$ reconstructions. Note that the md(2×4) (mixed-dimer(2×4)) cell has an As-Ga dimer in the top surface layer and 8 Ga atoms in the second layer (see Fig. 5.3(c) and Ref. [119]). Like for GaAs, the c(4×4) reconstruction was not considered. For these reconstructions all possible N substitutions have been calculated in the first and second anion layers. The thus determined most favorable substitutions are found to be the same as for GaAs(001) surface reconstructions.

5.8.1 Stability phase diagrams and concentration lines

The computed stability phase diagrams for InAs(001) are shown in Fig. 5.8(a) for N in InAs bulk, in Fig. 5.8(b)/(c) for the second/first anion surface layer. Very similar conclusions as for the GaAs(001) surface can be drawn: (*i*) All surfaces reduce the energy to substitute N. (*ii*) The second anion layer is energetically more favorable for N than the first. (*iii*) N exhibits a complex behavior at the surface and affects the stability of the reconstructions. (*iv*) Finally, a

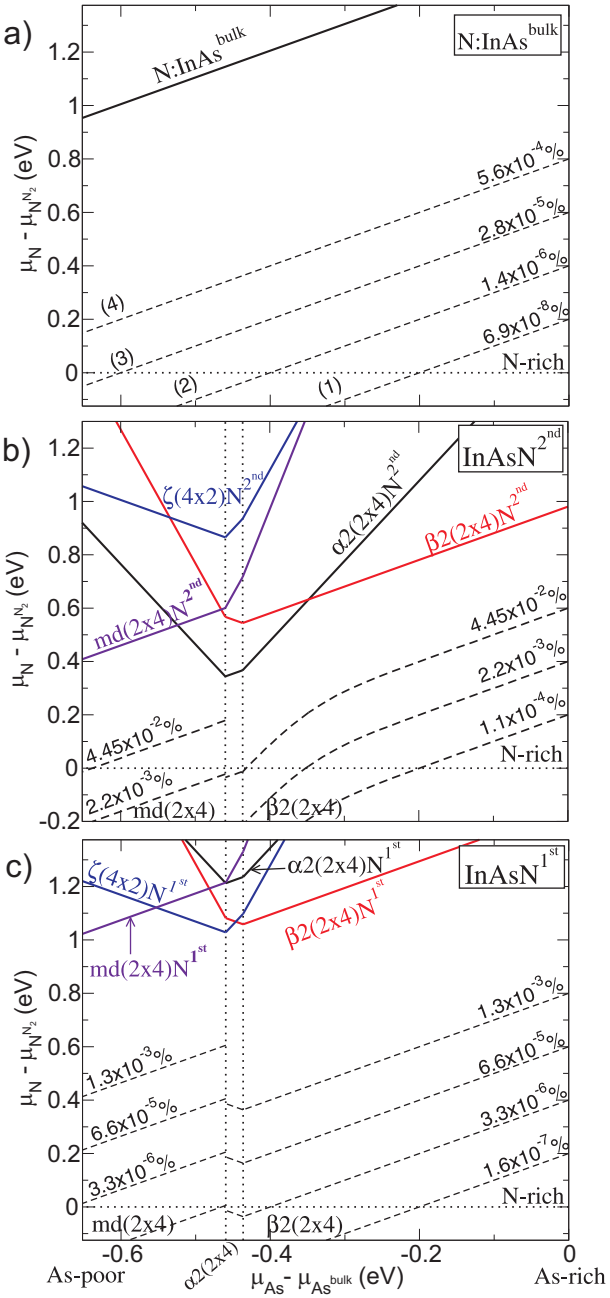


Figure 5.8: Same as in Fig. 5.6 but for InAs. The dotted lines indicate the phase boundaries between the $\beta 2(2\times4)$, $\alpha 2(2\times2)$, and $\mathrm{md}(2\times4)$ reconstructions.

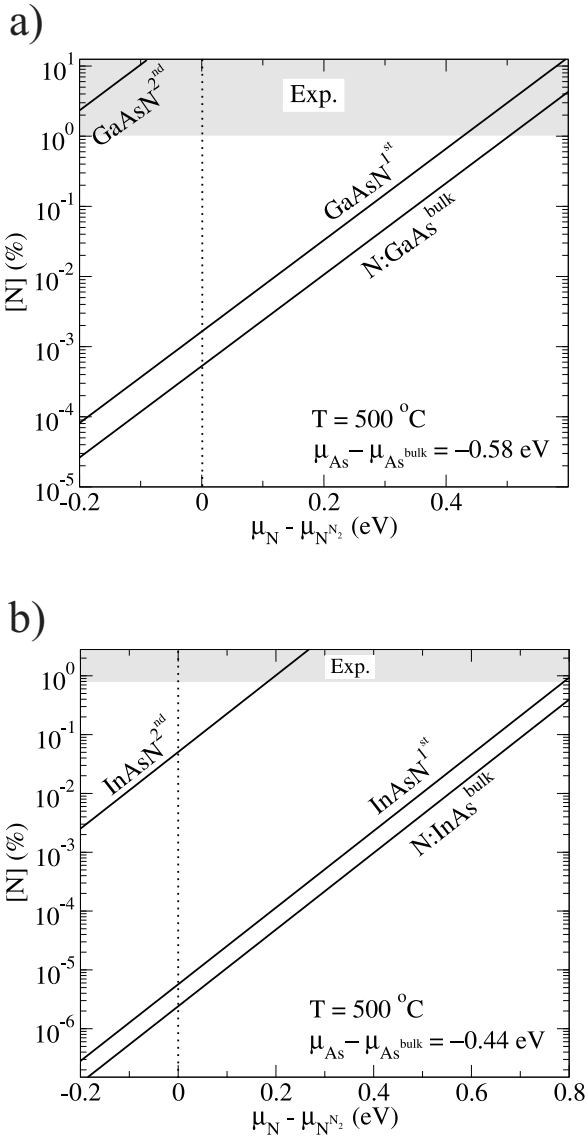


Figure 5.9: Calculated N concentration in bulk and at surface and subsurface layers as a function of N chemical potential and at a temperature of $T = 500\,^{\circ}\mathrm{C}$ for (a) GaAs at $\mu_{\mathrm{As}} - \mu_{\mathrm{As^{bulk}}} = -0.58\,\mathrm{eV}$, and for (b) InAs at $\mu_{\mathrm{As}} - \mu_{\mathrm{As^{bulk}}} = -0.44\,\mathrm{eV}$. The different lines represent the different incorporation mechanisms (see text). The shaded regions represent experimentally relevant concentrations (Refs. [34, 169, 35, 36] for GaAs and Refs. [12, 170, 171, 172] for InAs).

key difference is that the formation energies of N substitutions in InAs are generally higher than those in the corresponding GaAs case. A consequence is that all N substitutions in InAs are thermodynamically not stable.

Contour lines with constant concentration were also calculated (Fig. 5.8). They were chosen so that they coincide partially (at the As-rich conditions limit) with the corresponding lines in the GaAsN phase diagrams to allow for a convenient comparison between the two materials.

5.8.2 Results

From the concentration lines a slightly different behavior of N solubility than in the case of GaAs results. The differences are mainly related to the presence of an additional surface reconstruction (md(2×4)) that becomes stable under As-poor conditions. Nevertheless, very similar features as in the case of GaAs are found and can be summarized as follows: (*i*) The N solubility is strongly affected by surface reconstruction.(*ii*) The N solubility generally increases with decreasing As chemical potential. The discontinuities occur at phase transitions (here at the md(2×4) phase). In contrast to the GaAs case the discontinuity at the boundary does not lead to a reduction in the N solubility when going towards As-poor conditions. (*iii*) The solubility of N in the subsurface region (second anion layer) is higher by up to 4 orders of magnitude than the solubility at the uppermost surface layer (first anion layer), which is similar to the case of GaAs. (*iv*) The solubility at the subsurface is higher than that in the bulk by about 3–4 orders of magnitude, while no considerable solubility enhancement can be noticed in the case of the top surface layer. For both surface and subsurface incorporation the optimal As chemical potential (μ_{As}) to achieve a maximal N concentrations is around the region of the phase transition from $\beta2(2\times4)$ to $\alpha2(2\times4)$, i.e., $\mu_{\mathrm{As}} - \mu_{\mathrm{As^{bulk}}} = -0.44\,\mathrm{eV}$.

The N solubility as a function of the N chemical potential and at the optimal As chemical potential for the three different N incorporation mechanisms (surface, subsurface, and bulk (equilibrium) incorporation) has been calculated and is shown in Fig. 5.9(b). Experimentally, the achieved N concentration in InAs is low and in the range of $0.8 - 2.8\,\%$ [12, 170, 171, 172]. However, it is reported that isolated N substitutions (i.e. no N clusters) are present only for the low concentrations ($0.5\,\%$), at higher concentrations ($1.2\,\%$) di-nitrogen In complexes tend to form [173]. The experimentally relevant concentrations are indicated in Fig. 5.9(b) by the shaded region. The figure shows that under the condition $\mu_{\mathrm{N}} \leq \mu_{\mathrm{N^{N_2}}}$ (Eq. (5.14)) *none* of the incorporation mechanisms predicts concentrations that coincide with experimental values. As for GaAs, the incorporation mechanisms are discussed in Chapter 7 after discussing the kinetic barriers in the following chapter (Chapter 6).

Comparing Figs. 5.9(a) and (b), it becomes clear that the N solubility is generally less in InAs than in GaAs: For the incorporation in the bulk it is less by about one order of magnitude, in the first anion layer by ∼1–4 orders of magnitude, and in the second anion layer by ∼1–2 orders of magnitude. The significant differences have their origin in the interplay between chemical and strain energy; the lattice mismatch between GaAs/GaN is larger than that between InAs/InN while the binding energy of InN is less than that of GaN. Due to the better strain relaxation at the surface layers, the effect of the lower In-N binding energy becomes more pronounced/isolated

at the surface.

5.9 Summary

Based on the comprehensive study of the thermodynamics of N in GaAs and InAs we were able to determine its solubility for various incorporation mechanisms: in the surface, in the subsurface, and in the bulk, as a function of the growth conditions. First, we have shown that the the formation energy of N substitution is significantly affected by the structure of the surface (i.e. its reconstruction) due to the reconstruction-induced surface strain. Hence, the calculated stability phase diagrams of N substitutions showed that the solubility of N at surfaces is a complex function of growth conditions and depends strongly on the specific surface reconstruction and on the incorporation layer. Furthermore, contrary to what is expected from the pure bulk behavior, the maximal N concentration is not achieved under extreme Ga/In-rich conditions but rather under less Ga-rich conditions. The solubility of N is found to be highest at the subsurface for all considered surface reconstructions. Finally, our results showed that the solubility in InAs is generally lower compared to GaAs by 1–4 orders of magnitude. This becomes more pronounced at surfaces due to the more noticeable contributions of the chemical binding energy as a result of strain energy relaxation.

To accomplish a comprehensive understanding of the mechanisms that control N solubility in practical growth experiments, the knowledge of surface thermodynamics is not sufficient. Therefore, in the next chapter we address the kinetic mechanisms for N incorporation and determine the associated barriers. In Chapter 7 we discuss the interplay between thermodynamics and surface kinetics in determining N solubility.

Chapter 6

Kinetic Mechanisms and Barriers for N Adatoms at GaAs(001) Surface

6.1 Introduction

Molecular-beam epitaxy growth mode is operating in a regime far away far from thermodynamic equilibrium, and hence adatom kinetics play a central role. For instance, in the previous chapter we have shown that the solubility of N in the subsurface layer is highest, and it is up to 5 orders of magnitude higher than the equilibrium bulk solubility. By employing the appropriate kinetic pathway, i.e., subsurface incorporation and freezing-in mechanism, it is possible to achieve high N concentrations that are in accordance with experimental results. Therefore, it is essential to identify the accessible mechanisms that control N incorporation under typical growth temperatures, and hence the kinetic mechanisms that allow for subsurface incorporation and freezing-in during growth in order to employ them in practice. Important consequences of adatom kinetics on the properties of the MBE grown dilute nitride alloys will be discussed in the following chapter.

A main aim of this chapter is therefore to explore the kinetic mechanisms controlling the incorporation of N adatoms at the surface/subsurface. In a first step we concentrate on achieving a detailed understanding of the surface diffusion of N adatoms. Therefore, we first discuss the basic methodology to study adatom diffusion and employ it. Several challenges arise in the case of N adatoms on the GaAs(001) surface. These challenges are discussed and analyzed in detail. After developing and employing the appropriate methodology, we determine adatom binding sites, diffusion effective activation barriers, diffusion constants, and diffusion anisotropy. In a second step, we explore possible kinetic mechanisms resulting in subsurface incorporation of N and calculate the corresponding energy barriers. By a careful analysis of our results a comprehensive understanding of the incorporation kinetics of N adatoms is achieved. We note that throughout this chapter the $\beta 2(2 \times 4)$ reconstructed GaAs(001) is assumed as a prototype because of its stability over a wide range of chemical potentials and its importance in molecular beam epitaxy growth.

6.2 Adatom potential energy surface

In order to theoretically identify the diffusion paths and barriers for an adatom it is in principle required to identify its potential energy landscape E in a complex multi-dimensional configurational space. This space is generally defined not only by the adatom spacial coordinates $\mathbf{R}^{\mathrm{ad}} = (X, Y, Z)$ but also by the spacial configuration of surface atoms $\{\mathbf{R}^{\mathrm{surf}}\}$, i.e.,

$$E \equiv E(\mathbf{R}^{\mathrm{ad}}, \{\mathbf{R}^{\mathrm{surf}}\}). \tag{6.1}$$

In practice, it is exceedingly computationally demanding to explore this space using standard *ab-initio* methods. A common approach to simplify the problem is based on two assumptions. First, it is assumed that the adatom surface interaction potential energy is restricted to the spacial coordinates of the adatom, i.e.,

$$E \equiv E(X, Y, Z). \tag{6.2}$$

Second, in the direction perpendicular to the surface (z-direction) it is assumed that for each pair of adatom lateral coordinates $(x, y) \in (X, Y)$ a single energy minimum of the potential energy exists, say at $z = z_0$. At this position the adatom gets adsorbed initially and starts to move in the potential energy surface (PES) defined by Z_0. A single two-dimensional PES results $E(X, Y, Z_0)$ that is assumed to describe adatom diffusion. Using this notation, the PES experienced by the adatom can be defined as:

$$E_{\mathrm{PES}}(X, Y) = \min_{Z, \{\mathbf{R}^{\mathrm{surf}}_{\mathrm{relax}}\}} \left[E^{tot}_{\mathrm{system}}(\mathbf{R}^{\mathrm{ad}}, \{\mathbf{R}^{\mathrm{surf}}\}) \right] - E^{tot}_{\mathrm{slab}^{\mathrm{clean}}} - E^{tot}_{\mathrm{atom}}. \tag{6.3}$$

Here, $E^{tot}_{\mathrm{system}}$ refers to the total energy of the system (adatom + slab), $E^{tot}_{\mathrm{slab}^{\mathrm{clean}}}$ is the total energy of the adsorbate-free surface slab, and E^{tot}_{atom} is the total energy of an isolated adatom infinitely far away from the surface. The subset $\{\mathbf{R}^{\mathrm{surf}}_{\mathrm{relax}}\}$ indicates that the energy minimization in Eq. (6.3) takes into account only the relaxations of substrate atoms, in addition to the Z coordinates of the adatom (its height above the surface). Note that the energy definition in Eq. (6.3) corresponds to the *binding energy* of the adatom.

In order to accomplish the PES in practice constrained relaxations are applied: For a set of lateral mesh points on the surface covering its irreducible part, the adatom is placed at an initial height above the surface and allowed to relax only in the vertical direction, while the surface atoms are allowed to fully relax. Interpolating between the positions versus energy points results in a two-dimensional (2D) potential energy map (PES). Once the PES is known, the binding sites and transition states can be directly extracted from the minima and saddle points, which provides an immediate access into e.g. diffusion paths or diffusion coefficients [174]. This approach, which was first employed by Brocks et al. [175], has become very popular over the last years and has been successfully applied to address a wide variety of surface related questions [175, 176, 177, 178, 179, 180, 181, 182, 183, 153, 184].

The validity of the approximations necessary to construct a 2D-PES can be discussed con-

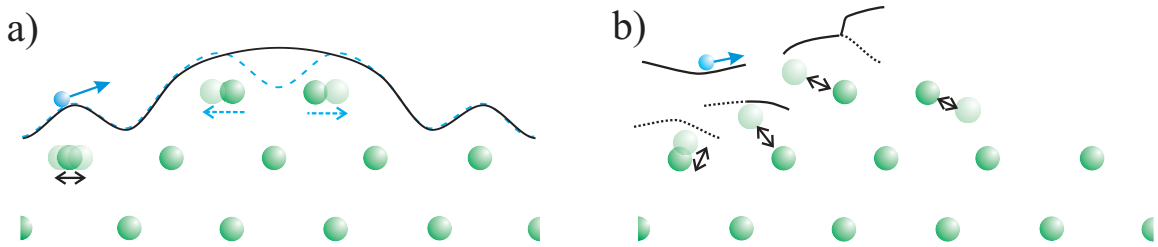


Figure 6.1: Cross sectional schematic representations of configuration profiles associated with the surface diffusion of adatoms. In (a), the conventional diffusion is characterized by smooth PES surfaces and associated with small relaxations of surface atoms. In (b), the diffusion of a highly reactive adatom is characterized by multivalued surfaces which are associated with large surface atomic relaxations (see text). The blue spheres represent adatoms while the green ones represent surface atoms.

sidering fundamental aspects of conventional adatom-surface interaction. To be more specific, consider first the surface growth of a simple elemental crystal. In this case the bond-strength between an adatom and a surface atom is identical to that in between surface atoms. Since surface atoms have higher coordination numbers than the adatoms they are more strongly bound and therefore they will be only modestly affected by the diffusing adatoms on the surface. A consequence is that the surface relaxations will be *small* (few tenths of Å) and restricted to the vicinity of the adatom. This implies that adatom induced atomic relaxations can be considered as small perturbation around the equilibrium position in the adatom-free case. Relaxation can then be regarded as a simple elastic (i.e. non-bond-breaking) mechanism. The condition of staying in the elastic regime is fundamental for the validity of the 2D-PES concept. If plastic relaxation (i.e. bond-breaking) occurs, the reduction from the full configuration space (Eq. (6.1)) to the 2D adatom space is no longer valid. The configuration profile in the elastic regime is schematically sketched in Fig. 6.1(a) by the solid black line. Following our discussion the diffusion of the adatom is accompanied by small relaxations of surface atoms (black double arrows), resulting in a single-valued smooth energy surface map.

A similar behavior may also be expected for the surface growth of typical binary materials, where the bond-strength between an adatom and surface atoms is similar to the bond-strength in between the surface atoms. However, the existence of different types of bonds and atoms at reconstructed surfaces may lead to additional complications. For instance, it was first reported by Kley et al. [111] that the conventional approach to map the PES is not sufficient to describe the migration of Ga and Al adatoms on the $\beta 2(2 \times 4)$ reconstruction of GaAs(001). For this surface, a second deeper energy minimum was found. This minimum forms in the center of the As dimers present at that surface. To access it the bond between the As-As dimer has to be broken before relaxing the system. For this surface, the actual PES is found to have an additional dimensionality: It is a function of not only the adatom lateral coordinates but also of the As-dimer bond configuration (broken or bounded). Thus, two PES's exist that can be associated with different metastable surface configurations. By constructing the PES that includes the new diffusion channels Kley et al. [111] were able to describe adatom diffusion, i.e., the concept of PES mapping remains useful when extending the original concept and adding a few (in this

case a single) dimensionalities. The diffusion profile of the adatom for this case is schematically sketched in Fig. 6.1(a) (dashed blue line), and the associated modifications in the configuration surface atoms (e.g. a broken surface bond) are indicated by the blue dashed arrows. Note that both PES profiles in Fig. 6.1(a) are smooth and differ only near that specific surface bond.

For compounds with more than two constituents (e.g. ternaries and quaternaries) a similar behavior is expected provided that the bond strengths between the various constitutes are comparable. If this is not the case, i.e., the adatom-surface bond becomes significantly strong, the situation becomes very different. In such a case, plastic deformation may occur, resulting in the breaking and forming of bonds. In this case, the binding energy of the adatom becomes a strong function of its position and of the local atomic configuration around it. For instance, relaxing the adatom starting from different initial heights above the surface at the same lateral position can result in different local energy minima, i.e., giving rise to a multivalued PES along the normal direction. Therefore, the overall adatom potential energy becomes complex. As a consequence, the potential energy in the 3D configuration space of the adatom appears as a set of multivalued discontinuous surface. This case is schematically expressed in Fig. 6.1(b) and highlights the breakdown of the conventional 2D-PES. The different lines in the figure correspond to the various minima as a result of the large surface relaxations (double arrows). The solid parts of the lines indicate the parts that will be detected when applying the conventional 2D-PES mapping. As becomes obvious from Fig. 6.1(b) these parts do not describe the actual adatom diffusion nor do they correspond to a single smooth 2D-PES.

Therefore, despite the successes of this approach, its underlying assumptions need to be carefully verified. For instance, in this work, the strongest bond a N adatom can form on the GaAs(001) surface is with a Ga atom with a binding energy of 2.24 eV, while the bond strength between Ga-As atoms is 1.63 eV, i.e., it is weaker by 27 % [106]. This strong adatom-surface interaction makes it essential to carefully examine the adatom-surface potential energy and the applicability of the 2D-PES mapping, as will follow.

6.2.1 Mapping the PES of a N adatom on GaAs(001)

In order to identify the binding sites and the diffusion barriers of N adatoms on the $\beta 2(2 \times 4)$ GaAs(001) surface we map in a first step the potential energy surface of a N adatom employing the conventional 2D adatom PES concept described above. A repeated 4×4 surface supercell was used in order to reduce the fictitious adatom-adatom interactions due to periodic boundary conditions.[1] The mapping was performed on the irreducible part of the surface cell[2] using a uniform rectangular mesh of spacing 0.815 Å in both directions, which corresponds to a 20×20 mesh in the 4×4 cell.

Performing the PES mapping, a strong dependence of the relaxed configuration (and hence its energy) on the initial vertical position of the N adatom (its height above the surface) is found at many mesh points. Selected examples of such cases are shown in Fig. 6.2. In each of the

[1] Further computational details can be found in Sec. 5.4.3

[2] Due to the mirror symmetry in the structure of the $\beta 2(2 \times 4)$ surface unit cell it becomes sufficient to map a 2×2 area.

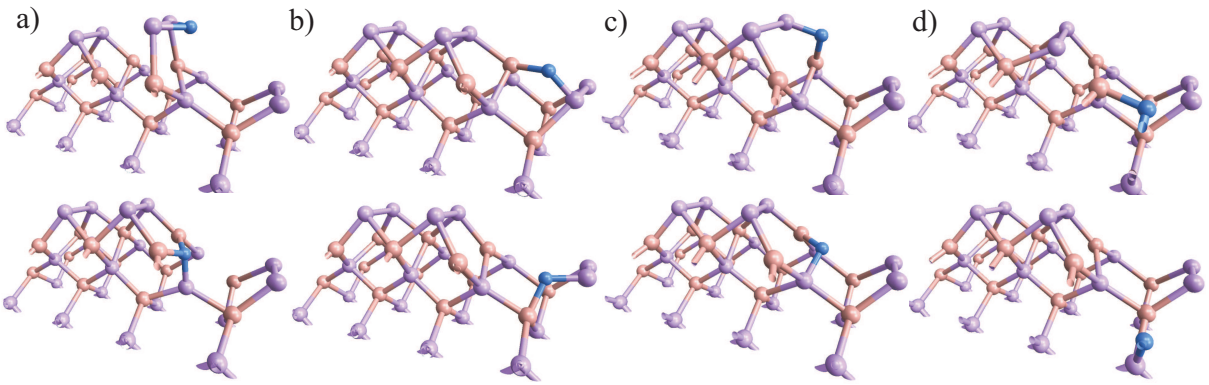


Figure 6.2: Examples showing the effect of the initial vertical position of the N adatom above the surface on the resulted relaxed geometry. In each subfigure (a, b, c, and d) the N adatom is fixed at a certain lateral position above the surface and allowed to relax vertically. The upper panel of each subfigure is for a higher adatom initial position, and the lower panel is for a lower initial position. Ga atoms are in brown, As are in purple, and N are in blue.

cases (a, b, c, and d) shown there, the configuration in the upper panel results when positioning the adatom far away from the surface whereas the configuration in the lower panel results from initial positions closer to the surface. In the first case (upper panel) the adatom relaxes to a first energy minimum causing strong outward relaxations of surface atoms, i.e., the adatom attracts surface atoms. In contrast, relaxing the structure with the adatom initially at a lower position (lower panel) results in an inward relaxation of the adatom and the occupation of a second energy minimum. The occurrence of multiple minima is associated with large local structural deformations as can be seen from the figure.

A straightforward extension of the 2D-PES concept is to assume that the multiple vertical minima are associated with different PES's. In this case we have layers of continuous PES's that can be easily visualized or analyzed. In order to check the applicability of such an approach we have constructed a PES by interpolating[3] the energy values of the *first* local minima for the adatom approaching the surface (corresponding to the *vertically highest* obtained adatom relaxed positions). The thus constructed PES is shown in Fig. 6.3.

To check of the validity of the underlying approximations we first examine the lateral forces acting on the N adatom at different positions. The force acting on an adatom at coordinates τ moving in the potential energy surface E_{PES} is given by

$$\begin{aligned} \mathbf{F}(\tau) &= -\nabla E_{\mathrm{PES}}(\tau) \\ &= -\left(\frac{\partial E_{\mathrm{PES}}(\tau)}{\partial x}\,\hat{i} + \frac{\partial E_{\mathrm{PES}}(\tau)}{\partial y}\,\hat{j} \right). \end{aligned} \tag{6.4}$$

The force in z-direction is *zero* by the definition of the PES (see Eq. 6.3). Therefore, the forces acting on the adatom should *ideally* correspond to the negative gradients of the interpolated PES at the corresponding positions without significant deviations.[4] On the other hand, using

[3] The mirror symmetry is applied to produce the energy map for the complete 2×4 unit cell. The interpolation is done using cubic splines and using the 2D periodicity of the surface unit cell.

[4] This is true provided that the mesh is dense enough (converged sampling).

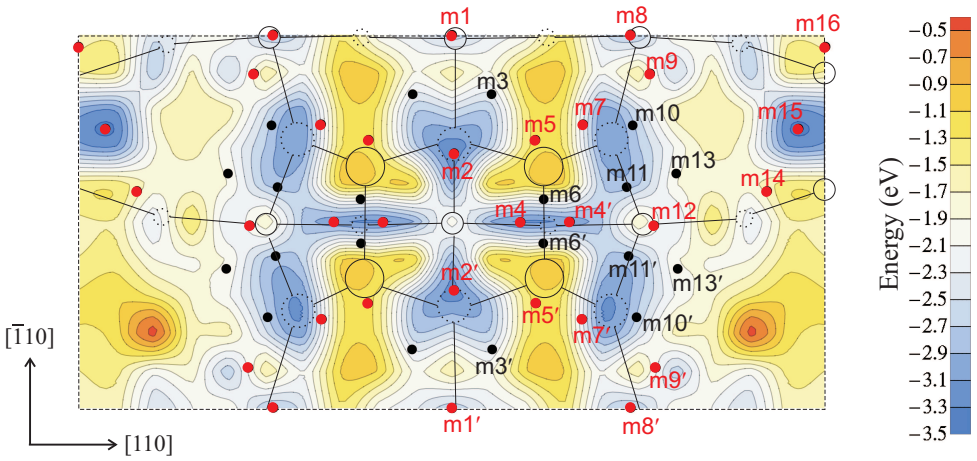


Figure 6.3: Top view of the PES of the N adatom on the $\beta 2(2 \times 4)$ reconstruction of the GaAs(001) surface obtained by the conventional 2D mapping, i.e., it is defined by the first local minima points of the N adatom approaching the surface in the vertical direction. The red points are the exact position of the binding sites obtained after performing full structural relaxation with the adatom around minima regions, while the black points are the additional binding sites obtained during successive NEB calculations (see text). The positions of Ga and As atoms of the clean surface are marked with circles in dotted and solid borders, respectively.

the forces acting on the adatom in the interpolation, together with the energies, should result in increasing its *accuracy*, without a significant change in its shape. In Fig. 6.4(a) we show the actual forces acting on adatom obtained from our calculations at all mesh points. As can be seen, a key result of this analysis is that at some points on the PES the actual forces do *not* match the negative gradients of the underlying PES. Inconsistent forces are indicated by the red arrows in Fig. 6.4(a). Consequently, using the forces to supply the actual gradients in the interpolation has dramatic consequences on the PES. This can be seen from the PES in Fig. 6.4(b) which features substantially different features than the original one in Fig. 6.4(a), in contradiction with the expectations. This strange and unexpected observation was the first indication that for the system considered here that the 2D mapping of the PES for the adatom is insufficient to describe its actual potential energy landscape. A mapping of the multiple vertical minima onto layers of continuous 2D-PES's is therefore not justified.

To understand the significance of the forces we make a schematic explanation in Fig. 6.4(c) for a one dimensional case. In the ideal case, it is expected that the gradients obtained from an interpolation that uses only the energy points (black dotted curve) will roughly correspond to the actual forces acting on the adatom (black arrows). By including the actual forces in the interpolation one can increase its accuracy without changing its shape considerably. On the other hand, the existence of forces at some points that are inconsistent with those obtained from the expected gradients of the energy curve (red arrows) indicates that the interpolation is *invalid*. Using these forces in the interpolation results in a potential energy that has very different features (red dashed curve). Therefore, it becomes not suitable to project the adatom potential energy onto a single potential energy curve/surface.

As a further consistency check of the 2D-PES in Fig. 6.3 we examine the minima points at

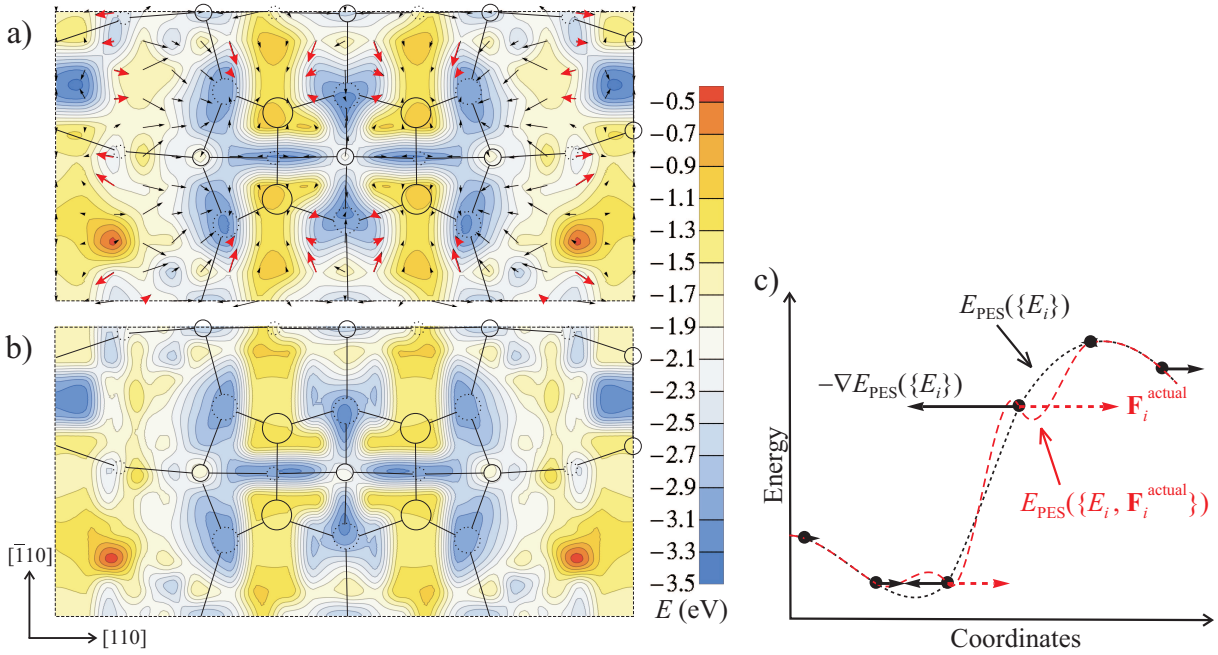


Figure 6.4: (a) The PES of the N adatom interpolated using the energy values only. The Forces acting on the adatom at the mesh points are shown as arrows. Red arrows indicate some of the forces that disagree in direction with the negative gradients of the interpolated PES. (b) The interpolated PES obtained by additionally using the *actual* gradients (obtained from the forces shown in (a)). The effect of including the forces in this case is explained in the schematic plot in (c) for a one dimensional case (see text). The black dots indicate the sampling points.

the PES. The exact positions and energies of the first set of minima points were identified after allowing a full relaxation of the adatom starting from mesh points that are close to the minima regions. The hence identified binding sites are shown in Fig. 6.3 as small *red* dots. For simplicity we shall hereafter refer only to the points in the half mirror-symmetric part. These are the points with the labels m1, m2, m4, m5, m7-9, m12, m14-16, and their primed counterparts. The energy values at these points are presented in Table 6.1, and their corresponding relaxed atomic structures are shown in Fig. 6.5. We note that the remaining labeled points in Fig. 6.3 (in black) will be discussed later.

It is interesting to see from Fig. 6.3 that some of the found minima points substantially

Minima label	E (eV)	Minima label	E (eV)
m1	−2.61	m9, m9′	−2.72, −2.76
m2, m2′	−3.22, −3.18	m10, m10′	−2.03, −1.96
m3, m3′	−3.21, −3.17	m11, m11′	−2.80, −2.76
m4, m4′	−3.14, −3.06	m12	−2.52
m5, m5′	−3.77, −3.76	m13, m13′	−2.42, −2.49
m6, m6′	−2.82, −2.79	m14	−2.67
m7, m7′	−3.12, −3.15	m15	−3.25
m8	−2.67	m16	−3.71

Table 6.1: Energy values at the PES minima points (binding sites), shown in Fig. 6.3.

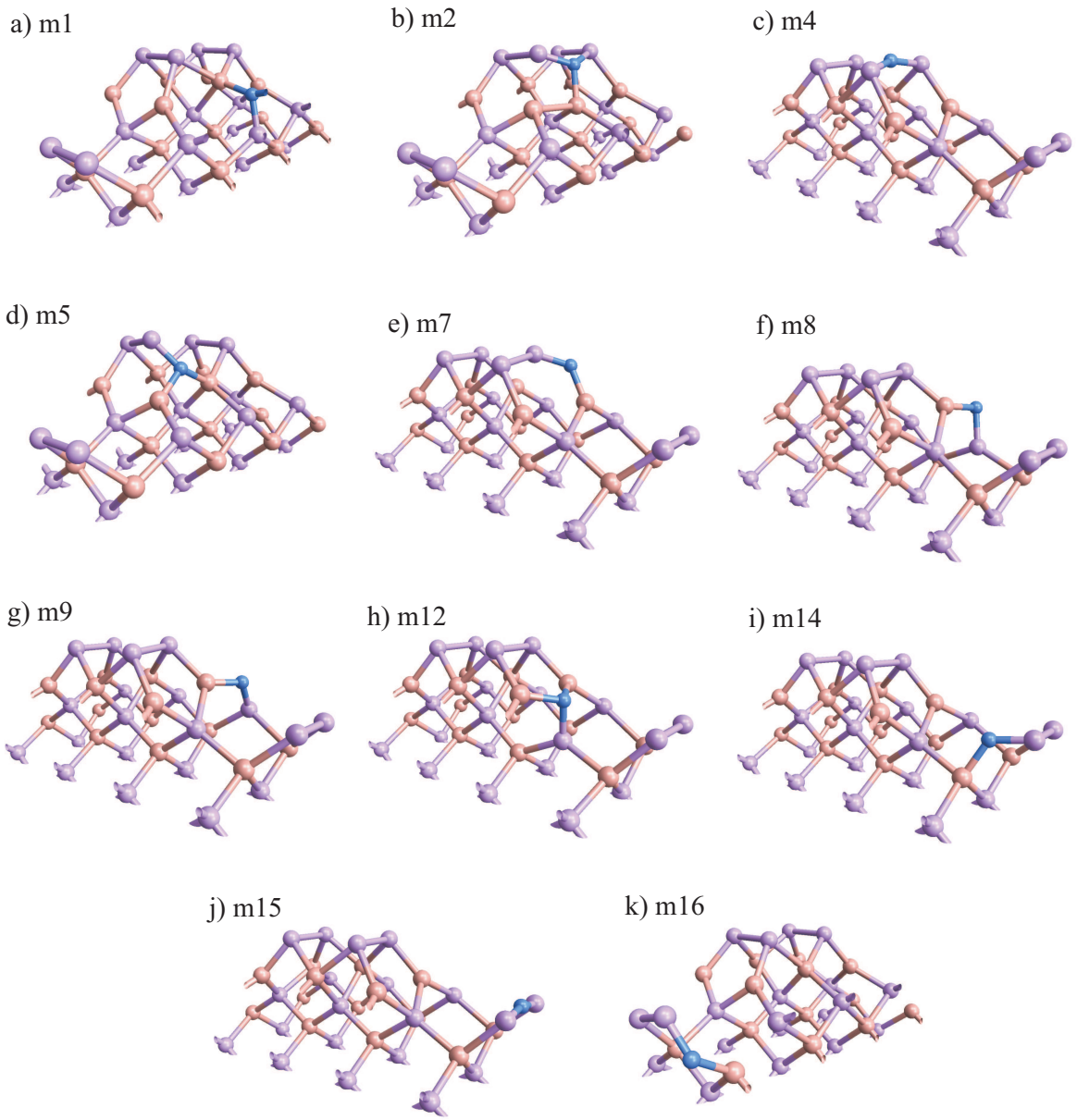


Figure 6.5: Atomic structures corresponding to the minima points on the PES in Fig. 6.3 as indicated.

deviate both in value and position with what is expected from the interpolated PES. First, energetically degenerate absolute minima points were found at subsurface sites below As dimers atoms. These points are labeled m5/5′ (Fig. 6.5(d)) and m16 (Fig. 6.5(k)). The m5/5′ points, which are located below the uppermost As dimer atoms, were obtained when fully relaxing the adatom from mesh points that are close to the expected positions of the minima m2/2′ (which are located in between the two surface dimers). The m16 point, which is located below an As atom of the lower dimer (trench dimer), was found when fully relaxing the adatom in the region close to a trench dimer atom in the $[\bar{1}\bar{1}0]$ direction (in this region a depression in the 2D-PES appears). Second, the positions of some minima points (e.g. m9/9′, m7/7′, m12, and m14) appear to be

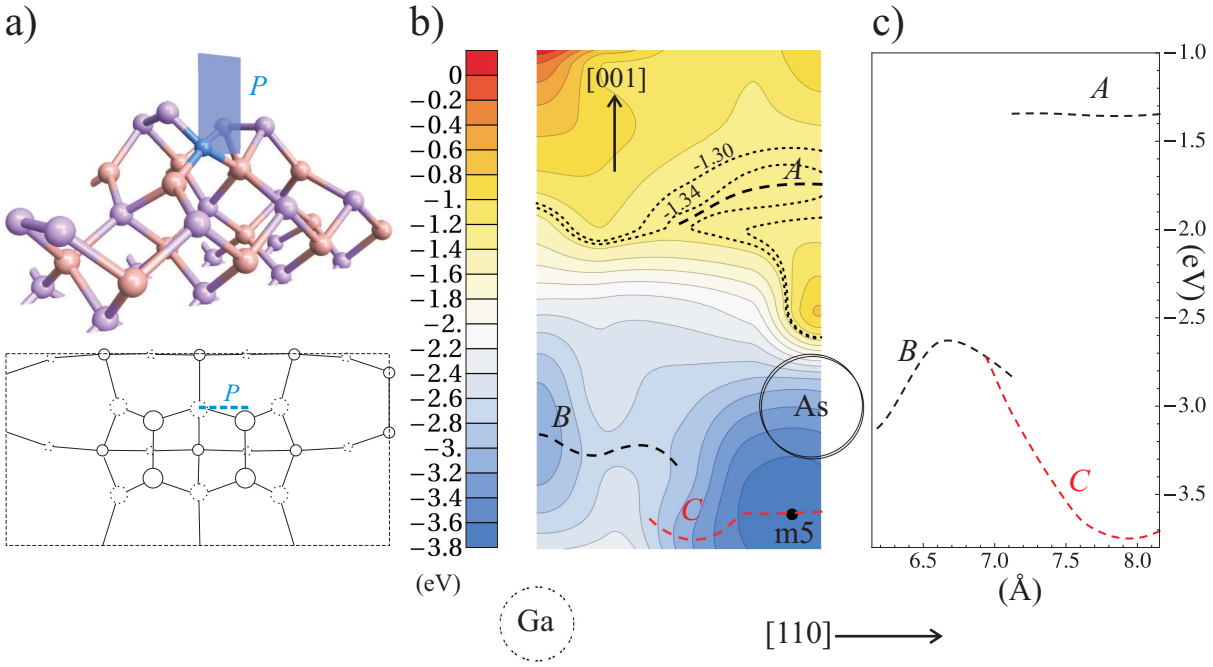


Figure 6.6: (a) Upper panel: A 3D structural model of the m5 binding configuration showing the cross-sectional plane (P) that is used to explore the N adatom energy around m5 vertically. In the lower panel the projection of P is shown in a top view by the dashed blue line. (b) The corresponding potential energy landscape in the plane P. The dashed black curves (A and B) show the position of first local minima for the N adatom in the vertical direction, while the dashed red curve C shows the positions of deeper potential energy minima. For reference, some atomic positions of the clean surface are projected in the plane P and shown as circles scaled inversely in diameter with their distance from the surface. In addition, extra contour lines around A, with their values in eV, are plotted for clarity. The energy values from (b) are also shown in the right panel (c).

inaccurately predicted by the presented PES; contrary to other minima (e.g. m1/1′, m2/2′, m4, m8/8′, and m15). Note that the interpolated PES using the 2D mesh gives completely different energy values at these points, and, even worse, provides absolutely no hint about the existence of the subsurface minima (m5/m5′ and m16). We note that the inconsistency between actual forces and PES derivatives is directly related to the existence of these "hidden" minima (the above two cases).

Let us now analyze in more detail the potential energy of the N adatom in more details around one of these hidden minima points by constructing a cross section through the 3D-PES (Eq. (6.2)). To do so we have chosen the region between the minima m5 and m2. We have then mapped the potential energy of the N adatom in the cross sectional plane (P) located in a $(\bar{1}10)$ plane and passing through m5, shown in Fig. 6.6(a). At each mesh point in this plane the adatom is kept fixed and the surface atoms were allowed to relax. The resulting PES is shown in Fig. 6.6(b).

Fig. 6.6(b) reveals a highly non-trivial behavior of the adatom-surface interaction. This cross section in the 3D configuration space of the adatom allows us to directly access the results of

the conventional mapping of the 2D-PES, i.e., the results that are obtained when performing a vertical relaxation of the adatom for a lateral 2D mesh of points on the surface. When probing the 2D-PES along the line at the upper border of the plane P the N adatom will fall in the first local minima in the vertical direction towards the surface. These first vertical energy minima points are indicated in Fig. 6.6(b) and (c) by the black dashed lines A and B. The red dashed line C represents the second vertical energy minima. This part is hidden from the conventional lateral PES probing, i.e., the conventional 2D-PES mapping is not able to detect such parts and structures. Based on this discussion we can derive three conclusions from Fig. 6.6. First, the conventional mapping of the lateral 2D-PES results in a PES that is *discontinuous* in energy (and structure), as indicated by the gap between curves A and B. A small deviation of the adatom at the end of A along the $[\bar{1}\bar{1}0]$ direction will result in a huge and abrupt change in energy and structure, corresponding to the starting point of B. At this discontinuity there is a discrepancy in the forces acting on the adatom as indicated from the energy gradients. Such discontinuities explain the observed inconsistencies between the actual forces and the gradients of the 2D-PES. Second, in the conventional mapping scheme the subsurface binding site at m5 will be completely dismissed. However, when a full relaxation of the adatom is allowed at points near the A-B discontinuity it will directly fall into m5; which explains how these minima points were found in the first place. Finally, the PES in Fig. 6.6 shows one example of the effect of the N adatom initial vertical position on the obtained relaxed structure. For instance, if the adatom is initially placed at a position just above curve A and close to the A-B discontinuity it relaxes to the first local minimum, i.e., to a point on A; while placing it at a slightly lower vertical position it relaxes on the second (deeper) minimum located on C (simply following the gradients of the energy surface). This insight allows therefore to understand the nature of the adatom surface interaction, and allows to easily explain all of the difficulties encountered during the 2D-PES mapping with respect to the effect of the initial vertical position of the adatom, the forces-gradient disagreements, and the hidden or not accurately predicted minima points.

To demonstrate that the above discovered complex adatom surface interaction is indeed a consequence of the strong N-Ga and N-As bonds we compare the potential energy of the N adatom versus that of an As adatom. We consider a 1D path only (i.e. one line of the 2D-PES) and choose the line passing along the lower As-dimer (the trench dimer). In order to isolate the effect of surface relaxations we have also performed calculations with keeping surface atoms frozen (unrelaxed). In Fig. 6.7 we show the energy, adatom relaxed vertical position, and the maximum single relaxation of the surface as functions of the adatom position along the diffusion path and for the fully relaxed and frozen surfaces. Note that the surface atom relaxation is defined for each atom i as norm of the displacement vector of the specific surface atom from its equilibrium position in the clean surface, i.e.,

$$\Delta r_i = |\mathbf{r}_i - \mathbf{r}_i^{\text{clean}}|. \tag{6.5}$$

For the N adatom compared to the As adatom, the main results of Fig. 6.7 can be summarized as follows:

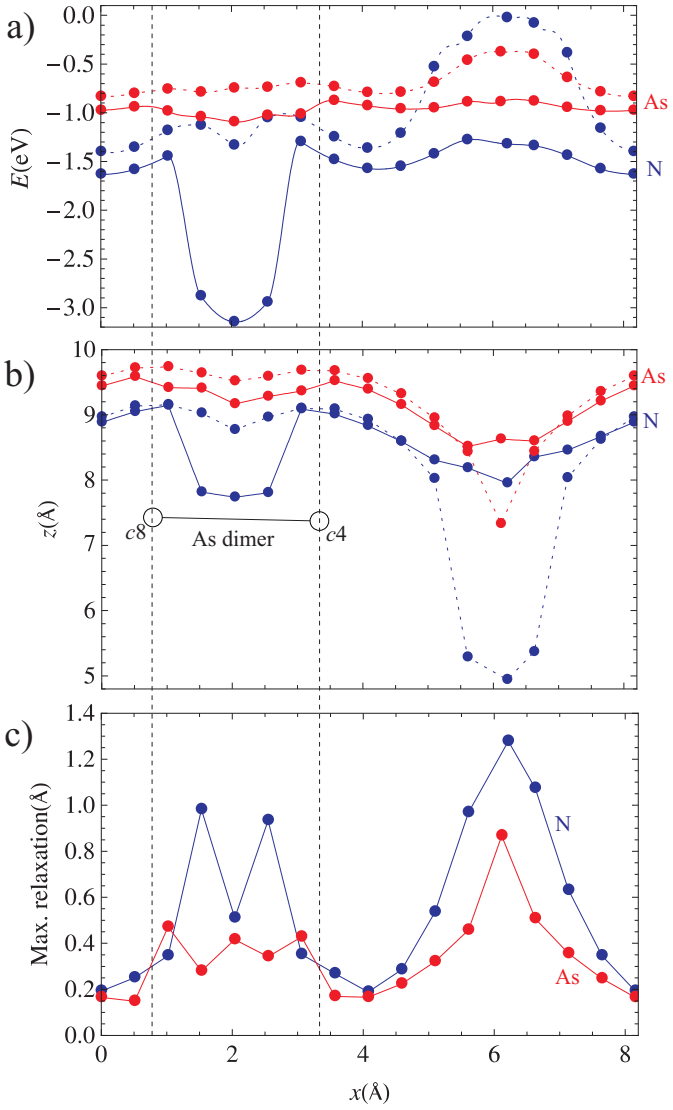


Figure 6.7: Binding energy (a), adatom relaxed vertical position/height (b), and the maximum single relaxation of surface atoms (c) for N (blue) and As (red) adatoms along a straight line in the [$\bar{1}$10] direction which is passing along the trench As dimer (labeled $c4$-$c8$ in Fig. 5.5(a)). For reference, the positions As dimer atoms are marked with the dashed vertical lines, and additionally by the empty circles in (b). The solid lines are used for the fully relaxed systems, and the dashed curves are used for the frozen (unrelaxed) surfaces. In (b) and (c) lines are drawn to guide the eye.

(i) Significantly larger and spatially more rapid energy variation as a function of adatom position.

(ii) Surface relaxation has a far more significant effect on the adatom energy and its relaxation.

(iii) The relaxation of surface atoms can be significantly larger.

For point (iii) The maximum value of surface atomic relaxations along this line is found to be 1.28 Å for the N adatom versus 0.88 Å for the As adatom. In fact, for the full 2D-PES of the N adatom shown in Fig. 6.3 we have identified a much larger value of the largest relaxation of surface atoms of 2.38 Å. These results reflect the strong interaction of the N adatom with the surface atoms and its capability in breaking and forming bonds confirms the large surface relaxations.

In conclusion, it becomes clear that the N adatom shows complex interactions with the GaAs surface which cannot be described in the 2D-PES picture. The adatom-surface interaction occurs in the multi-dimensional configurational space of the adatom and the surface atoms, resulting in

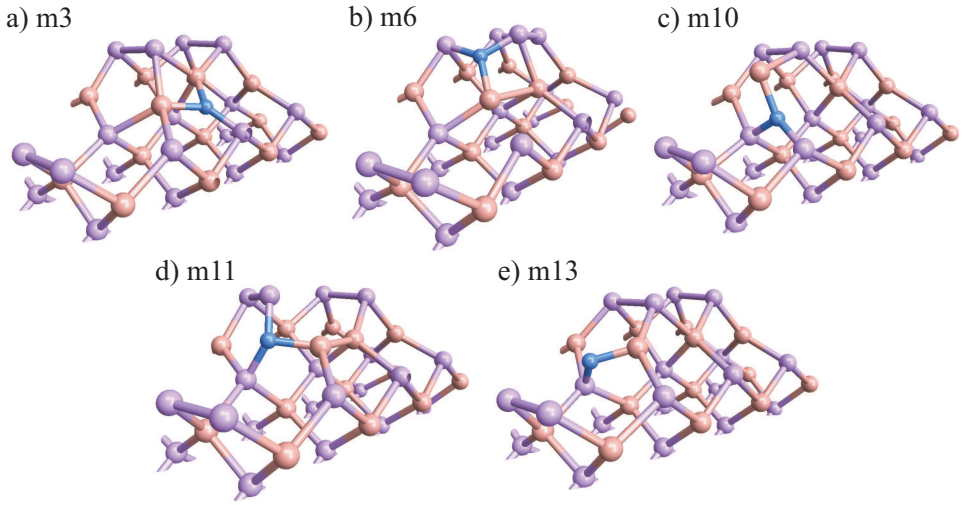


Figure 6.8: Atomic structures corresponding to the additional minima points found during the NEB calculations for the minimum energy paths.

many local minima in this space and hence in complex diffusion channels which are challenging to identify. This behavior is attributed mainly to the high reactivity of the N adatom, which is diffusing through or within the surface rather than on top of the surface, breaking and forming new bonds. The conventional 2D-PES mapping of the adatom therefore fails dramatically in this case. The potential energy of the N adatom in its 3D configurational space is therefore seen as a set of discontinuous multivalued 2D surfaces rather than a single continuous 2D-PES, which is due to the additional dependence of surface atoms configuration. Due to this it becomes very challenging to (i) identify the "complete" set of multivalued 2D surfaces, and (ii) to extract the necessary information on the actual adatom diffusion channels and their energy barriers. Moreover, this picture of the multivalued surfaces can be still incomplete due to the other degrees of freedom of the surface atoms, which should be also considered for a full description of adatom diffusion.

6.3 Calculating energy barriers

As discussed in the last subsection, due to the multi-dimensional adatom-surface interaction the identification of the complete adatom potential energy on the $\beta 2(2 \times 4)$ surface is highly challenging. Nevertheless, performing this step we have identified many subsurface binding sites for the N adatom. A possible option to proceed is therefore to estimate the energy barriers between these binding sites by employing the nudged elastic band (NEB) method [185]. In this method, the minimum energy path (MEP) is identified by constructing a series of "images" of the system between the initial and final states. Spring forces between the adjacent images are included to imitate an elastic band which is then optimized. The series of images are typically generated by employing a linear interpolation between initial and final states.

We note here that additional challenges appear in the present case in applying this method. Because of the local deformations of the surface around the N adatom, and of the general com-

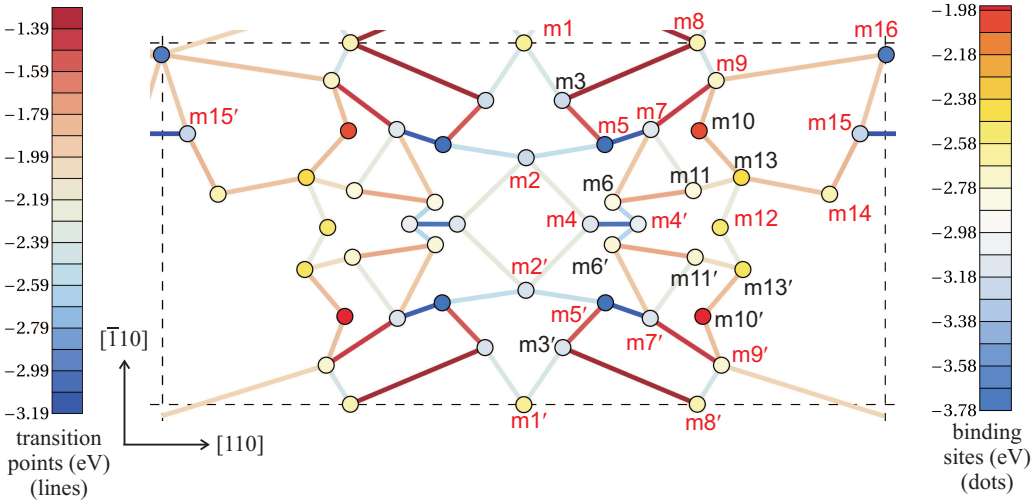


Figure 6.9: Top view of the binding sites (dots) and the transitions between them (lines) for a N adatom on the $\beta 2(2\times4)$ GaAs(001) surface (corresponding to Fig. 6.3). The energy values at the binding sites are represented by the color scale on the right hand side, and their exact values are presented in Table 6.1. For clarity, the shape of the minimum energy paths connecting each two binding sites and the exact positions of the transition state points are not shown, and instead represented by straight lines colored according to the value of the saddle point energy given by the left side color scale. The values of the transition state energies are given in Table 6.2.

plexity of the energy landscape (which depends also on surface atoms positions) setting the band of images needs additional care. It turned out that in most cases it is not suitable to generate the structures of the images by simply interpolating between the two end point. For instance, this applies when the structures at the two end points of a certain energy path differ significantly, or when the adatom's binding configurations are deep in the surface. A bad initial guess for one of the images makes it difficult for the algorithm to find the minimum energy path. Therefore, in some cases several trial calculations had to be performed first.

In order to calculate the minimum energy paths between minima points we have used the identified binding sites and constructed the network of possible adatom moves between them. The MEP's were calculated for each network line employing the NEB method. We have used 4 to 6 images for calculating each of the MEP's. During the optimization of some minimum energy paths *additional* energy minima between the starting and final points were identified. These minima points were *not* previously predicted within the conventional PES mapping due to the complexity of the adatom potential energy. These are the minima m3, m6, m10, m11, m13, and their primed counterparts, and are shown in Fig. 6.3, with their energy values in Table 6.1. The corresponding relaxed atomic structures for these points are shown in Fig. 6.8. Finally, it is important to note that due to the complexity of the adatom-surface interaction it can still be possible that the extended set of binding sites is yet not complete. An identification of the whole set of binding sites is not straightforward and computationally demanding. While it is not possible to quantify the consequences, it turns our that most of the diffusion barriers are indecisive for adatom diffusion, as will be shown later in this chapter (Sec. 6.4.1).

After including these additional minimum energy points and computing the diffusion barriers

Transition state	E (eV)	Transition state	E (eV)	Transition state	E (eV)
$T_{1,3}$	-2.40	$T_{7,9}$	-1.44	$T_{9,10}$	-1.87
$T_{3,5}$	-1.53	$T_{9,16}$	-1.89	$T_{10,13}$	-1.86
$T_{3,8}$	-1.29	$T_{9',16}$	-1.99	$T_{13,12}$	-2.24
$T_{2,5}$	-2.53	$T_{4,4'}$	-3.01	$T_{13,14}$	-1.89
$T_{2,4}$	-2.26	$T_{4',6}$	-2.59	$T_{14,15}$	-1.83
$T_{5,7}$	-3.11	$T_{6,11}$	-1.78	$T_{15,16}$	-1.92
$T_{7,11}$	-2.28	$T_{11,13}$	-2.04	$T_{15,15'}$	-3.19
$T_{7,6}$	-1.91	$T_{8,9}$	-2.44		

Table 6.2: Transition state energies between the binding sites of the N adatom on the $\beta 2(2\times 4)$ reconstruction of GaAs(001). The numbers in the subscripts give the labels of the binding sites (minima) such that the transition state from minimum m_i to m_j is denoted with $T_{i,j}$.

in between we have constructed a full reaction network containing of adatom single jumps. This network is shown in Fig. 6.9. As seen, there is a large number of adatom binding sites associated with local structure deformations, which makes it not trivial to determine the relevant complete network of adatom single jumps. For the assumed network the minimum energy paths were optimized within the NEB scheme and the saddle point energies were extracted. The values of the resulting saddle point energies are presented in Table 6.2. These values are used to to color code the corresponding network lines in Fig. 6.9. To reduce computational effort the minimum energy paths are calculated for a reduced set of minima points (restricting on the unprimed points in Fig. 6.3). The remaining set of paths are assumed to have similar saddle point energies due to the very similar structural topology.

6.4 Calculating activation barriers: Kinetic Monte-Carlo

Provided the network of binding and transition states of the N adatom on the $\beta 2(2\times 4)$ surface of GaAs(001) presented above, an efficient and effective approach to extract the effective activation diffusion barriers for this highly complex network are kinetic Monte-Carlo (KMC) simulations. A very brief introduction to the KMC algorithm is presented in Appendix A.

The KMC simulations are usually performed at several temperatures in the relevant range and for sufficiently long time. After that, the time-dependent mean square displacements are calculated and averaged over a sufficiently high number of independent KMC runs at the different temperatures. The obtained mean square displacements are expected to scale linearly with time t. Using this relation, the diffusion coefficients D in a certain *direction* (e.g. x-direction) can be extracted according to Einstein's diffusion equation:

$$D_x = \lim_{t\to\infty} \frac{\langle x^2(t)\rangle}{2t}, \tag{6.6}$$

where $\langle x^2\rangle$ is the mean square displacement in the x-direction. The diffusion is expected to satisfy the known Arrhenius relation:

$$D(T) = D_0\, e^{-E_A/k_B T}. \tag{6.7}$$

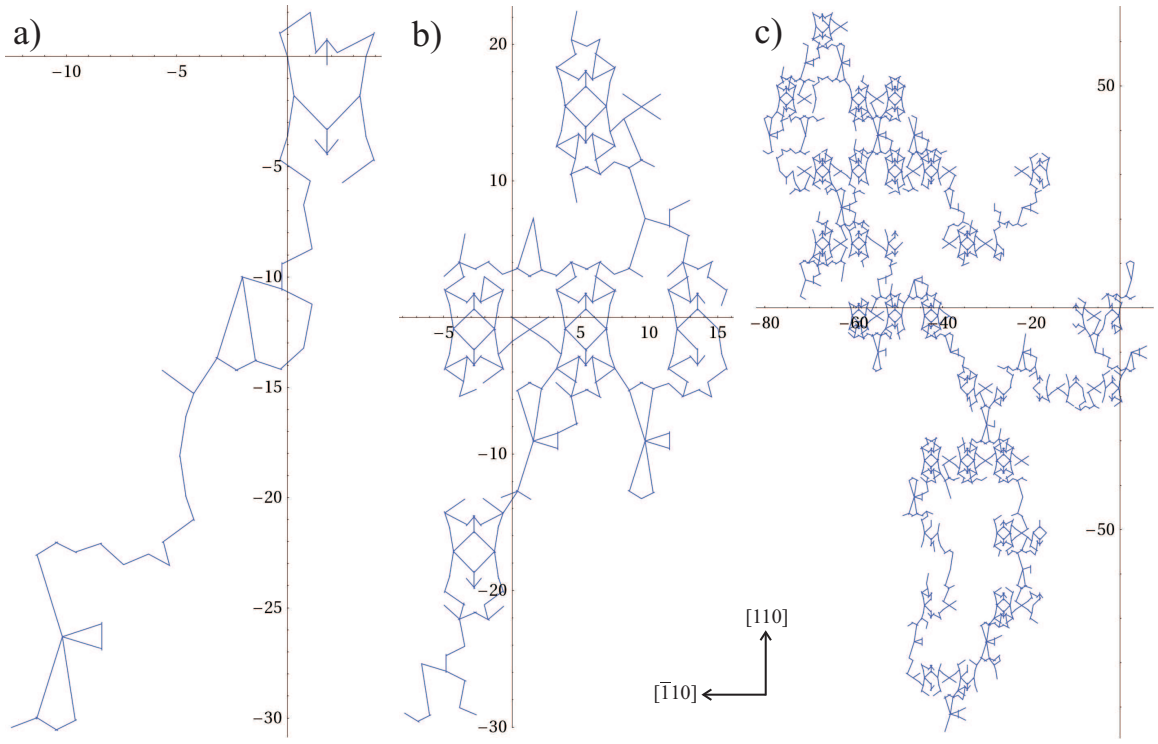


Figure 6.10: Snapshots of adatom trajectories in single KMC simulations running for 2.5×10^4 time steps at 3 different temperatures. (a) At $T = 1200\,^{\circ}\mathrm{C}$, (b) at $T = 2000\,^{\circ}\mathrm{C}$, and (c) at $T = 3000\,^{\circ}\mathrm{C}$. The axes are in Å. The network of adatom jumps in the single unit cell is defined in Fig. 6.9.

Here E_A is the activation energy barrier and D_0 is the prefactor, which can be calculated by plotting the diffusion coefficients against $1/T$.

We note that the KMC calculations require calculating the probabilities of the adatom to jump from the initial state to the neighboring states (see Appendix A), which are obtained through calculating the corresponding jumping rates. The rate of an adatom crossing the transition state from an initial state (site i) to a nearby final state (site j) is given according to the transition state theory [186] by

$$\Gamma_{ij} \;=\; \Gamma_0\, e^{-(E_{i,j}^{\mathrm{trans}} - E_i)/k_B T}, \tag{6.8}$$

where $E_{i,j}^{\mathrm{trans}}$ is the energy at the transition state point $\mathrm{T}_{i,j}$ and Γ_0 is the frequency of jump trials. The latter is calculated according to

$$\Gamma_0 \;=\; \frac{1}{2\pi}\frac{\prod_{n=1}^{N}\omega_n^i}{\prod_{n=1}^{N-1}\omega_n^{\mathrm{trans}}}, \tag{6.9}$$

where ω_n denote the vibrational frequencies. Since the rate equation depends exponentially on the energies it is very sensitive to small variations or uncertainties. Therefore, it is often enough to use an order of magnitude estimate of Γ_0. This is usually taken to be in the order of magnitude of maximum phonon frequencies (optical phonons) which are of $\sim 10^{13}\mathrm{s}^{-1}$.

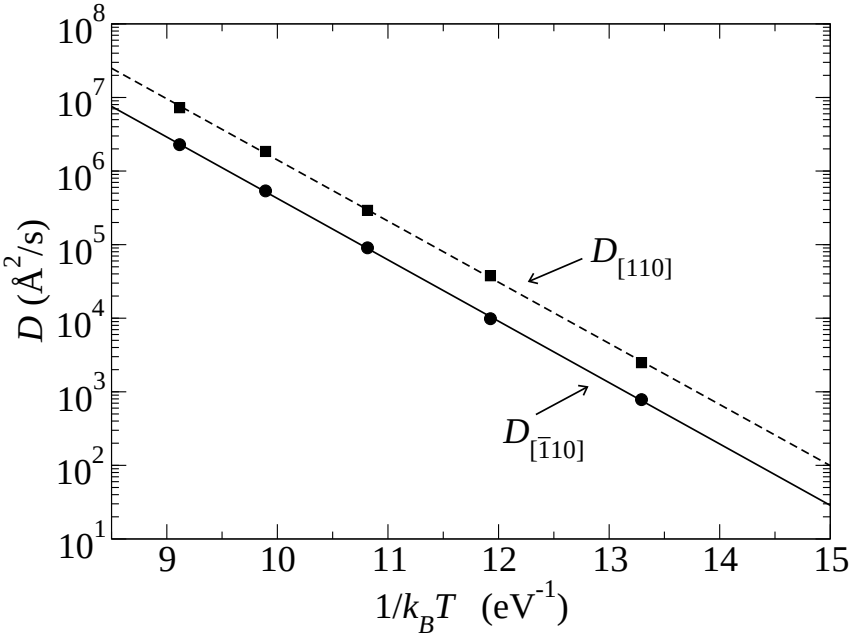


Figure 6.11: Diffusion coefficients obtained from KMC calculations along $[\bar{1}10]$ (dots) and the [110] directions (squares), plotted on a logarithmic scale against $1/k_BT$ and fitted according to the Arrhenius law. Note that $1/k_BT = 15.01\,\mathrm{eV}^{-1}$ corresponds to $T = 500\,^\circ\mathrm{C}$.

In order to extract the activation barriers in our system we performed the KMC simulations for temperatures in the range 600–1000 C°. The calculations were performed for a sufficiently large number time steps in the range of 1×10^5 to 3×10^6 steps. A number of a 2×10^3 to 2×10^4 independent KMC simulations was used to obtain the time averages. We note that the initial adatom positions were drawn randomly from a Boltzmann distribution of occupations according to the energy values of the minima. In addition, the calculations were allowed to equilibrate by running for sufficient steps (∼10 to 20 % of the total number of steps) before recording the adatom trajectories. Fig. 6.10 shows a few selected samples of the obtained adatom trajectories. The trajectories were attained in a single KMC run at temperatures of $T = 1200$, 2000, and 3000 °C; and after 2.5×10^4 time steps.

The obtained diffusion coefficients are plotted in Fig. 6.11 against $1/k_BT$ on a logarithmic scale. The data in this temperature window is found to be almost linear and can thus be fitted to an Arrhenius law. For the $[\bar{1}10]$ direction the fit gives a prefactor of $D_0 \approx 9.31 \times 10^{13}\,\text{Å}^2/\text{s}$ and an effective activation diffusion barrier of $E_A = 1.92\,\text{eV}$, and for the [110] direction it gives $D_0 \approx 2.88 \times 10^{14}\,\text{Å}^2/\text{s}$ and $E_A = 1.91\,\text{eV}$. Hence the adatom diffusion is almost isotropic. Comparing the N adatom activation energies to those calculated on the same surface for the Ga adatom of 1.5 eV and 1.2 eV and for the Al adatom of 1.6 eV and 1.3 eV, for the [110] and $[\bar{1}10]$ directions; respectively [111], we find that the values for the N adatom follows the trend in cohesive bond energies with values of 1.63 eV, 1.89 eV, and 2.24 eV per Ga-As, Al-As, and Ga-N bond, respectively.

6.4.1 Simplified kinetic model for understanding the KMC results

Although the adatom jumps network in Fig. 6.9 is complex, the following analysis provides a simple yet detailed understanding. Because the diffusion depends exponentially on the barrier energy we first assume that mainly the highest barrier in the minimum energy path in each direction determines the adatom diffusion. Therefore, by carefully examining the energy barriers, we first identify the adatom *minimum* energy paths in each of the lateral directions. The hence identified minimum energy paths are indicated by the light-green thick lines in Fig. 6.13. It turns out that the highest barrier along these paths and in both directions is at the transition state $T_{10,13}$ (between m10 and m13) and at the other symmetric/semi-symmetric counterparts. The configuration at the transition state $T_{10,13}$ is shown in Fig. 6.12. These transition states are indicated by the arrows in Fig. 6.13. The energy barrier between the lowest energy binding site m5 and this transition state reads $-1.86-(-3.77) = 1.91\,\text{eV}$. This gives an approximate estimation of the activation barrier in both directions. This value is in a surprisingly good agreement with the activation barriers obtained from the full KMC simulation. We can therefore conclude that the diffusion rates in both directions are mainly determined by a single energy barrier corresponding to a single barrier configuration, despite the complexity of the diffusion network.

It is straightforward to systematically improve the matching to the KMC results (also at artificially elevated temperatures) by including all the barriers that are close in energy. In the following we derive a basic model that includes a set of barriers to describe adatom diffusion and allows to extract the effective activation energies. Assuming thermodynamic equilibrium, the probability of the adatom occupying the binding site i is given by Boltzmann statistics

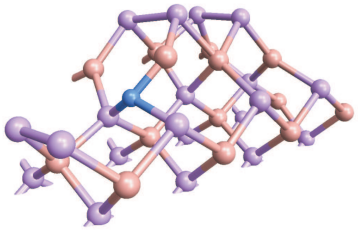


Figure 6.12: The atomic structure corresponding to the transition state $T_{10,13}$.

$$p_i = \frac{e^{-E_i/k_BT}}{\sum_k e^{-E_k/k_BT}}, \tag{6.10}$$

where E_i is the adatom binding energy at site i. In equilibrium, the adatom flow or "current" from i to j (denoted by I_{ij}) equals to that from j to i. This quantity is obtained by multiplying the probability of occupation of i, given by Eq. (6.10), by the escape rate given by the Eq. (6.8). This gives

$$\begin{aligned} I_{ij} &= \frac{e^{-E_i/k_BT}}{\sum_k e^{-E_k/k_BT}}\,\Gamma_0\, e^{-(E^{\text{trans}}_{i,j}-E_i)/k_BT} \\ &= \frac{\Gamma_0}{\sum_k e^{-E_k/k_BT}}\, e^{-E^{\text{trans}}_{i,j}/k_BT}. \end{aligned} \tag{6.11}$$

The first term is constant at a certain temperature. Hence, the adatom current is defined directly by the energy at the transition point relative to the total occupation probability. Note that a constant shift in the energies cancels out.

First, let us assume that only the transition point at $T_{10,13}$ determines the effective barrier.

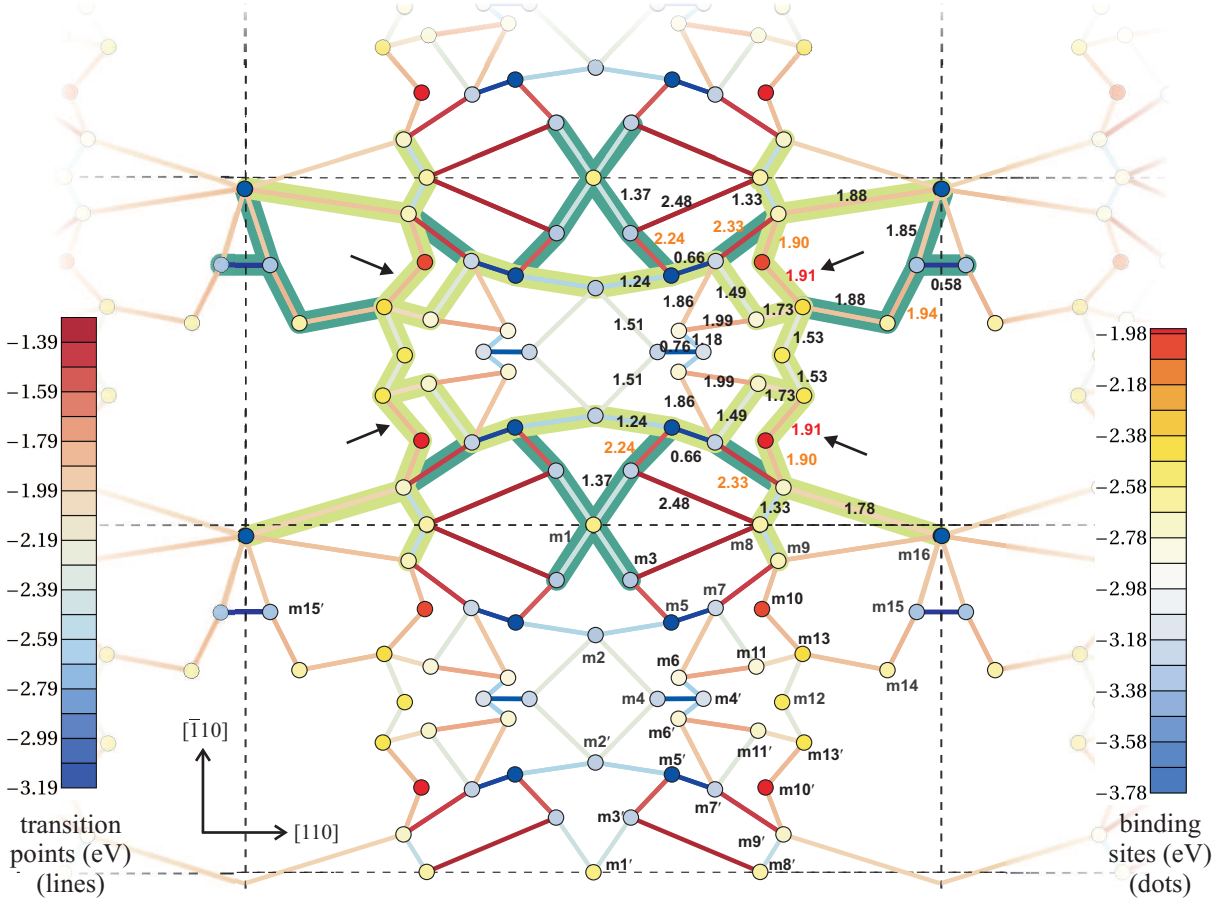


Figure 6.13: Adatom diffusion network indicating the primary minimum energy diffusion paths. The numbers show the barrier energies with respect to the lowest energy binding site m5, i.e., obtained by subtracting $E_{\mathrm{m5}} = -3.77\,\mathrm{eV}$ from the transition state energies (given in Table 6.2), or $E^{\mathrm{trans}}_{i,j} - E_{\mathrm{m5}}$. The right- and left-hand side color scales give the binding and transition energies respectively (see Fig. 6.9). The light green lines indicate the lowest minimum energy paths along both lateral directions. The highest energy barrier along these paths reads 1.91 eV as indicated with the red numbers and by the arrows. The darker green paths indicate next-lowest energy paths additionally included in the kinetic model. All the considered energy barriers in the kinetic model are indicated in orange (and red). The dashed lines mark the surface (2×4) unit cells.

Now, the diffusion coefficient can be calculated by multiplying the current (Eq. (6.11)), the square of the distance traveled by crossing this barrier once, and the number of channels with that barrier in each direction. The traveled distance equals $\frac{1}{2}\,a_{[\bar{1}10]}$ for the $[\bar{1}10]$ diffusion and $\frac{1}{2}\,a_{[110]}$ for the [110] diffusion (see Fig. 6.13). Thus we get

$$D_{[110]}(T) \;=\; 2\left(0.5\,a_{[110]}\right)^2 \frac{\Gamma_0}{\sum_k e^{-E_k/k_BT}}\; e^{-E^{\mathrm{trans}}_{10,13}/k_BT}, \tag{6.12}$$

and

$$D_{[\bar{1}10]}(T) \;=\; 2\left(0.5\,a_{[\bar{1}10]}\right)^2 \frac{\Gamma_0}{\sum_k e^{-E_k/k_BT}}\; e^{-E^{\mathrm{trans}}_{10,13}/k_BT}. \tag{6.13}$$

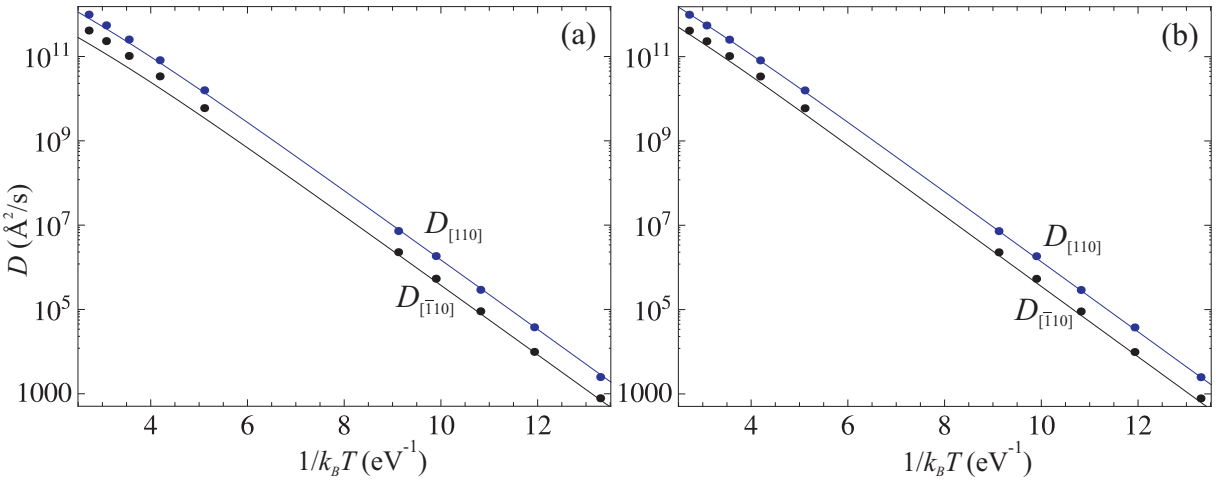


Figure 6.14: The lines indicate diffusion coefficients obtained by a simplified model using (a) Eqs. (6.12) and (6.13) that consider a single main diffusion path with a single barrier, and (b) Eqs. (6.14) and (6.15) which consider more diffusion paths and barriers. The dots indicate the diffusion constants obtained using KMC simulations, as presented previously.

Note that the factor 2 arrives from the fact that there are 2 diffusion channels in each direction. The above equations are plotted together with the diffusion coefficients obtained from the KMC simulations and shown in Fig. 6.14(a). To further extend the plot range, additional KMC results performed at artificially elevated temperatures (2000–4000 C°) have been included. As can be seen, already this very simple model gives results in very good agreement with the KMC simulations, particularly in the low temperature region. The difference in the lattice parameters of the surface unit cell in both directions gives a factor of 4 between the [110] and the [$\bar{1}$10] diffusion coefficients, which only affects the prefactors.

A closer inspection of Fig. 6.14(a) shows that this diffusion model starts to deviate from the KMC data with increasing the temperature. This indicates that additional diffusion channels start to be more probable at higher temperatures and hence have to be included in more accurate models. In order to do that we additionally include the next-lowest energy paths for adatom diffusion which have maximum barriers that are larger by $< 0.5\,\mathrm{eV}$ than for the lowest energy path. These are two additional paths, and they are indicated in Fig. 6.13 with the darker green thick lines. The value of the highest barrier along each of these paths is shown in an orange color in the figure, and it is taken to be the effective barrier for that path. In addition, as this modified model allows to include more than a single barrier, we have additionally included the barrier energy at $T_{9,10}$ which has a value very close to that at the $T_{10,13}$ (0.01 eV lower), which assists to improve the results also for the low temperature region.

In order to calculate the total diffusion coefficients we note that adatom diffusion flux on the surface can be regarded as equivalent to an electric current in a circuit. Each transition state has a temperature dependent "resistance" of a value $e^{E_{i,j}^{\mathrm{trans}}/k_BT}$ (compare to Eq. (6.11)). Recalling from simple circuit physics, we now sum the currents from the different paths to obtain the total current using a serial and parallel sum of resistances. Following the procedure in deriving

Eq. (6.12) and (6.13), we hence approximate the diffusion constants with (see also Fig. 6.13):

$$D_{[110]}(T) = \left(0.5\, a_{[110]}\right)^2 \frac{\Gamma_0}{\sum_k e^{-E_k/k_BT}} \Big[2\, e^{-E^{\mathrm{equiv,ser}}_{9,13}/k_BT} + 2\, e^{-E^{\mathrm{trans}}_{7,9}/k_BT} + e^{-E^{\mathrm{trans}}_{14,15}/k_BT}\Big], \tag{6.14}$$

and

$$D_{[\bar{1}10]}(T) = \left(0.5\, a_{[\bar{1}10]}\right)^2 \frac{\Gamma_0}{\sum_k e^{-E_k/k_BT}} \Big[2\, e^{-E^{\mathrm{equiv,ser}}_{9,13}/k_BT} + 2\, e^{-E^{\mathrm{trans}}_{7,9}/k_BT} + e^{-E^{\mathrm{trans}}_{14,15}/k_BT} + 2\, e^{-E^{\mathrm{trans}}_{3,5}/k_BT}\Big]. \tag{6.15}$$

Here $e^{E^{\mathrm{equiv,ser}}_{9,13}/k_BT} = e^{E^{\mathrm{trans}}_{9,10}/k_BT} + e^{E^{\mathrm{trans}}_{10,13}/k_BT}$ is the equivalent resistance for the transition between m9 and m13, obtained using a serial sum. The other remaining terms between the square brackets in Eqs. (6.14) and (6.13) come from summing the parallel currents. Again, in analogy to an electrical circuit we obtain: $1/e^{E^{\mathrm{equiv,par}}/k_BT} = 1/e^{E^{\mathrm{trans}}_{i,j}/k_BT} + 1/e^{E^{\mathrm{trans}}_{k,l}/k_BT} + \cdots$.

Fig. 6.14(b) shows the calculated diffusion coefficients using Eqs. (6.14) and (6.15) together with the KMC data. As can be seen, including the additional channels significantly improves the results in the high temperature region, leading to an excellent agreement to KMC simulations in the whole temperature range (600–4000 °C). While the agreement in the high temperatures is physically completely irrelevant, including the additional channels (Eqs. (6.14) and (6.15)) has additionally improved the agreement with the KMC data in the lower temperature region. For instance, using our formulas to generate the diffusion constants for the the low temperatures considered in the KMC calculations (600–1000 °C) and fitting these values using Arrhenius law gives $D_0 = 7.77 \times 10^{13}\,\text{Å}^2/\text{s}$ and an effective activation diffusion barrier of $E_A = 1.92\,\text{eV}$ for the $[\bar{1}10]$ direction, and $D_0 = 2.73 \times 10^{14}\,\text{Å}^2/\text{s}$ and $E_A = 1.91\,\text{eV}$ for the [110] direction, in excellent agreement with the fitting of the KMC results presented in Sec. 6.4.

Finally, we conclude from Fig. 6.9 that the diffusion barriers around the As-dimers are generally much lower than the effective barrier for the lateral transport. This indicates that a N adatom spends a long time close to these regions before making the jump to one of neighboring surface cells. At typical growth temperatures of the ternary $GaAs_{1-x}N_x$ system of ∼500 C° and a growth rate of ∼ 1 monolayer per second, the N adatoms are *not* able to fully equilibrate due to the very small diffusion constants of order 10–100 $\text{Å}^2/\text{s}$, as indicated from Fig. 6.11. However, a limited local equilibration is realized. For these small adatom diffusion constants other kinetic mechanisms (e.g. N substitutions) can become important and will be discussed in the next section.

6.5 Subsurface incorporation of N: Two mechanisms

The incorporation of N in the subsurface region during growth is of great interest since it results in a strongly increased N solubility compared to the bulk solubility, as we have shown in the previous chapter. Although previous studies has assumed that the subsurface incorporation mechanism is functional [50] to explain the experimentally measured N concentrations, the actual mechanisms

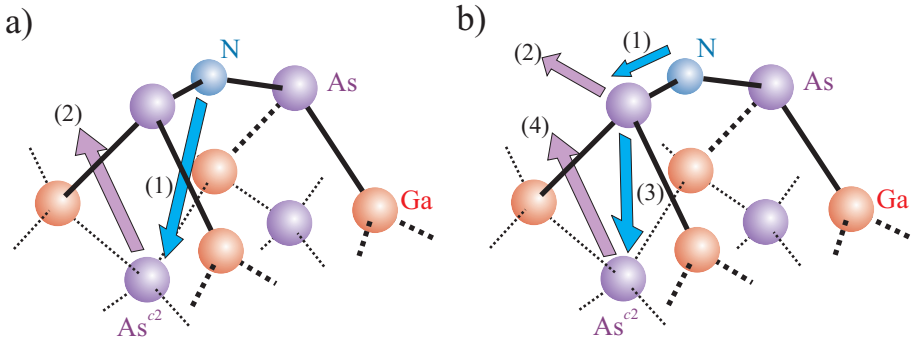


Figure 6.15: Schematic representation of the considered subsurface incorporation mechanisms of N, initially bound at an As-dimer. (a) The direct subsurface incorporation. (b) The indirect subsurface incorporation. The numbers indicate the sequence.

and the associated kinetic barriers for this incorporation process are not identified so far. In this section we therefore investigate the possible incorporation process of N in the subsurface layer of the $\beta 2(2 \times 4)$ GaAs(001) surface reconstruction and obtain an approximate estimate of the associated activation barriers, which allows to draw conclusions regarding their occurrence.

We shall concentrate on the incorporation of N in the energetically *most favorable* subsurface layer (c-layer) substitutional site, which is the $c2$ site (recall the used notations in Fig. 5.5). As we have shown in the previous chapter, the occupation of this site would allow to easily explain the high N concentrations obtained in experiments.

In order to study the incorporation mechanisms and reaction paths we first identify the initial interaction steps for the incoming N adatom towards this subsurface site. This is done by making use of the calculated binding sites of the N adatom discussed earlier in this chapter. Among the most interesting low energy binding sites for the incoming N adatoms are the short bridge sites, i.e., the sites in the center of As-dimer atoms at the topmost layer. These are the m4/4′ sites according to our earlier notations in Fig. 6.3 and Fig. 6.5(c). For each of these sites the incoming N adatom directly breaks the As-As dimer bond (without a barrier), forming two new directional bonds with the underlying As atoms. The relaxed structure with the N adatom bound at one of these binding sites, namely the m4 site, is also shown in the upper panel of Fig. 6.16(a). These binding sites are also interesting because they are close to the most favorable subsurface substitutional position; the $c2$ position. In addition, the structure just below them is relatively open, which makes them very reasonable starting points for subsurface incorporation and allows to simplify the problem.

Starting from the binding configuration at an As-dimer center, we identify the possible reaction paths for the N atom. Specifically, we study two possible mechanisms that may lead to its incorporation in the $c2$ position in the subsurface. These mechanisms are: (i) a direct subsurface substitution, and (ii) an indirect subsurface substitution. In the first process, the N atom directly diffuses into the surface and substitutes the As atom at the subsurface position (As^{c2}) by a kick-out mechanism. In the second process, the N atom kicks-out a neighboring As-dimer atom and gets incorporated in the topmost surface layer, before diffusing into the subsurface to substitute the subsurface As^{c2} atom. A schematic representation of these mechanisms is shown

in Fig. 6.15. These mechanisms will be discussed in the following.

6.5.1 Direct subsurface N incorporation

Starting from the short bridge binding site (at the As dimer center), one possible mechanism for N atom to be incorporated at the $c2$ subsurface position is by directly diffusing towards the subsurface region to replace the As atom at that site. In order to simplify estimating the overall energy barrier for this reaction path we split it into two successive processes:

1. The diffusion of the N atom in the normal/vertical direction into the surface to a subsurface interstitial region (close to the $c2$ site), followed by:

2. The kick-out substitution of the subsurface As^{c2} atom, which diffuses to outside the surface.

These processes are analyzed in the following.

1. The diffusion of the N atom to the subsurface layer:

Here we discuss the first part of this reaction path which is the diffusion of the N atom in the vertical direction towards the interstitial region close to the subsurface substitutional site. In order to calculate the energy barrier for this process one in principle needs to map the full configuration space of the adatom-surface. This is however computationally expensive and time consuming. To simplify the problem we restrict ourselves to a cross-sectional plane in the 3D configuration space of the N atom. The chosen plane is going from the first layer to the subsurface layer and passing through the center of the a-layer As dimer (i.e. through the initial binding site). This plane is shown in Fig. 6.16(a).

The resulting PES of the N atom in this plane is shown in the lower panel of Fig. 6.16(a). The lowest energy configuration is at the point labeled A, and it corresponds to the binding configuration at the As-dimer center. Since we are interested in the diffusion close to the $c2$ subsurface substitutional site, we calculate the energy barrier for the N atom to reach the point in the lowermost left corner on the PES; labeled B'. This point is the closest interstitial point to the subsurface site (in the same same plane). The minimum energy path from A to B' is schematically shown by the dashed arrow. The barrier energy along this path reads $\sim$1.6 eV.

We note that at the interstitial position B' the N atom is bound to the As^{c2} atom, which is displaced from its original place to an off-plane position (the symmetry is broken).

2. Kick-out and replacement of the subsurface As^{c2} atom:

The next step to fully accomplish the subsurface incorporation is the kick-out and the replacement of the As^{c2} atom. Therefore, to obtain the barrier for this part we map the PES of the As^{c2} atom after the diffusion of the N atom to the interstitial position. For the sake of simplicity we also sample a 2D cross sectional PES located in the same vertical plane. Recall that for the starting configuration (B') the As^{c2} lies outside this plane. Therefore, we have first displaced the As^{c2} atom back to an in-plane position (same plane in which the N atom is located) and relaxed the

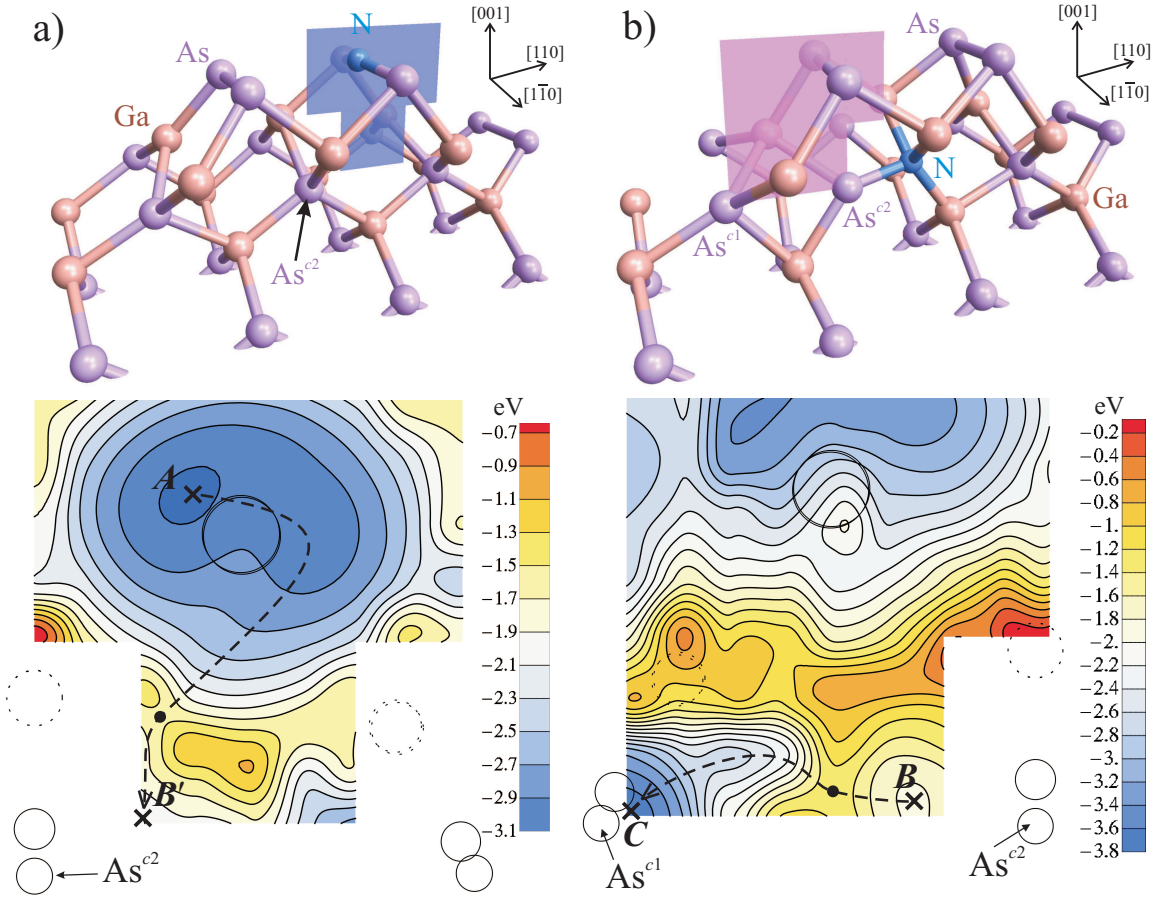


Figure 6.16: Reaction paths and barriers for the first (a) and second (b) parts of the direct subsurface incorporation of N. The upper panels show the cross-sectional planes for which the PES was sampled for (a) the N atom bound at the As-dimer center, and for (b) the As^{c2} atom after the diffusion of the N atom to a subsurface interstitial position. The lower panels show the corresponding PES's. The shown structures correspond to the minimum at point A in (a), and to the minimum B in (b). The dashed arrows indicate the relevant minimum energy paths (approximate) with the saddle points indicated by the dots. For reference, the positions of Ga (As) atoms of the *clean* surface are projected on the 2D planes and are represented by circles with dotted (solid) borders, with their size inversely scaled with their distance from the surface.

structure with the As^{c2} allowed to relax in-plane only, keeping the N fixed. The resulting new position of the As^{c2} defines the starting point to sample its PES. This configuration is found to be 0.45 eV higher in energy than the one with the fully relaxed As^{c2} (the B' configuration). The new configuration and the plane chosen to sample the PES of the As^{c2} atom are shown in the upper panel of Fig. 6.16(b).

The lower panel of Fig. 6.16(b) shows the resulting PES. The next minimum for the As^{c2} atom is located at B, at which the N atom takes an off-plane position. As can been, the lowest energy barrier for the As^{c2} atom to be removed from its interstitial site is located along the path connecting the minimum at the point B to the point C. The energy barrier along this path reads 0.54 eV. For the configuration at C the N atom completely occupies the $c2$ position, with the As^{c2} atom almost substituting the As^{c1} atom originally located close to that position (as marked

on the PES).

The final step of the reaction is the migration of the As^{c1} atom on the surface to an As reservoir. However, it is not relevant here to calculate the energy barriers for this process since already before this step the N has completely replaced the As^{c2} atom. However, in order to obtain an estimate of the energy of the system after this process we assume that the chemical potential of the As reservoir is at the mid-point of the thermodynamically allowed range of μ_{As}, denoted as $\mu_{As}^{1/2}$. Note that at this chemical potential the considered surface reconstruction ($\beta 2(2\times 4)$) is still stable.

Summary

Finally, for an overview of the full mechanism for the direct subsurface incorporation of N, we summarize the calculated energetics in the reaction path in Fig. 6.17(a) (Mechanism I). Note that the notations in the reaction coordinates (horizontal axes) correspond to the same configurations described above. The dashed parts of the curve imply that the actual energies for these parts were not calculated and drawn only to guide to the eye. These parts correspond to the energy barrier to change the configuration at B' to the configuration at B, and the migration barrier of the As^{c1} atom on the surface. Assuming that the former barrier is small and can be ignored, the energy barrier from B' to C (ignoring B' to B) reads 0.84 eV. Taking into account only the barriers until C, since at this point N is already incorporated in the subsurface site, the overall activation barrier for this reaction path reads $\Delta E = 1.6 + 0.84 = 2.44\,\text{eV}$.

A rough estimation of the rate of this reaction can be made using $\Gamma_\circ e^{-\Delta E/k_B T}$ with $T = 500\,^\circ\text{C}$ and assuming a typical prefactor value of $\Gamma_\circ = 10^{13}\,\text{s}^{-1}$. This gives a rate of order $10^{-2}\,\text{s}^{-1}$. However, since this is not essentially the minimum energy path for this direct subsurface substitution process, the actual rate for this event is possibly higher. This indicates that this reaction path can still be possible at typical growth rates in MBE. However, since this is not the only possible reaction path for N to be incorporated in the subsurface position $c2$, other possible reaction paths should be considered as well.

6.5.2 Indirect subsurface N incorporation: A trapping mechanism

As another possible reaction mechanism for subsurface incorporation of N we have considered an indirect incorporation mechanism. In this mechanism the N atom, bound at the As-dimer center (short bridge site), first substitutes an As atom in the topmost surface layer before diffusing and substituting the subsurface As^{c2} atom. To study the associated energetics of this mechanism we split it into two successive processes:

1. The substitution of an As atom at the first surface layer.
2. The diffusing of the surface-incorporated N atom to the subsurface to kick-out and replace the subsurface layer As^{c2} atom.

In the first process, the N atom bound at the surface is closest to the neighboring As dimer atoms. Hence, a potential reaction path is the diffusion of one of the As dimer atoms in a kick-

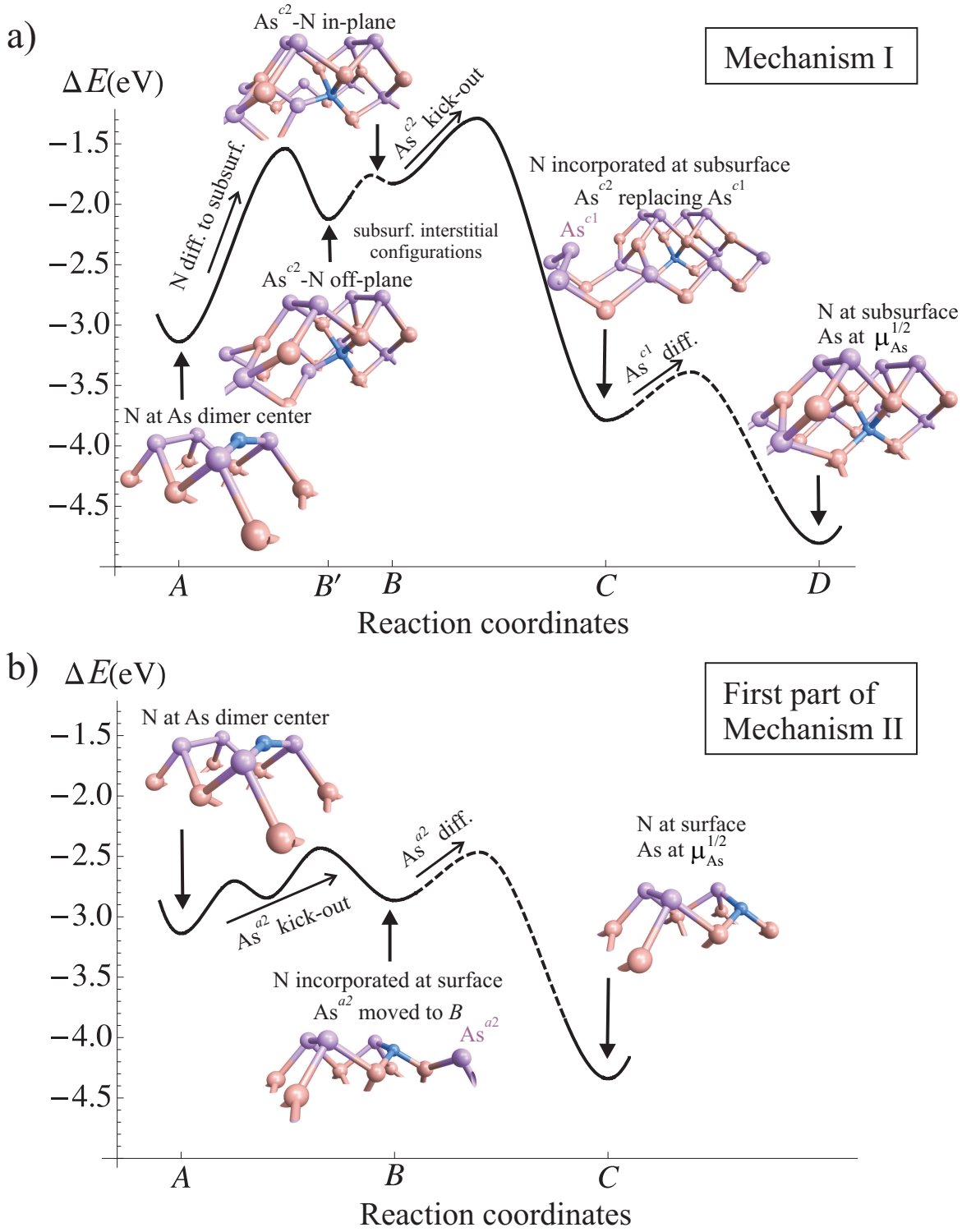


Figure 6.17: Summary of the kinetic barriers for the proposed reaction mechanisms for N incorporation (a) at the subsurface in a direct process and (b) at the topmost surface layer during the indirect subsurface incorporation. The dashed parts of the curves imply that the actual energies for these regions were not calculated explicitly, and hence they are drawn only to guide the eye. Note that the reaction coordinates labeled on the horizontal axes correspond to the points indicated in Figs. 6.16(a) and (b) for Mechanism I, and Fig. 6.16 for Mechanism II.

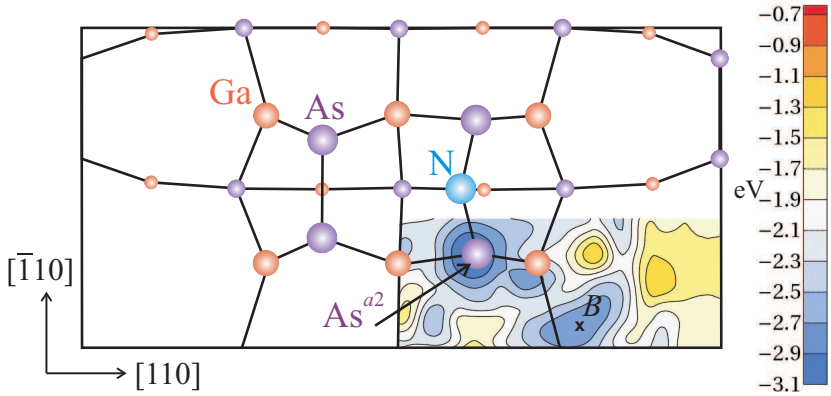


Figure 6.18: Top view of the PES of the As^{a2} atom with the N adatom bound at the m4 minimum position (see Fig. 6.3) as shown. The underlying surface atoms are shown as spheres as indicated. This structure corresponds to the one shown in Fig. 6.16(a).

out process, to be substituted by the N atom. In order to calculate the activation barrier of this kick-out process we map the PES of the As atom at the $a2$ site, denoted As^{a2} (recall notations in Fig. 5.5), with the N atom bound at the As-dimer center position. The PES has been mapped using a rectangular mesh that covers the relevant area for the diffusion of the As^{a2} atom. The relevant area was identified after performing preliminary calculations on a coarser mesh covering the whole cell and using a reduced cell size (2×4 instead of the 4×4 used for the more accurate calculations).

The resulting PES is shown in Fig. 6.18. As can be seen from the figure, the diffusion barrier of the As^{a2} atom is relatively low. After the As^{a2} atom moves from its initial position to another binding site on the surface, e.g. to the point B in Fig. 6.18, the N atom occupies its position (i.e., substitutes it) by simple relaxation. Along the minimum energy path towards B, the As^{a2} atom has to overcome two barriers. The total activation barrier for this process is hence $0.44 + 0.41 = 0.85\,\text{eV}$. After that, the As^{a2} atom diffuses on the surface to an As reservoir. As in the previous incorporation mechanism (Sec. 6.5.1), the actual barrier for this part was not calculated, and to estimate the final energy we take the chemical potential at the As reservoir to be the mid-point value of the allowed range ($\mu_{\text{As}}^{1/2}$).

Note that if the As atom would return back it will not occupy its original position by replacing the N atom, but it will rather bind at another position which is energetically more favorable. The new position is directly above the N atom, and it corresponds to the configuration at the minimum m5 (see Fig. 6.3 and Fig. 6.5(d))

The relevant local minima and barriers for this part of the energy path are summarized in Fig. 6.17(b) (Mechanism II). The low energy barrier (0.85 eV) required for the incorporation of N atom at the topmost surface layer is significantly lower than the barrier required for the N atom to move directly to the subsurface replacing the As^{c2} atom. A rough estimate of the reaction rate at $T = 500\,^{\circ}\text{C}$ (as before) gives $\sim 10^7\,\text{s}^{-1}$, which is about 9 orders of magnitude *higher* than the rate of the direct subsurface incorporation. Therefore, this reaction path is much more dominant; an incoming N adatom interacting with the surface As dimers has a much

higher probability to kick-out a surface As dimer atom to substitute it in the first place. We therefore conclude that N adatoms are much more easily incorporated in the first surface layer than directly in the subsurface.

The second process in the indirect subsurface incorporation mechanism is the diffusion of the N atom incorporated in the first surface layer vertically towards the subsurface to replace the As^{c2} atom. Rather than explicitly mapping the full path for this reaction we perform a simple estimation: This reaction is in principle similar to the case of the direct subsurface incorporation discussed above (Sec. 6.5.1, and "Mechanism I" in Fig. 6.17(a)). The barrier for the actual path is likely to be even *higher* since the N atom substituted at $a2$ has more covalent bonds with surface atoms. We therefore assume that the barrier is $> 2.44\,\text{eV}$.

The overall barrier for the full indirect subsurface incorporation mechanism is therefore expected to be $> 2.44 + 0.85\,\text{eV}$, resulting in an extremely low reaction rate of $< 10^{-9}\,\text{s}^{-1}$. Hence, the incoming N adatom is easily incorporated and *trapped* at the As^{a2} dimer surface site. The first surface layer N incorporation therefore acts as a trapping mechanism which suppresses the direct diffusion of the N adatom to subsurface, and hence prohibiting subsurface incorporation.

6.6 Surface kinetics combined: The incorporation of N

With the above results it becomes possible to establish a full picture for N diffusion and incorporation (substituting As atoms) under practical MBE growth conditions. To do so we additionally include the incorporation mechanisms as additional channels in our network for surface diffusion discussed in Sec. 6.4 and shown in Fig. 6.9. The additional channels are the direct subsurface incorporation channels (Sec. 6.5.1) and the direct surface incorporation channels (first part of the indirect incorporation channel, Sec. 6.5.2), and they emerge from the binding sites at the centers of As-dimers. For the sake of simplicity we have treated the successive processes involved in each of the As-N substitution processes as a single one with the same overall barrier. After that, we have performed Kinetic Monte-Carlo simulations as described in Sec. 6.4 for a growth temperature of 500 °C. In specific, we have recorded the time required for the N atom to be incorporated and the associated overall displacement, for 700 independent KMC runs (i.e. 700 N atoms). Each of the KMC calculations has required up to $\sim 10^9$ time steps to achieve N incorporation.

Our KMC results revealed that none of the N atoms was able to incorporate in the subsurface substitutional site, and all of them were incorporated at the topmost surface layer substitutional site. This observation is expected from the previously estimated 9 orders of magnitude lower rates for subsurface incorporation compared to surface incorporation (see Sec. 6.5.2), which makes it not possible to obtain a statistical sample for subsurface incorporation employing KMC calculations in a realistic time. However, subsurface incorporation of N is realized only when assuming very high temperatures. This shows the effectiveness of the surface trapping mechanism in prohibiting the subsurface incorporation in practice.

Fig. 6.19 shows the obtained required time for all of the N atoms to get incorporated at the surface. Using these results the average incorporation time is 1.9 ms. This allows us to conclude

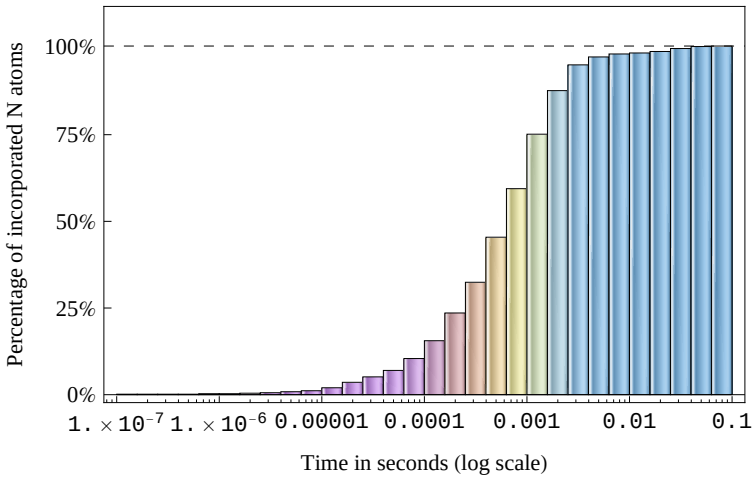


Figure 6.19: Incorporation time of N atoms at GaAs(001) $\beta 2(2 \times 4)$ surface calculated using KMC simulations for 700 N atoms at $T = 500\,^{\circ}\mathrm{C}$. None of the N atoms was able to incorporate in the subsurface and all atoms were incorporated at the topmost surface layer.

that for typical MBE growth rates/conditions the N adatoms have sufficient time to achieve substitutional incorporation in the topmost surface layer. On the other hand, the calculated average displacement of only 3.6 Å indicates a very limited diffusion of N adatoms; considering that the length of the 1×1 surface cell edge is ~ 4.08 Å. These results support our conclusions that N atoms are *unable* to achieve thermodynamic equilibrium on the surface, and only a local equilibration can be achieved.

6.7 Summary and conclusions

The diffusion of N adatoms on GaAs(001) surface is found to be highly different from the established behavior of adatoms on semiconductor surfaces. The fundamental origin of the unusual behavior is the strong interaction of the N adatom with the surface. This interaction is realized by the surface-adatom bonds that are substantially stronger than the normal surface-surface bonds, which results in breaking and forming new bonds and in large relaxations of surface atoms from their equilibrium positions in the clean surface up to few Angstroms. The adatom-surface system considered here may be thus regarded as a prototype for the surface diffusion in other similar bond-strength mismatched alloys.

As a consequence, the actual potential energy profile for the N adatom is complex and multi-dimensional (multi-valued) in nature and hence cannot be mapped onto the two dimensional (lateral) configuration space of the adatom. This leads to the dramatic failure of the concept of the potential energy surface for the N adatom in identifying its binding sites and diffusion paths.

Employing the NEB method we were able to identify an extended set of adatom biding sites and to calculate the minimum energy paths. By constructing the adatom diffusion network and employing KMC simulations we extracted diffusion constants and activation barriers. The N diffusion is found to be almost isotropic, with an effective activation barrier of $\sim$1.9 eV. A further analysis through a simplified diffusion model showed that the activation barrier is mainly

determined by the highest barrier along the minimum energy path responsible for the migration of the adatom to neighboring surface cells. The small diffusion constants indicate that the N adatoms are rather localized at the typical growth temperatures and the equilibration is possible only locally because the barriers are lower *inside* the surface unit cell.

In the next step we have investigated possible mechanisms for the subsurface incorporation of N, which is appealing to significantly enhance their solubility as shown in the previous chapter. Our calculated energy barriers for these mechanisms indicate that the N adatoms are easily incorporated in the *topmost* surface layer by substituting an As dimer atom. This incorporation acts as trapping mechanism that prohibits the subsequent incorporation of N in the subsurface.

In order to see how N incorporation kinetics work in practice we have performed KMC simulations combining the surface and subsurface incorporation channels with adatom surface diffusion channels. The results confirmed that only the incorporation in the topmost surface layer can be realized in practice. Furthermore, despite their very limited diffusion, the N atoms are able to achieve full substitutional incorporation under typical MBE growth rates.

These results provide for the first time a detailed understanding of the kinetic mechanisms that control the surface diffusion and the incorporation of N adatoms during the growth of ternary $GaAs_{1-x}N_x$ systems, and thus a detailed understanding of the mechanisms limiting eventually N solubility.

Chapter 7

The Interplay Between Thermodynamics and Kinetics: Consequences on Compositional Correlations and on N Solubility in Dilute Nitride Alloys

7.1 Introduction

In the previous chapters we have shown that while the subsurface incorporation of N is thermodynamically significantly more favorable, its occupation is prohibited due to large activation barriers which cannot be overcome under typical MBE growth conditions. The N adatoms are rather incorporated and trapped at the topmost surface layer. This has significant consequences on the growth and properties of dilute nitride alloys.

In this chapter we consider two cases for which the delicate interplay between thermodynamics and surface kinetics has crucial consequences. In particular, in the first case we focus on the compositional anticorrelation between In and N atoms in the quaternary $In_xGa_{1-x}As_{1-y}N_y$ dilute nitride alloys, which has been experimentally observed but not understood. Based on our theoretical analysis we are able to identify the origin of this compositional anticorrelation, which allows to experimentally control it. In the second case we focus on N solubility in GaAs and InAs and the role of surface kinetics. In specific we discuss the theoretically predicted maximal N solubilities resulting from the various incorporation mechanisms (surface, subsurface, bulk), in comparison with experiment. Based on this analysis we are able to identify the mechanism that controls N solubility in practice and thus to explain experimental results.

7.2 Compositional correlation and anticorrelation in quaternary $In_xGa_{1-x}As_{1-y}N_y$ alloys

Due to the large lattice mismatch between GaAs and GaN, increasing the concentration of N in $GaAs_{1-x}N_x$ alloys may result in clustering, phase separation, or spinodal decomposition [23, 24]. As a consequence, drastic degradation of material quality is observed. As a promising approach to overcome these problems Kondow et al. [21, 22] proposed the growth of quaternary $In_xGa_{1-x}As_{1-y}N_y$ alloys. The incorporation of In allows for the compensation of the tensile strain induced by the N atoms, promising an enhanced N solubility and material stability. A further benefit in growing quaternary alloys is that they can be tuned to be lattice matched to GaAs.

A specific problem in the growth of these quaternary materials is the existence of compositional fluctuations. For the quaternary $In_xGa_{1-x}As_{1-y}N_y$ alloy, four types of nearest-neighbor bonds are possible (In-N, Ga-N, In-As, and Ga-As). Hence, simply providing the macroscopic composition (given by x and y) is *not* sufficient to uniquely determine the number of these bonds, contrary to the case of ternary material systems. For instance, if the the formation of In-N bonds *in the system* is more favorable than Ga-N bonds, N atoms would tend to go into In-rich regions where they can predominantly form In-N bonds, while in the opposite case of more favorable Ga-N bonds N atoms would tend to segregate into In-deficient regions. The first (second) scenario will lead to compositional correlation (anticorrelation) between In and N. The existence of larger degrees of freedom for the short range ordering (i.e., the spatial arrangement of elements and bonds) strongly affects the physical properties of these quaternary alloys [31, 32]

Experimentally, compositional fluctuations in InGaAsN as a function of growth temperature and thermal annealing and its effect on optical properties has been intensively studied over the last few years. The results can be summarized as follows: Post-growth annealing leads to preferred formation of In-N bonds [37] and In-N spatial correlation [29]. On the other hand, Kong et al. [28] reported a preferred formation of Ga-N and In-As bonds at moderate growth temperatures of 430 °C (i.e. in our terminology In-N anticorrelation). Further, an increased N-concentration at the GaAs/InGaAsN interface has been observed by several groups [29, 30] and can be again regarded as In-N anticorrelation. The mechanisms behind the anticorrelation, however, have not been identified so far.

Therefore, identifying compositional correlation/anticorrelation and understanding the underlying physical principles is essential to eventually technologically employ it. Hence, in this chapter, we identify the atomic mechanisms that can lead to this behavior based on *ab-initio* calculations and by utilizing the previously calculated surface thermodynamics and kinetics mechanisms.

7.2.1 Identification of the In-N anticorrelation: Experimental findings

The In-N anticorrelation behavior in quantum wells grown by MBE can be clearly identified in recent experimental data by Albrecht et al. [26], presented in Fig. 7.1. In this experiment, a stack of three single GaAsN, InGaAsN and InGaAs quantum wells has been grown at 410 °C by

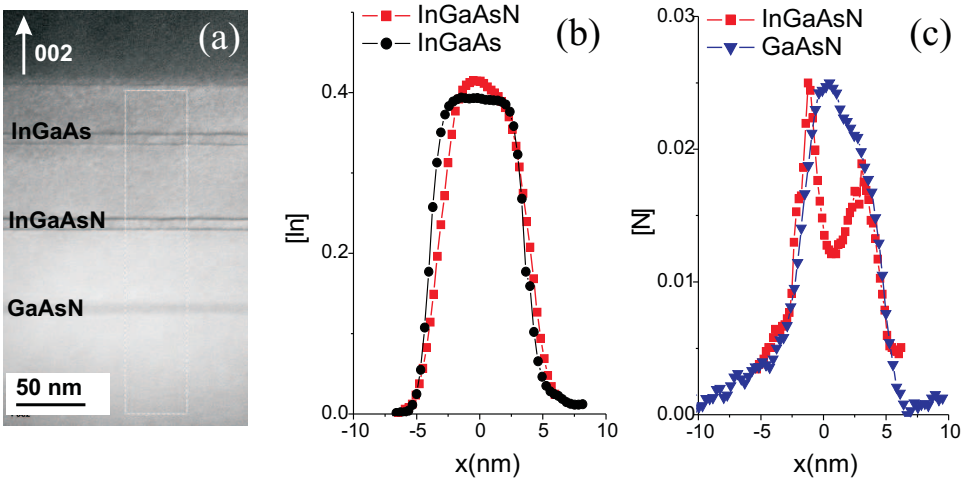


Figure 7.1: (a) Cross sectional transmission electron micrograph (TEM) of successive GaAsN, InGaAsN, and InGaAs quantum wells as indicated. (b) shows the obtained In concentration profile in the InGaAs and InGaAsN quantum wells, and (c) shows the obtained N concentration profiles in GaAsN and InGaAsN quantum wells. After Albrecht et al. [26].

molecular beam epitaxy (MBE) on a semi-insulating GaAs(001) substrate (further details on the growth process can be found in Ref. [47]). This geometry guarantees identical growth conditions for the incorporation of In, N and In+N in GaAs and thus makes correlation effects immediately accessible. The quantum wells were embedded into 50 nm thick GaAs barriers. A radio frequency plasma source was used as a nitrogen source. Nominal In-concentrations of $x = 39\,\%$ and N-concentrations of $y = 1.9\,\%$ were intended. From this sample electron transparent cross sectional samples in $\langle 100 \rangle$ projection were prepared. The In- and N-concentration profiles in these samples were analyzed by a combined evaluation (i) of the tetragonal lattice distortion of the unit cell obtained from experimental high resolution transmission electron micrographs (TEM) and (ii) of the intensity of the chemically sensitive (002) reflection in respective dark field images of the identical areas (the detailed characterization procedure can be found in Ref. [187]). The experimental error for quantification of In- and N-concentration, given by the fluctuations in the (002) dark field contrast are estimated to be within 10 % for N and 5 % for In.

Fig. 7.1(a) shows a typical dark field micrograph of the sample taken with the chemical sensitive (002) reflection. The GaAsN quantum well can be seen at the bottom by reduced (002) dark field intensity. It is followed by the InGaAsN and the InGaAs quantum wells, which are characterized by reduced dark field intensity in between narrow dark lines. Note that these lines are not the interfaces but related to a combination of In- and N-concentration that reduces the (002) intensity to zero. From the micrograph it can be seen that the quantum wells grow two-dimensionally without measurable thickness and lateral compositional fluctuations at the projected sample thickness of 30 nm. Fig. 7.1(b) shows the In-profiles of the InGaAs and the InGaAsN quantum wells. A maximum In-concentration of $x = 39\,\%$ is found for the InGaAs well with almost abrupt interfaces. In case of the quaternary InGaAsN well the maximum In-concentration is slightly higher $x = 41\,\%$ but the In-profile is smeared out and shows a Gaussian-like profile, which leads to identical average In-concentrations integrated over the profile.

It is most important here to examine the N-profiles of the GaAsN and the InGaAsN quantum wells shown in Fig. 7.1(c). The N-concentration in the GaAsN well shows a Gaussian shaped profile with a maximum N-concentration of $y = 2.5\,\%$, while an M-shaped profile is found in the InGaAsN quantum well. In case of the quaternary alloy the maximum N-concentration is reached at the interfaces. It has a value of about 2.5 % for an In-concentration of about 16 % and drops with increasing In-concentration in the center of the well to a value of about $y = 1.2\,\%$. M-shaped N-profiles have been found for all as-grown InGaAsN quantum wells with In-concentration $x > 25\,\%$, independent of growth mode (two dimensional or three dimensional), growth temperature (between 400 and 450 °C) and growth rate. Gaussian shaped N-profiles are found for In-concentrations below 20 %. These experimental results thus suggest an anticorrelation between N and In-incorporation during growth of InGaAsN at In-concentrations $x > 20\,\%$.

7.2.2 Theoretical modeling

In Chapter 5 we have shown that the formation energy of N substitution is strongly affected not only by the incorporation layer (surface, subsurface, bulk) but also by the specific *position* within that layer. Hence, for the case of quaternary alloys it is interesting to examine the *pair* energies of incorporated In-N pairs in these different atomic environments in the system (surface or bulk). This allows, together with our knowledge of kinetic mechanisms of the N atoms, to theoretically identify the physical origin of the In-N anticorrelation behavior.

Therefore, we consider the formation energy of a substitutionally incorporated pair of In and N atoms replacing a pair of Ga and As atoms. To do that, and since we are interested in the interaction between In and N rather than in their total concentrations, only differences in the formation energies are needed, i.e., the differences arising from the various modifications of the spatial arrangements of In and N. The differences are calculated with respect to a certain In-N configuration (a reference system). Since we have the same number of In and N atoms, the differences in surface formation energies reduce to differences in total energies (see Eq. (5.12)):

$$\Delta H^f_{\mathrm{InGaAsN}} \quad = \quad E^{tot}_{\mathrm{InGaAsN}} - E^{tot}_{\mathrm{InGaAsN^{ref}}}. \tag{7.1}$$

Here, $E^{tot}_{\mathrm{InGaAsN}}$ is the total energy of the considered system and $E^{tot}_{\mathrm{InGaAsN^{ref}}}$ is the total energy of the corresponding reference system. We chose as reference the same host structure as for the system of interest (i.e. bulk or surface, respectively) with the In- and N- substitutions fully separated to avoid any interaction between them. The separation is modeled by calculating the total energy of three systems: The host system with an In-substitution, with a N-substitution, and the pure GaAs host, i.e.:

$$E^{tot}_{\mathrm{InGaAsN^{ref}}} \quad = \quad E^{tot}_{\mathrm{InGaAs^{ref}}} + E^{tot}_{\mathrm{GaAsN^{ref}}} - E^{tot}_{\mathrm{GaAs^{ref}}}. \tag{7.2}$$

Using this definition the energy zero of $\Delta H^f_{\mathrm{InGaAsN}}$ corresponds to the case where In and N have no interaction (correlation). It also ensures that the chemical potentials do not enter since the stoichiometry of the interacting and the reference system is identical.

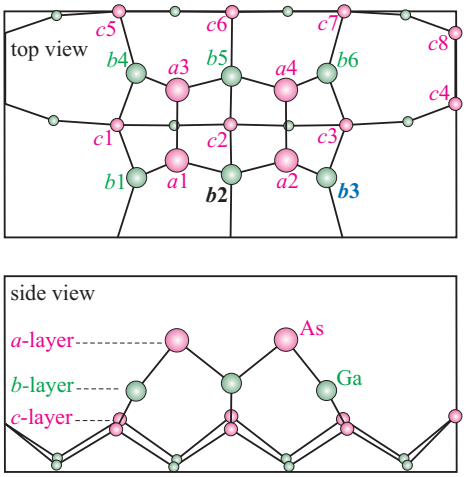


Figure 7.2: Top (upper panel) and side (lower panel) views of the $\beta 2(2\times 4)$ reconstruction. The labels give the N and In substitutional sites discussed in the text. Recall that the a- and c-layers consist of group V atoms, while the b-layer consist of group III-atoms.

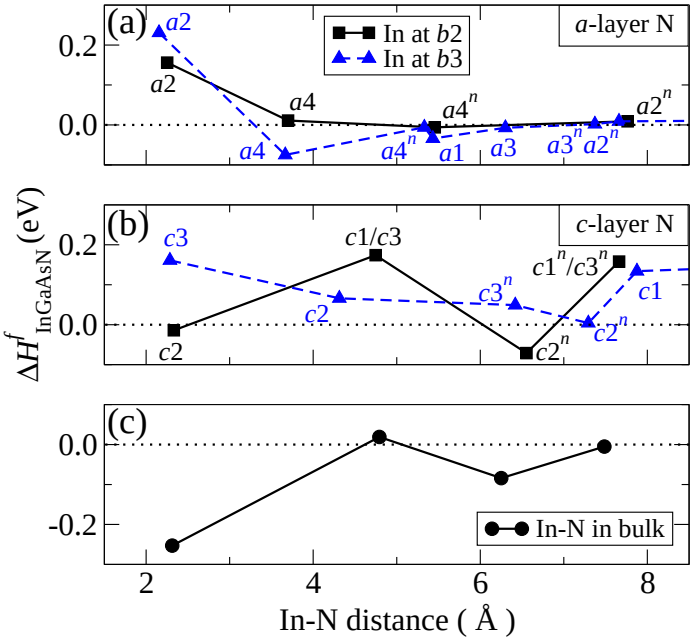


Figure 7.3: Formation energies of InGaAsN as a function of the In-N distance at the surface (a and b) and in the bulk (c). The labels give the positions of In and N as marked in the schematic plot of the $\beta 2(2\times 4)$ reconstruction of GaAs(001) surface shown in Fig. 7.2. The labels with n in the superscripts represent corresponding positions in an adjacent $\beta 2(2\times 4)$ cell.

In order to study the In-N pair energies at the GaAs(001) surface we have considered the $\beta 2(2\times 4)$ reconstruction as a prototype (see Fig. 7.2). Recall that this reconstruction has been considered in details in the previous chapters specially for the study of surface kinetics because of its stability over a wide range of chemical potentials both in experiment and theory [154]. At that surface we substituted a Ga atom in the uppermost cation layer (b-layer in Fig. 7.2) by an In atom, and an As atom in the first (a-layer) or in the second (c-layer) anion layer (Fig. 7.2) by a N atom. As can be seen in Fig. 7.2, two classes of cation sites exist: corner sites ($b1$, $b3$, $b4$, $b6$) and center sites ($b2$ and $b5$). For the In-substitution we therefore considered two cases. As an example of a corner and a center site we considered the $b3$ and $b2$ site, respectively. For the N-substitution *all* first and second anion layer sites have been considered. As reference systems for the surface calculations (the fully separated In and N substitutions, as described above) we have used the most favorable configuration (lowest energy) for each the single In- and N-substitutions in the corresponding surface layer. These are the $a4$ site for N in the first anion layer, the $c2$ site for N in the second anion layer, and the $b3$ site for In.

7.2.3 Results

First, for the interaction between the incorporated In-N pair in bulk GaAs, we have computed the formation energies as defined in Eq. (7.1) as a function of the In-N separation. The results

are shown in Fig. 7.3(c). As can be seen, the formation energy is lowest if In and N are nearest neighbors and increases by ∼0.25 eV as the In and N are fully separated. That is, In and N atoms have a strong attractive interaction between each other in bulk GaAs as expected due to strain compensation. As a result of this, a clear In-N compositional correlation would be observed. The fact that experimentally a clear anticorrelation is usually observed implies that this mechanism is not operational at characteristic MBE conditions considered here.

The absence of any compositional correlation indicates that short-range order is not a result of bulk thermodynamics but rather must be driven by kinetics. Therefore, we discuss now the pair energies of In-N substituted at the surface, which is essential to understand their interaction in the earlier steps of their incorporation. The corresponding calculated formation energies as a function of the In-N separation are summarized in Fig. 7.3(a) for N in the first anion layer (a-layer), and in Fig. 7.3(b) for N in the second anion layer (c-layer). Solid lines connecting the formation energies have been included as a guide for the eye. Different lines have been drawn for the two different In sites for N in both layers. Note that the data points corresponding to $c5$, $c6$, and $c7$ N-substitutions are omitted because they are much higher in energy (∼1.0–1.5 eV) than those of the other c-layer sites. Also the data points corresponding to $c4$ and $c8$ N-substitutions are excluded because the In-N interaction is found to be small (<0.02 eV) for these sites.

Let us first consider the case of N in the first anion layer. As can be seen in Fig. 7.3(a) forming an In-N bond at the surface is energetically highly unfavorable irrespective of whether In occupies a corner or a center site. Separating the In and N atoms, i.e., replacing the In-N bond with In-As and Ga-N bonds, reduces the formation energy by about 0.15–0.3 eV. Thus, for N in the top surface layer a strong compositional anticorrelation is expected.

In the final case for N in the second anion layer, the results presented in Fig. 7.3(a) show that the energy trends are not as pronounced as in the case of N in the first anion layer or in bulk GaAs. This is mainly due to the large dependence of the formation energy of N substitution on the specific subsurface substitution site (as can be seen also in Chapter 5). Nevertheless, a small repulsive interaction can be seen for both In sites. For In at the corner site ($b3$) a monotonous decrease (i.e. stabilization) of the formation energy with increasing distance is found. For the center site ($b2$) a non-monotonous behavior is found, i.e., moving the N away from the In first increases and then decreases the formation energy, which is due to the fact that the surface sites $c2$ and $c3$ are not equivalent. Therefore, to extract the In-N interaction only the energy of equivalent sites (i.e. same sites in different cells) should be compared. Thus for both $c2$ and $c3$, compared to $c2$ and $c3$ in the adjacent (2×4) unit cell (denoted by $c2^n$ and $c3^n$), a small repulsive interaction of 0.06 and 0.02 eV is observed, respectively.

7.2.4 Discussion

The results above show that the In-N interaction is strongly attractive in bulk GaAs, strongly repulsive at the top surface layer and weakly repulsive in the second surface layer. The origin of the qualitatively different behavior is an intricate interplay/competition between the chemical bond energy and the elastic strain energy. Chemical bond energies in the InGaAsN alloy system can be estimated from the respective cohesive energies of the binaries [106] which show the

	GaN	InN	GaAs	InAs
Cohesive energy E^c (eV)	8.96	7.72	6.52	6.20
Lattice constant (ZB) (Å)	4.52	4.98	5.65	6.03

Table 7.1: Cohesive energies between the different types of atoms in the binary semiconductor bulk systems and the corresponding lattice parameters (taken from Table 4.2).

following trend: $E^c_{\mathrm{GaN}} > E^c_{\mathrm{InN}} > E^c_{\mathrm{GaAs}} > E^c_{\mathrm{InAs}}$. The trend in elastic strain energy follows from the lattice mismatch between the equilibrium lattice constants of the constituting compounds: $a_{\mathrm{InAs}} > a_{\mathrm{GaAs}} > a_{\mathrm{InN}} > a_{\mathrm{GaN}}$ (see Table 7.1 for the actual values). The chemical component tends to stabilize Ga-N bonds and destabilize the In-N bonds, and thus favors In-N compositional anticorrelation. In contrast, the strain component leads to an attractive interaction between In and N: According to the sequence in lattice constants, embedding an In atom in GaAs results in compressive strained In-As bonds, while embedding a N results in tensile strained Ga-N bonds.

Using our results we get a detailed understanding of the interplay between chemical and strain energy. First, for the In-N interaction in the bulk, as an In and a N atom are brought closer together in the GaAs host the compressive and the tensile strain fields overlap and largely cancel, which results in an attractive interaction that significantly dominates over the repulsive chemical interaction in GaAs, thus resulting in a net attractive interaction.

In the subsurface region, there exist intrinsic compressive *and* tensile strained bonds resulting from the specific surface reconstruction. This has a strong effect on the stability of substituting atoms, as discussed in Chapter 5. This leads to a more complicated interaction behavior since the strain energy components of In-N bonds cannot relax completely. However, a weak net repulsive interaction can be noticed when comparing the energies at the same sites in different cells.

At the topmost surface layer, the atoms are under-coordinated and more flexible to relax, which means that the atoms can reduce strain more efficiently. Thus, in the top surface layer the chemical contribution dominates and hence a strong In-N repulsion can be seen. This can also be expected by looking at the solubility of N at GaAs (001) surfaces compared to at InAs (001) surfaces, as shown in Chapter 5. Significantly higher solubilities in GaAs have been found despite the fact that the lattice mismatch between InAs and InN is less than that that between GaAs and GaN. Hence, the solubility difference is mainly due to the chemical energy contribution which results in clearly more favorable GaN bonds than InN and hence higher solubilities of N at GaAs surfaces. The presence of In on the surface will make it less favorable for N to stick to the surface implying an inverse relation between the N concentration and the In concentration (i.e., the N concentration decreases with increasing In concentration).

Comparing these results to the experimental data which are clearly showing an In-N compositional anticorrelation, we conclude that under characteristic MBE growth conditions the local compositional order is predominantly determined by the surface and is not a bulk property. An immediate consequence of this finding is that the as-grown InGaAsN quaternary alloys are not in thermodynamic equilibrium but kinetically stabilized by freezing in the local bond distribution as formed at the surface. This also can be seen from our data regarding the kinetics barriers of N

diffusion in Chapter 6 between the top surface layer and the subsurface region, which show that the bulk and the subsurface incorporation mechanisms are not operational. The substitution of N in the top surface layer was found to be a dominant trapping mechanism that prevents the N from incorporating in the subsurface, which leads to freezing-in the surface configuration. This insight allows us to additionally explain the puzzling experimental investigations regarding the growth and annealing of InGaAsN: Low temperature growth freezes in the surface configuration (i.e. In-N anticorrelation), while annealing drives the system towards thermodynamic equilibrium (In-N correlation) [188, 189, 39].

7.3 Solubility of N in GaAs and InAs

In Chapter 5 we have shown that the maximal solubility of N in GaAs and InAs can be achieved through their incorporation in the subsurface layer. On the other hand, we have shown in Chapter 6 that the subsurface incorporation mechanism is prohibited under typical growth temperatures. This is further confirmed by the observed compositional anticorrelation between In and N in quaternary InGaAsN alloys discussed in the previous section. This issue makes it challenging to explain N concentrations reported in MBE experiments, which are significantly higher than the theoretical predictions for the concentrations resulting from topmost surface layer incorporation (as provided in Chapter 5). In the following we provide a detailed discussion in order to achieve a comprehensive understanding for this apparent contradiction between theory and experiment.

7.3.1 Solubility of N in GaAs

Incorporation mechanisms and solubility limits

A consequence of the results in Chapter 5 is that N solubility exhibits a complex behavior and depends strongly on the active incorporation mechanism. The first mechanism (bulk incorporation) is active under complete thermodynamic equilibrium conditions, i.e., all surface layers and the bulk are in thermodynamic equilibrium with the chemical reservoirs in the growth chamber, and an unrestricted exchange of all atoms is allowed. As our results show, this mechanism leads to a very poor N solubility ($\sim 10^{-3}$ % N; see Fig. 5.6(a) under As poor and N rich conditions) and cannot explain the experimentally achieved N solubility.

The second mechanism is N incorporation in the subsurface region of GaAs(001), i.e., the second anion layer is at thermodynamic equilibrium with the upper layers but not in equilibrium with the bulk. For this scenario, a N concentration as high as 12.5% can be achieved at moderate As and N-rich conditions (Fig. 5.9(a)), and can easily explain experimental solubilities. However, it is essential at this point to identify if this mechanism is functional in practice. Our theoretical calculations of the kinetic barriers and mechanisms at surfaces discussed in Chapter 6, showed that N atoms cannot diffuse from the top surface layer to incorporate in the subsurface layers because of high energy barriers which cannot be overcome at a temperature of 500 °C. In addition, our theoretical study for the $In_xGa_{1-x}As_{1-y}N_y$ quaternary alloys combined with experimental results discussed in Sec. 7.2 (also in Ref. [26]) further confirm that this mechanism is not active

under typical growth conditions. Therefore, this incorporation mechanism is not active and will be excluded from further discussion.

The last mechanism is the incorporation of N in the top surface layer: i.e., thermodynamic equilibrium is restricted solely to the uppermost surface layer. Although this mechanism leads to a higher N solubility than the bulk equilibrium solubility, it is still well below the N concentrations which are routinely achieved in MBE experiments. This can bee seen in Fig. 5.9(a) by the discrepancy between the gray shaded area and the first layer concentration at the upper limit of N chemical potential μ_{N} as defined in Eq. (5.14), i.e., at $\mu_{\mathrm{N}} - \mu_{\mathrm{N^N_2}} = 0$; indicated by the dotted line. The zero value means that the surface is in thermodynamic equilibrium with N_2 molecules in the gas phase. However, realistic MBE growth involves procedures and conditions that significantly affect the upper limit of the N chemical potential, which in turn has direct consequences on the calculated solubility in the topmost surface layer, as will be discussed in the following.

N chemical potential limits in MBE and topmost surface layer N solubility

At the equilibrium condition between N_2 gas molecules and the surface (at $\mu_{\mathrm{N}} = 0$), excess N atoms at the surface must be able to form N_2 molecules without kinetic limitations. In a realistic growth environment, this thermodynamic equilibrium with the N_2 gas will be hard/impossible to realize: First, in MBE growth a radio-frequency (rf) *N plasma* source is usually used, which provides several species in the plasma: excited N_2 molecules (N_2^*), electrons, *atomic* N, diatomic N, and ionized N species. These activated species indeed correspond to a much higher N chemical potential than that of the N_2 molecules, up to $\sim 5\,\mathrm{eV}$ per atom for atomic N compared to N in N_2 molecules. For instance, it has been shown experimentally that the characteristic vibrational temperature for the excited N_2 molecules in a N plasma source is $> 10^4\,\mathrm{K}$ (i.e. corresponding to a thermal energy of $\sim$0.86 eV) and the dissociation fraction of the N_2 molecules can be up to 70 % [190], which means the that the upper limit of the chemical potential of N is no longer limited by $\mu_{\mathrm{N_2}}$. Second, a spontaneous formation of N_2 molecules at the surface can be achieved only if the N atoms are sufficiently mobile and provided that the repulsive interaction between N atoms is sufficiently small so that it can be overcome at characteristic growth temperatures. However, as we have shown in Chapter 6, the diffusion of N adatoms is very slow and an equilibration cannot be achieved. Both mechanisms indicate that under typical MBE growth conditions the actual N chemical potential at the surface can be shifted by a few tenths of an eV towards higher energies.

It is important here to recall the additional entropy contributions to the free energy of N_2. As introduced earlier in Sec. 5.3.1, at a temperature of 500 °C and under MBE pressures of $\sim 10^{-5}\,\mathrm{atm}$, the entropy contributions lower the total energy of the N_2 molecule, thus shifting down the upper limit of μ_{N} by $\sim$1.26 eV (per N atom). However, for a plasma N source the entropic contributions for the different N species are different. Nevertheless, we note that these growth conditions apply on the growing substrate *not* on the incoming N species flow since they are not in equilibrium. The properties of the incoming N species are mostly determined by the plasma source itself. The experimental setups of the plasma source such as the applied rf-power and gas flow rate, have been found to have significant consequences on the structural and

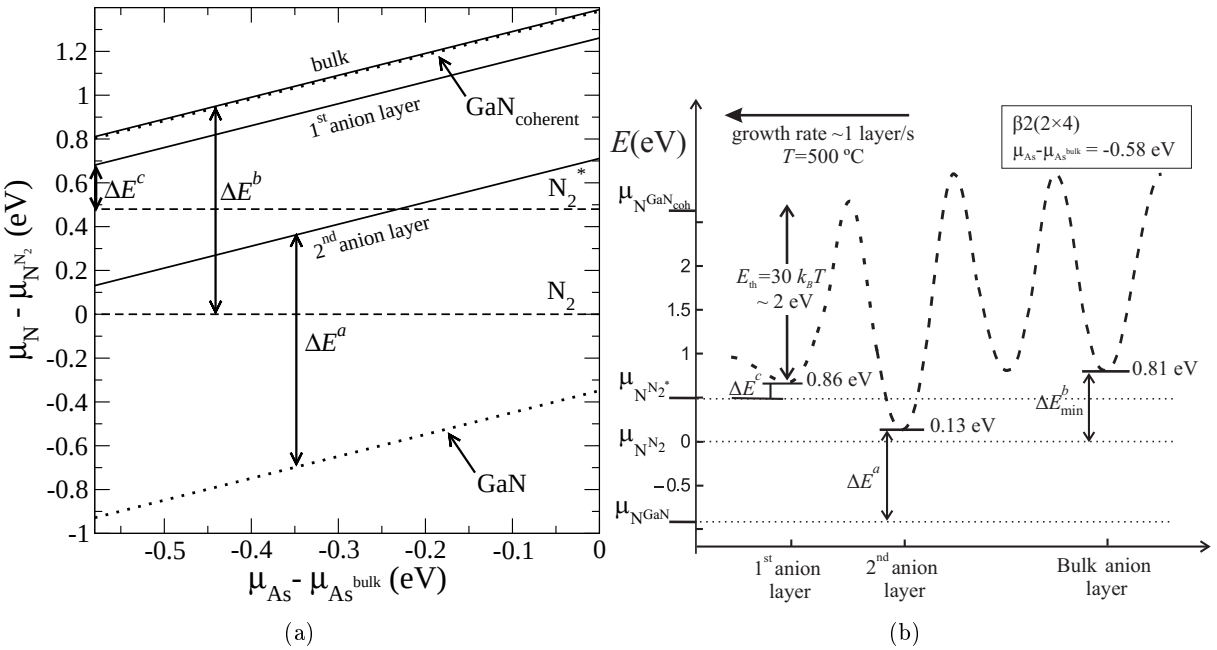


Figure 7.4: (a) Stability phase diagram of N in bulk GaAs and in the surface layers of the $\beta 2(2\times 4)$ reconstruction of GaAs(001) (solid lines). The dotted lines define the stability regions for bulk GaN and the coherent bulk GaN (at the lateral lattice constant of GaAs). Note that the latter line almost coincides with the stability line of N in bulk GaAs. The dashed lines indicate the chemical potential of N_2 and excited N_2 molecules (N_2^*). The arrows indicate the mechanism and formation energy used to calculate N solubility in Ref. [50] (ΔE^a), Ref. [51] (ΔE^b), and present work (ΔE^c). μ_{As} is restricted to the relevant region for N incorporation, i.e., to the region where the $\beta 2(2\times 4)$ reconstructed surface is stable. (b) Summary of the results taken from (a) at the more As poor limit $\mu_{\mathrm{As}} - \mu_{\mathrm{As^{bulk}}} = -0.58\,\mathrm{eV}$ (i.e. at the left border) in a representation corresponding to that in Fig. 1.2. E_{th} is the threshold energy as defined in Eq. (7.4) at 500 °C, and $\Delta E^b_{\mathrm{min}}$ is the lowest limit of ΔE^b (see text). The curve is dashed to indicate that the barriers were not identified so far.

optical properties of diluted nitride alloys due to the different kinetics of the various N species [191, 190, 192, 193, 194]. Therefore, the plasma source itself is the most important factor here.

Another potential upper limit of the N chemical potential is the formation of parasitic GaN. Earlier theoretical studies to calculate the solubility of N in GaAs have assumed the chemical potential reservoir to be that of bulk GaN [50]. Zhang and Wei [51] then proposed that μ_{N} can be further increased, since a relaxed GaN phase can form only after the formation of a N-rich cluster of critical size for dislocation formation, which in turn can form only after a spontaneous formation of a N-rich layer at the GaAs surface. The upper limit for μ_{N} was raised accordingly, i.e., to the point where the formation energy of N substitution at the surface equals to zero. In order to explain this and based on our results (Fig. 5.6) we show in Fig. 7.4(a) the stability phase diagram for N substitution in bulk GaAs and in the surface layers of the $\beta 2(2\times 4)$ reconstruction, showing also the stability regions for bulk GaN, the coherent GaN phase (at the lateral GaAs lattice constant), and of the N_2 molecules. It becomes clear that the formation of a coherent GaN layer or precipitate is unfavorable. Only when assuming strongly non-equilibrium conditions with $\mu_{\mathrm{N}} \gg \mu_{\mathrm{N_2}}$ sufficiently high N concentrations maybe reached to form a coherent

GaN cluster. Only if this cluster exceeds a critical size/thickness a fully relaxed GaN precipitate can form. Therefore, under typical MBE conditions, we expect that μ_N can be increased beyond μ_{N_2} *without* the formation of the relaxed GaN phase. Following this assumption and based on our results in the region where the $\beta 2(2\times 4)$ reconstruction is stable (relevant for MBE growth), μ_N can be increased by 0.68 eV above μ_{N_2} before spontaneously forming a coherent GaN layer. In contrast, to incorporate about 2% of N (assuming the first anion layer incorporation mechanism), it is sufficient to increase μ_N by 0.48 eV above μ_{N_2} (indicated in Fig. 7.4 as a dashed line marked with N_2^*).

Results in view of previous studies

In previous studies, two incorporation mechanisms have been proposed to explain the high N concentration which can be realized in experiment and which is much higher than the bulk N solubility. While these models had an important impact to qualitatively gain insight into realistic dopant incorporation, the underlying surface models where rather approximate. Based on the stability phase diagram summarized in Fig. 7.4 we are able to carefully reexamine these proposals and to identify the solubility for each of them.

In the early work by Zhang and Zunger [50] it has been assumed that N substitutes in the second anion surface layer, i.e., in our terminology that the subsurface incorporation mechanism is active. Using the valance force field approach, a (2×1) reconstruction, and GaN as the chemical potential reservoir they obtained a substitution energy of $\Delta E = 0.7\,\text{eV}$. With $T = 727\,^\circ\text{C}$ and assuming 3 sites of this type per layer of the actual $\beta 2(2\times 4)$ cell they calculated a concentration of $\sim 0.09\,\%$. Replacing the approximate (2×1) reconstruction by a realistic $\beta 2(2\times 4)$ cell, the corresponding substitution energy increases slightly ($\Delta E^a = 1.06\,\text{eV}$ in Fig. 7.4). For actual MBE growth conditions ($T = 500\,^\circ\text{C}$) this translates into a N concentration of $\approx 10^{-6}\,\%$, i.e., too low to explain experiment.

A second mechanism proposed by Zhang and Wei [51] assumes the third incorporation mechanism (bulk incorporation) to be the active mechanism. In that study the chemical potential limits were extended to the point where the formation of N substitution at the top surface layer becomes stable. With this they obtained $\mu_N^{\max} = 0$ and $\mu_{As}^{\min} = -0.44\,\text{eV}$, and hence a substitution energy of $\Delta E = 0.24\,\text{eV}$. Using $T = 650\,^\circ\text{C}$ they calculated a N concentration of 4 %. However, based on our calculations the first layer incorporation does not become stable in the thermodynamically allowed range. Using their values of the chemical potential limits we get $\Delta E = 0.95\,\text{eV}$ (indicated by ΔE^b in Fig. 7.4). This value can be reduced to 0.65 eV by going to more As poor conditions (indicated by $\Delta E^b_{\min}$). The large discrepancy is most likely related to using a reduced plane-waves cutoff energy of 25 Rydberg and LDA pseudopotentials. Using our data, this mechanism gives a N solubility of $\approx 0.006\,\%$ at $T = 500\,^\circ\text{C}$, again too low to explain experiment.

In contrast, assuming an incorporation in the first anion surface layer and an excited N_2 source which increases the upper limit of the N chemical potential, the lowest formation energy to incorporate N at the surface (indicated by ΔE^c in Fig. 7.4) gives solubilities in accordance with experiment.

These results are also summarized in Fig. 7.4(b) for the optimum As conditions at $\mu_{\mathrm{As}} - \mu_{\mathrm{As}^{\mathrm{bulk}}} = -0.58\,\mathrm{eV}$ as a function of the reaction coordinates (similar to the representation in Fig. 1.2 but using the actual energies). Let us define the thermal *threshold energy* according to Arrhenius law:

$$E_{\mathrm{th}} \quad = \quad k_B T \ln \frac{\Gamma}{\Gamma_0}, \tag{7.3}$$

where k_B is Boltzmann's constant, T is the temperature, and Γ_0/Γ is the prefactor/rate for the specific interlayer diffusion event. Let us assume that the adatom can be regarded as mobile if the hopping rate between surface layers is $\Gamma = 1\,\mathrm{s}^{-1}$, which is comparable to MBE growth rates. Therefore, using a common attempt frequency of $\Gamma_0 = 10^{13}\,\mathrm{s}^{-11}$ the threshold thermal energy becomes:

$$E_{\mathrm{th}} \quad \approx \quad 30\, k_B T. \tag{7.4}$$

In Fig. 7.4(b) the various chemical potential limits of the N species as previously discussed are indicated. In addition, the threshold energy as defined in Eq. (7.4) at 500 °C (of about $\sim 2\,\mathrm{eV}$) is indicated. The incorporation mechanism operational at a given growth temperature can be determined after the knowledge of the kinetic barriers. Based on the calculated kinetic barriers discussed in Chapter 6, in our case the only active mechanism under typical MBE conditions is the incorporation at the topmost surface layer.

7.3.2 Solubility of N in InAs versus GaAs

For the case of N incorporation at InAs(001), the actual kinetic barriers for the subsurface incorporation of N have not been calculated in this work. Hence, definite conclusions about the operational N incorporation mechanisms cannot be directly drawn. Since for this case a plasma N source is also employed, the upper limit of N chemical potential can be increased beyond that of the N_2 gas, as discussed above. For instance, let us assume that the incorporation of N at the InAs(001) follows the case of that at the GaAs(001) surface, i.e., only the first anion surface layer incorporation is possible. In this case, as can be seen in Fig. 5.9(b), based on the same arguments discussed in the case of GaAs the upper limit of the N chemical potential μ_{N} should be higher than $\mu_{\mathrm{N}^{\mathrm{N}_2}}$ in practice by $\gtrsim 0.8\,\mathrm{eV}$ in order to achieve concentrations in accordance with experiments.

For comparison, we show in Fig. 7.5 the calculated N concentrations resulting from the N incorporation of N at the first anion layer in both GaAs and InAs and at the same As chemical potential $\mu_{\mathrm{As}} - \mu_{\mathrm{As}^{\mathrm{bulk}}} = -0.44\,\mathrm{eV}$. The figure shows an approximately ~ 30 times lower N concentration at InAs than at GaAs surface at identical μ_{N}. This comes in a very good agreement with experimental observations. In particular, it has been reported that the N content in InAs is *systematically lower* than in GaAs, and that $\mathrm{InAs}_{1-x}\mathrm{N}_x$ alloys exhibit a N concentration which is roughly 25 times smaller than in $\mathrm{GaAs}_{1-x}\mathrm{N}_x$ under identical growth conditions ($T = 420\,^{\circ}\mathrm{C}$)

[1] This value is in the order of magnitude of optical phonons vibrations.

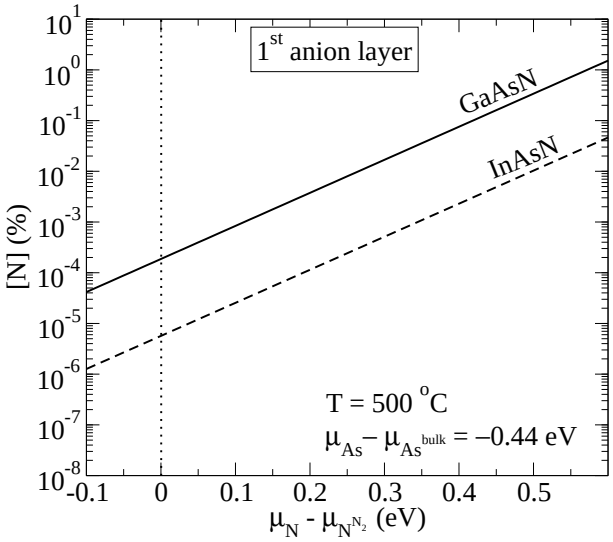


Figure 7.5: Comparison of the calculated N concentrations in InAs and GaAs as a function of the chemical potential of N at $T = 500\,^{\circ}\mathrm{C}$ and $\mu_{\mathrm{As}} - \mu_{\mathrm{As}^{\mathrm{bulk}}} = -0.44\,\mathrm{eV}$, assuming N incorporation in topmost surface layer.

[12]. These findings agree well with the results obtained here assuming N incorporation at the first anion layer and at the same chemical potentials and temperatures (Fig. 7.5).

7.4 Summary and conclusions

In the first part of this chapter we have shown that the compositional anticorrelation between In and N in quaternary $In_xGa_{1-x}As_{1-y}N_y$ quantum wells grown by MBE is mainly driven by surface kinetics rather than by bulk thermodynamics. The interaction between incorporated In-N pairs in bulk GaAs shows an attractive behavior and thus a correlated composition is predicted. On the other hand, the interaction at the surface is clearly repulsive for N incorporated at the top surface layer, and weakly repulsive for N at the subsurface layer, and hence an anti-correlated composition is expected at the surface. This repulsive interaction at the surface can also be predicted from the much lower solubilities of N at InAs surfaces compared to GaAs surfaces, indicating that N tend to bind to the Ga-rich regions at the surface which affects their distribution. This situation is very different from the simple bulk thermodynamics picture according to which the addition of In should enhance the N solubility. Therefore, the In-N anticorrelation is not a bulk property but a kinetically stabilized surface property.

These results are further supported by the previously calculated large energy barriers for N diffusion between the surface and the subsurface which prevent the equilibration with bulk under the typical MBE growth temperatures, confirming that the growth of these material systems indeed proceeds far from thermodynamic equilibrium. Our results further show that the annealing of these alloys under high temperatures allows for atomic rearrangements and thus drives the system towards equilibrium, which reduces the compositional fluctuations and In-N anticorrelation in as-grown alloys, in agreement with experimental observations.

In the second part of the chapter we have discussed the role of surface kinetics in determining

the solubility of N in GaAs and InAs. Based on our study of surface kinetics, supported by the experimentally observed In-N anticorrelation in quaternary $In_xGa_{1-x}As_{1-y}N_y$, the incorporation of N under typical growth temperatures is restricted to the uppermost surface layer, and the equilibration of N in the deep layers/bulk with the top surface layer is kinetically suppressed. The N bulk concentration is therefore solely dictated by the N solubility of the first surface layer. Furthermore, the results show that high N concentrations can be achieved under conditions far away from thermodynamic equilibrium and by using a plasma assisted source to supply the N species as routinely employed in MBE. The plasma assisted N source leads to a significant increase in the upper limit of the chemical potential of N which hence allows to increase its concentration. From these results we predict the solubility of N in InAs to be systematically lower than in GaAs by a factor of 30, which comes in a good agreement with recent experiments. Therefore, these results allow us to explain the experimentally obtained N concentrations and to identify the optimum growth conditions for achieving high N concentrations.

Chapter 8

Conclusions

While dilute nitride alloys; such as $GaAs_{1-x}N_x$, $InAs_{1-x}N_x$, and $In_xGa_{1-x}As_{1-y}N_y$; are appealing for a wide range of applications, the fabrication of high quality systems for device applications has been hindered by several challenges in their growth. These challenges include e.g. extremely low equilibrium solubility of N in GaAs and InAs, compositional fluctuations and inhomogeneities, and phase separation. In order to identify the driving mechanisms leading to these challenges we have employed extensive first principles calculations and developed appropriate methodologies and concepts. Based on our results we have been able to achieve a detailed understanding of the relevant physical processes that control the epitaxial growth of these alloys at the atomic level, which allows to improve their growth.

In the first step we have performed a detailed study of the thermodynamics of N in bulk GaAs and InAs and at surface and subsurface layers in order to identify the solubility limits and the optimal growth conditions for N incorporation. Our results showed that the substitution of N in the subsurface layer is energetically most favorable. This is found to be mainly due to an excellent cancellation effect of the tensile strain around N atoms because of the reconstruction-induced surface strain, which lowers their energy. In order to quantify the achievable N concentrations we have constructed the stability phase diagrams for N substitution in bulk and surface layers of the various relevant surface reconstructions as functions of the chemical potentials. Based on this we have developed an efficient scheme to calculate N concentration as a function of growth conditions (chemical potentials and temperatures). Using the computed phase diagrams for N solubility key conclusions have been made:

- The solubility of N in the subsurface layers is maximal, and it is up to 5 order of magnitudes larger than the bulk solubility.
- The solubility of N in InAs is generally lower than in GaAs under similar conditions, by up to 4 orders of magnitude; depending on the incorporation layer.
- The solubility of N is found to depend strongly on the specific surface reconstruction.
- A consequence of the previous point is that the maximal N solubility at surfaces is *not* achieved under extreme cation-rich conditions, contrary to what is expected from the pure

bulk behavior. Rather, our results clearly demonstrated that moderate cation-rich conditions are optimal for N incorporation.

The growth conditions must therefore be carefully controlled to achieve the maximal N solubility. The most promising approach is to identify kinetic regimes that permit subsurface incorporation but prevent equilibration with bulk. Such a kinetic regime allows to freeze the enhanced subsurface concentrations in the bulk during growth. A major part of this thesis was therefore devoted to develop theoretical concepts and tools to study the complex surface kinetics under these far from equilibrium growth conditions.

In a first step, we have focused on the diffusion of N adatoms at the GaAs(001) $\beta 2(2 \times 4)$ surface. The diffusion is found to show an exceptionally interesting behavior, which is unusual and different from the conventional understanding of adatom surface diffusion. This behavior leads to a dramatic failure of the classical concept of describing adatom kinetics by mapping the corresponding potential energy surface (PES). By carefully analyzing this behavior we were able to identify the underlying reasons. Due to the high reactivity and electronegativity of the N adatoms, in addition to their large size mismatch to GaAs substrate, they can easily break and form new covalent bonds at the surface, resulting in significant surface relaxations. Therefore, the potential energy landscape for the N adatom is complex and multi-dimensional in nature, i.e., it depends not only on the coordinates of the adatom but also of the neighboring surface atoms.

To overcome that difficulty and to identify the minima (binding sites) and the transition states at the complex energy landscape of the N adatom we have developed an approach that takes into account the degrees of freedom of the surface atoms. Based on these calculations an extended set of minima and transition states could be identified. Most of these states were completely unreachable using the conventional 2D-PES mapping approach. Using the minima and saddle point energies we have constructed the diffusion network for the N adatom. The resulting master equation was solved performing kinetic Monte-Carlo (KMC) simulations. As a result, the diffusion constants and the activation barriers could be derived. Performing a complex analysis of the results we developed a simplified approach that accurately reproduced the KMC results. The underlying simple analytical formulas allow for a detailed analysis and understanding of the complex diffusion of N on GaAs(001):

- The diffusion of N adatom is almost isotropic, with an activation energy of about 1.9 eV.
- Despite the complex diffusion network, a single energy barrier is found to be the main responsible one for determining the activation barrier.
- Within the surface unit cell the diffusion barriers are considerably lower.

A consequence of these results is that while the large activation barrier forbids a full equilibration of the N adatoms at the surface under typical growth temperatures, a local equilibration is realized, which guarantees their incorporation at the favorable substitutional sites at the surface.

Based on the identified adatom binding sites, we were able to propose possible kinetic reaction pathways that allow for the subsurface incorporation mechanism. The minimum energy paths

for these reactions have been identified using lateral and cross-sectional potential energy maps. Key conclusions based on these results are:

- The subsurface incorporation of N atoms is prohibited under typical growth temperatures by a trapping mechanism that leads to their incorporation only at the topmost surface layer.
- The average time for their incorporation the topmost surface layer is 1.9 ms, with average displacements of only 3.6 Å. Substitutional incorporation is therefore easy to realize under typical growth rates.

The insight gained by these studies is in clear contrast to what has been previously believed and assumed when explaining the experimentally observed N concentrations of $\gtrsim 2\,\%$. Furthermore, our calculated solubility based on the topmost surface layer incorporation mechanism predicts a much lower concentration, estimated at the N-rich limit when assuming equilibration with a gaseous N_2 atmosphere. These apparent contradictions with experiment could be resolved by carefully inspecting realistic growth conditions. For example, using a plasma N source as common raises the chemical potential of N considerably beyond that of the N_2 molecules, leading to a stronger N incorporation. This allows for the first time to predict N solubilities that are in accordance with experiments.

Based on the above results an understanding of the puzzling properties of the quaternary $In_xGa_{1-x}As_{1-y}N_y$ alloys has been achieved. Since the equilibration with the bulk is found to be forbidden based on the calculated barriers for subsurface N incorporation, the surface properties are kinetically stabilized and frozen-in during growth. While the strain compensation between In-N pairs in the bulk of GaAs leads to a net attractive interaction, the chemistry component prevails over the strain component at the surface resulting in a net repulsive interaction. This can be also seen from the lower N solubility at InAs surfaces compared to GaAs, which leads N to prefer Ga-rich regions. Hence, the In-N repulsive interaction at the surface leads to compositional anticorrelation between them in the as-grown alloys, contrary to what is expected from the pure bulk interaction. The use of thermal annealing after growth drives the alloy towards thermodynamic equilibrium, which improves the compositional anticorrelation and fluctuations. Consequently, in order to control the In-N compositional anticorrelation in practice, their interaction should be controlled by employing surface engineering, e.g. via controlling growth conditions, changing the composition, and adding surfactants.

In conclusion, based on *ab-initio* simulations we were able to construct a comprehensive understanding to address many challenging questions in the growth of dilute nitride alloys, which allows to experimentally optimize their growth and improve their properties. The developed concepts and methodologies are general and can be employed to address and realize key challenges and mechanisms for the epitaxial growth of highly mismatched multi-component systems.

Appendix A

Monte Carlo Methods

Monte Carlo methods refer to a class of methods that solve problems through using random numbers. In this work the Metropolis Monte Carlo method has been used to calculate N concentrations at surfaces under certain conditions. The kinetic Monte Carlo method has been applied to calculate the activation energy for the complex N adatom diffusion on the GaAs surface. In the following we briefly describe the algorithms of these methods.

A.1 Metropolis Monte Carlo

The Metropolis Monte Carlo method (MMC) [168] offers a powerful and elegant algorithm to sample the configuration space of a certain ensemble. It applies for systems under equilibrium, i.e., for systems that do not evolve dynamically, to obtain their statistical averages. This method is of significant importance in the cases for which it is difficult to determine the partition function of the system.

The Metropolis Monte Carlo algorithm can be summarized as follows:

1. Pick *any initial* configuration of the system, say ξ_n.
2. Pick a *trial* configuration, say ξ_t. This is usually taken to be similar to the configuration ξ_n (i.e. slightly modified).
3. Calculate the ratio of the probabilities:

$$R = \frac{P(\xi_t)}{P(\xi_n)}, \tag{A.1}$$

   where, e.g., according to Boltzmann's distribution $P(\xi_t) = c_0\, e^{-E(\xi_t)/k_B T}$.
4. Pick a random number r of a value between 0 and 1 (i.e., $r \in (0, 1]$).
5. If $r \leq R$ update the configuration to ξ_t, otherwise don't update. That is:

$$\xi_{n+1} = \begin{cases} \xi_t & r \leq R \\ \xi_n & \text{otherwise} \end{cases}. \tag{A.2}$$

Hence, the probability of accepting the new configuration is

$$P_{\text{accept}} = \min\left(1, R = \frac{P(\xi_t)}{P(\xi_n)}\right). \tag{A.3}$$

6. Go to step 2 and repeat for a sufficiently large number of times, until an equilibrium is achieved in the calculated quantity.

A.2 Kinetic Monte Carlo

The kinetic Monte Carlo (KMC) method is applied for systems that dynamically evolve from state to sate. Hence, it is very useful for surface physics applications such as adatom diffusion and growth simulations. KMC simulations provide an excellent option to overcome the time-scale problem and simulate times up to seconds and beyond, contrary to molecular dynamics (MD) simulations which require time steps in the femtoseconds region ($\sim 10^{-15}$ s) to resolve atomic vibrations and hence the total simulated time is very limited ($\lesssim 1$ microsecond) and hence cannot simulate processes that take place on much longer time scales. However, for KMC simulations the previous knowledge of the all possible processes and their rates is required.

The KMC algorithm can be summarized as follows:

1. Pick up a certain configuration of the system (e.g. a certain adatom position). Set the time $t = 0$.

2. For that configuration identify the *all* possible single processes/events N, and calculate their corresponding rates. For example, for an adatom at position (minimum) i jumping to the next N possible positions $\{j\}$ the rates can be calculated according to the transition state theory:

$$\Gamma_{ij} = \Gamma_0\, e^{-(E_{i,j}^{\text{trans}} - E_i)/k_B T}, \tag{A.4}$$

where $E_{i,j}^{\text{trans}}$ is the energy of the transition state between i and j.

3. Carry out an event randomly. To do that, first, the cumulative sums of the rates of the possible events are constructed, i.e.,

$$\Gamma_i^k = \sum_{j=1}^{k} \Gamma_{ij}, \quad \text{for } k = 1, \ldots, N. \tag{A.5}$$

Then, the event k for which $\Gamma_i^{k-1} < r\Gamma_i^{\text{tot}} \leq \Gamma_i^k$ is chosen, where $\Gamma_i^{\text{tot}} = \sum_{j=1}^{N} \Gamma_{ij}$ and r is a random number $r \in (0, 1]$.

4. Update the time $t = t + \Delta t$, where

$$\Delta t = -\frac{\ln r'}{\Gamma_i^{\text{tot}}}. \tag{A.6}$$

Here, r' is a new random number $r' \in (0, 1]$. Note that Δt has an average value of $\langle \Delta t \rangle = 1/\Gamma_i^{\text{tot}}$.

5. Update/change the list of events according to the new configuration and repeat the loop starting from point 2.

Bibliography

[1] URL http://www.internetworldstats.com/stats.htm.

[2] O. B. Shchekin and D. G. Deppe, Appl. Phys. Lett. **80**, 3277 (2002).

[3] M. Weyers, M. Sato, and H. Ando, Jpn. J. Appl. Phys., Part 2 **31**, L853 (1992).

[4] J. Neugebauer and C. G. Van De Walle, Phys. Rev. B **51**, 10568 (1995).

[5] S.-H. Wei and A. Zunger, Phys. Rev. Lett. **76**, 664 (1996).

[6] S. R. Bank, H. Bae, L. L. Goddard, H. B. Yuen, M. A. Wistey, R. Kudrawiec, and J. S. Harris, IEEE J. Quantum Electron. **43**, 773 (2007).

[7] K. Volz, D. Lackner, I. Nemeth, B. Kunert, W. Stolz, C. Baur, F. Dimroth, and A. W. Bett, J. Cryst. Growth **310**, 2222 (2008).

[8] S. R. Kurtz, A. A. Allerman, E. D. Jones, J. M. Gee, J. J. Banas, and B. E. Hammons, Appl. Phys. Lett. **74**, 729 (1999).

[9] K. M. Yu, W. Walukiewicz, J. Wu, J. W. Beeman, J. W. Ager, E. E. Haller, I. Miotkowski, A. K. Ramdas, and P. Becla, Appl. Phys. Lett. **80**, 1571 (2002).

[10] K. M. Yu, W. Walukiewicz, J. Wu, W. Shan, J. W. Beeman, M. A. Scarpulla, O. D. Dubon, and P. Becla, Phys. Rev. Lett. **91**, 246403 (2003).

[11] M. de la Mare, Q. Zhuang, A. Krier, A. Patane, and S. Dhar, Appl. Phys. Lett. **95**, 031110 (2009).

[12] V. Sallet, L. Largeau, O. Mauguin, L. Travers, and J. C. Harmand, Phys. Status Solidi B **242**, R43 (2005).

[13] J. W. Ager and W. Walukiewicz, Semicond. Sci. Technol. **17**, 741 (2002).

[14] D. K. Shih, H. H. Lin, and Y. H. Lin, Electron. Lett. **37**, 1342 (2001).

[15] A. Joullié, P. Christol, A. Baranov, and A. Vicet, *Mid-infrared 2-5 μm heterojunction laser diodes* (2003), URL http://dx.doi.org/10.1007/3-540-36491-9_1.

[16] K. Uesugi, N. Morooka, and I. Suemune, Appl. Phys. Lett. **74**, 1254 (1999).

[17] J.-Y. Duboz, J. A. Gupta, Z. R. Wasilewski, J. Ramsey, R. L. Williams, G. C. Aers, B. J. Riel, and G. I. Sproule, Phys. Rev. B **66**, 085313 (2002).

[18] W. Shan, W. Walukiewicz, J. W. Ager, E. E. Haller, J. F. Geisz, D. J. Friedman, J. M. Olson, and S. R. Kurtz, Phys. Rev. Lett. **82**, 1221 (1999).

[19] W. Shan, K. M. Yu, W. Walukiewicz, J. Wu, J. W. A. III, and E. E. Haller, J. Phys.: Condens. Matter **16**, S3355 (2004), URL `http://stacks.iop.org/0953-8984/16/S3355`.

[20] E. P. O'Reilly, A. Lindsay, P. J. Klar, A. Polimeni, and M. Capizzi, Semicond. Sci. Technol. **24**, 033001 (2009).

[21] M. Kondow, K. Uomi, A. Niwa, T. Kitatani, S. Watahiki, and Y. Yazawa, Jpn. J. Appl. Phys., Part 1 **35**, 1273 (1996).

[22] M. Kondow, T. Kitatani, S. Nakatsuka, M. C. Larson, K. Nakahara, Y. Yazawa, M. Okai, and K. Uomi, IEEE J. Sel. Top. Quantum Electron. **3**, 719 (1997).

[23] I. hsiu Ho and G. Stringfellow, J. Cryst. Growth **178**, 1 (1997).

[24] D. Schlenker, T. Miyamoto, Z. Pan, F. Koyama, and K. Iga, J. Cryst. Growth **196**, 67 (1999).

[25] X. Kong, A. Trampert, and K. H. Ploog, Micron **37**, 465 (2006).

[26] M. Albrecht, H. Abu-Farsakh, T. Remmele, L. Geelhaar, H. Riechert, and J. Neugebauer, Phys. Rev. Lett. **99**, 206103 (2007).

[27] M. Albrecht, V. Grillo, T. Remmele, H. P. Strunk, A. Y. Egorov, G. Dumitras, H. Riechert, A. Kaschner, R. Heitz, and A. Hoffmann, Appl. Phys. Lett. **81**, 2719 (2002).

[28] X. Kong, A. Trampert, E. Tournié, and K. H. Ploog, Appl. Phys. Lett. **87**, 171901 (2005).

[29] J.-M. Chauveau, A. Trampert, K. H. Ploog, and E. Tournie, Appl. Phys. Lett. **84**, 2503 (2004).

[30] D. Litvinov, D. Gerthsen, A. Rosenauer, M. Hetterich, A. Grau, P. Gilet, and L. Grenouillet, Appl. Phys. Lett. **85**, 3743 (2004).

[31] P. R. C. Kent and A. Zunger, Phys. Rev. Lett. **86**, 2613 (2001).

[32] K. Kim and A. Zunger, Phys. Rev. Lett. **86**, 2609 (2001).

[33] Z. C. Feng, ed., *III-Nitride Semiconductor Materials* (Imperial College Press, 2006).

[34] A. R. Kovsh, J. S. Wang, L. Wei, R. S. Shiao, J. Y. Chi, B. V. Volovik, A. F. Tsatsul'nikov, and V. M. Ustinov, J. Vac. Sci. Technol. B **20**, 1158 (2002).

[35] Q. X. Zhao, S. M. Wang, M. Sadeghi, A. Larsson, M. Friesel, and M. Willander, Appl. Phys. Lett. **89**, 031907 (2006).

[36] W. G. Bi and C. W. Tu, Appl. Phys. Lett. **70 (12)**, 1608 (1997).

[37] V. Lordi, V. Gambin, S. Friedrich, T. Funk, T. Takizawa, K. Uno, and J. S. Harris, Phys. Rev. Lett. **90**, 145505 (2003).

[38] V. Lordi, H. B. Yuen, S. R. Bank, M. A. Wistey, J. S. Harris, and S. Friedrich, Phys. Rev. B **71**, 125309 (2005).

[39] K. Uno, M. Yamada, I. Tanaka, O. Ohtsuki, and T. Takizawa, J. Cryst. Growth **278**, 214 (2005).

[40] E. M. Pavelescu, J. Wagner, H. P. Komsa, T. T. Rantala, M. Dumitrescu, and M. Pessa, J. Appl. Phys. **98**, 083524 (2005).

[41] Z. Pan, L. H. Li, W. Zhang, Y. W. Lin, R. H. Wu, and W. Ge, Appl. Phys. Lett. **77**, 1280 (2000).

[42] S. Karirinne, E. M. Pavelescu, J. Konttinen, T. Jouhti, and M. Pessa, New Journal Of Physics **6**, 192 (2004).

[43] T. Bouragba, M. Mihailovic, F. Reveret, P. Disseix, J. Leymarie, A. Vasson, B. Damilano, M. Hugues, J. Massies, and J. Y. Duboz, J. Appl. Phys. **101**, 073510 (2007), URL `http://link.aip.org/link/?JAP/101/073510/1`.

[44] H. Riechert, A. Y. Egorov, D. Livshits, B. Borchert, and S. Illek, Nanotechnology **11**, 201 (2000).

[45] N. Tansu, J. Y. Yeh, and L. J. Mawst, IEEE J. Sel. Top. Quantum Electron. **9**, 1220 (2003).

[46] H. Riechert, A. Ramakrishnan, and G. Steinle, Semicond. Sci. Technol. **17**, 892 (2002).

[47] A. Y. Egorov, D. Bernklau, B. Borchert, S. Illek, D. Livshits, A. Rucki, M. Schuster, A. Kaschner, A. Hoffmann, G. Dumitras, et al., J. Cryst. Growth **227-228**, 545 (2001).

[48] G. Steinle, H. Riechert, and A. Y. Egorov, Electronics Letters **37**, 93 (2001).

[49] J. Tersoff, Phys. Rev. Lett. **74**, 5080 (1995).

[50] S. B. Zhang and A. Zunger, Appl. Phys. Lett. **71**, 677 (1997).

[51] S. B. Zhang and S.-H. Wei, Phys. Rev. Lett. **86**, 1789 (2001).

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

[53] W. Kohn and L. J. Sham, Phys. Rev. **140**, 1133 (1965).

[54] J. Grotendorst, S. Blügel, and D. Marx, eds., *Computational Nanoscience: Do It Yourself!* (John von Neumann Institute for Computing (NIC), 2006), available online, URL `http://www.fz-juelich.de/nic-series/volume31/`.

[55] M. Born and R. Oppenheimer, Annalen Der Physik **84**, 0457 (1927).

[56] B. Grabowski, Master's thesis, University of Paderborn (2005).

[57] R. P. Feynman, Phys. Rev. **56**, 340 (1939).

[58] L. H. Thomas, Proceedings Of The Cambridge Philosophical Society **23**, 542 (1927).

[59] E. Fermi, Atti Reale Accademia Nazionale Dei Lincei **6**, 602 (1927).

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

[61] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).

[62] R. O. Jones and O. Gunnarsson, Rev. Mod. Phys. **61**, 689 (1989).

[63] A. D. Becke, Phys. Rev. A **38**, 3098 (1988).

[64] J. P. Perdew, Phys. Rev. B **33**, 8822 (1986).

[65] C. T. Lee, W. T. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).

[66] J. P. Perdew and Y. Wang, Phys. Rev. B **45**, 13244 (1992).

[67] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[68] B. Hammer, L. B. Hansen, and J. K. Norskov, Phys. Rev. B **59**, 7413 (1999).

[69] D. R. Hamann, M. Schlüter, and C. Chiang, Phys. Rev. Lett. **43**, 1494 (1979).

[70] A. Zunger and M. L. Cohen, Phys. Rev. B **20**, 4082 (1979).

[71] G. P. Kerker, J. Phys. C **13**, L189 (1980).

[72] G. B. Bachelet, D. R. Hamann, and M. Schluter, Phys. Rev. B **26**, 4199 (1982).

[73] D. Vanderbilt, Phys. Rev. B **41**, 7892 (1990).

[74] A. M. Rappe, K. M. Rabe, E. Kaxiras, and J. D. Joannopoulos, Phys. Rev. B **41**, 1227 (1990).

[75] N. Troullier and J. L. Martins, Phys. Rev. B **43**, 1993 (1991).

[76] J. S. Lin, A. Qteish, M. C. Payne, and V. Heine, Phys. Rev. B **47**, 4174 (1993).

[77] S. G. Louie, S. Froyen, and M. L. Cohen, Phys. Rev. B **26**, 1738 (1982).

[78] M. Fuchs, J. L. F. Da Silva, C. Stampfl, J. Neugebauer, and M. Scheffler, Phys. Rev. B **65**, 245212 (2002).

[79] L. Kleinman and D. M. Bylander, Phys. Rev. Lett. **48**, 1425 (1982).

[80] M. Fuchs and M. Scheffler, Comput. Phys. Commun. **119**, 67 (1999).

[81] C. Kittel, *Introduction to Solid State Physics* (John Wiley & Sons, 2005), 8th ed.

[82] N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Holt, Rinehart and Winston, 1976).

[83] A. Baldereschi, Phys. Rev. B **7**, 5212 (1973).

[84] D. J. Chadi and M. L. Cohen, Phys. Rev. B **8**, 5747 (1973).

[85] H. J. Monkhorst and J. D. Pack, Phys. Rev. B **13**, 5188 (1976).

[86] K. Shiraishi, J. Phys. Soc. Jpn. **59**, 3455 (1990).

[87] URL `www.sphinxlib.de`.

[88] G. Kresse and J. Furthmuller, Comp. Mater. Sci. **6**, 15 (1996).

[89] M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, and J. D. Joannopoulos, Rev. Mod. Phys. **64**, 1045 (1992).

[90] C. G. Van De Walle and P. E. Blöchl, Phys. Rev. B **47**, 4244 (1993).

[91] O. K. Andersen, Phys. Rev. B **12**, 3060 (1975).

[92] D. D. Koelling and G. O. Arbman, J. Phys. F: Metal Physics **5**, 2041 (1975).

[93] P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).

[94] N. A. W. Holzwarth, G. E. Matthews, R. B. Dunning, A. R. Tackett, and Y. Zeng, Phys. Rev. B **55**, 2005 (1997).

[95] C. M. Goringe, D. R. Bowler, and E. Hernandez, Rep. Prog. Phys. **60**, 1447 (1997).

[96] J. C. Slater and G. F. Koster, Phys. Rev. **94**, 1498 (1954).

[97] D. Porezag, T. Frauenheim, T. Kohler, G. Seifert, and R. Kaschner, Phys. Rev. B **51**, 12947 (1995).

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

[99] T. Frauenheim, G. Seifert, M. Elstner, Z. Hajnal, G. Jungnickel, D. Porezag, S. Suhai, and R. Scholz, Phys. Status Solidi B **217**, 41 (2000).

[100] L. Ismer, Ph.D. thesis, University of Paderborn (2008), available online, URL `http://ubdok.uni-paderborn.de/servlets/DocumentServlet?id=8295`.

[101] S. Nikolov, M. Petrov, L. Lymperakis, M. Friák, C. Sachs, H.-O. Fabritius, D. Raabe, and J. Neugebauer, Adv. Mater. **21**, 1 (2009).

[102] J. Elsner, Ph.D. thesis, University of Paderborn (1998), available online, URL `http://ubdok4.uni-paderborn.de/servlets/DocumentServlet?id=5002`.

[103] M. Sternberg, Ph.D. thesis, University of Paderborn (2001).

[104] M. Elstner, Ph.D. thesis, University of Paderborn (1998), in German, URL `http://ubdok4.uni-paderborn.de/servlets/DocumentServlet?id=5003`.

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

[106] W. A. Harrison, *Electronic Structure and the Properties of Solids* (Dover, New York, 1989).

[107] A. V. Podolskiy and P. Vogl, Phys. Rev. B **69**, 233101 (2004).

[108] R. Pariser, J. Chem. Phys. **24**, 250 (1956).

[109] R. G. Parr and R. G. Pearson, J. Am. Chem. Soc. **105**, 7512 (1983).

[110] URL `http://www.dftb.org`.

[111] A. Kley, P. Ruggerone, and M. Scheffler, Phys. Rev. Lett. **79**, 5278 (1997).

[112] F. D. Murnaghan, Proc. Natl. Acad. Sci. U.S.A. **30**, 244 (1944).

[113] S. Kotochigova, Z. H. Levine, E. L. Shirley, M. D. Stiles, and C. W. Clark, Phys. Rev. A **56**, 5191 (1997), URL `http://physics.nist.gov/PhysRefData/DFTdata/Tables/ptable.html`.

[114] S. Kotochigova, Z. H. Levine, E. L. Shirley, M. D. Stiles, and C. W. Clark, Phys. Rev. A **55**, 191 (1997), URL `http://physics.nist.gov/PhysRefData/DFTdata/Tables/ptable.html`.

[115] M. Bernasconi, G. L. Chiarotti, and E. Tosatti, Phys. Rev. B **52**, 9988 (1995).

[116] R. J. Needs, R. M. Martin, and O. H. Nielsen, Phys. Rev. B **33**, 3778 (1986).

[117] D. R. Lide, ed., *CRC Handbook Of Chemistry and Physics* (CRC Press, 2005), 86th ed.

[118] M. Haugk, J. Elsner, and T. Frauenheim, J. Phys.: Condens. Matter **9**, 7305 (1997).

[119] W. G. Schmidt, Appl. Phys. A: Mater. Sci. Process. **75**, 89 (2002).

[120] V. P. LaBella, H. Yang, D. W. Bullock, P. M. Thibado, P. Kratzer, and M. Scheffler, Phys. Rev. Lett. **83**, 2989 (1999).

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

[122] G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).

[123] X. Gonze, J. M. Beuken, R. Caracas, F. Detraux, M. Fuchs, G. M. Rignanese, L. Sindic, M. Verstraete, G. Zerah, F. Jollet, et al., Computational Materials Science **25**, 478 (2002).

[124] X. Gonze, G. M. Rignanese, M. Verstraete, J. M. Beuken, Y. Pouillon, R. Caracas, F. Jollet, M. Torrent, G. Zerah, M. Mikami, et al., Zeitschrift Fur Kristallographie **220**, 558 (2005).

[125] J. Neugebauer and M. Scheffler, Phys. Rev. B **46**, 16067 (1992).

[126] H. J. Beister, K. Strossner, and K. Syassen, Phys. Rev. B **41**, 5535 (1990).

[127] C. Stampfl and C. G. Van de Walle, Phys. Rev. B **59**, 5521 (1999).

[128] A. Trampert, O. Brandt, and K. H. Ploog, Gallium Nitride (GaN) I **50**, 167 (1998).

[129] M. R. Ranade, F. Tessier, A. Navrotsky, and R. Marchand, J. Mater. Res. **16**, 2824 (2001).

[130] M. D. Pashley, Phys. Rev. B **40**, 10481 (1989).

[131] K. B. Ozanyan, P. J. Parbrook, M. Hopkinson, C. R. Whitehouse, Z. Sobiesierski, and D. I. Westwood, J. Appl. Phys. **82**, 474 (1997).

[132] L. J. Whitman, P. M. Thibado, S. C. Erwin, B. R. Bennett, and B. V. Shanabrook, Phys. Rev. Lett. **79**, 693 (1997).

[133] P. Laukkanen, M. P. J. Punkkinen, H. P. Komsa, M. Ahola-Tuomi, K. Kokko, M. Kuzmin, J. Adell, J. Sadowski, R. E. Perala, M. Ropo, et al., Phys. Rev. Lett. **100**, 086101 (2008).

[134] G. P. Srivastava, Appl. Surf. Sci. **252**, 7600 (2006).

[135] S. Z. Wang, S. F. Yoon, and W. K. Loke, J. Appl. Phys. **94**, 2662 (2003).

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

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

[138] S. B. Zhang and S. H. Wei, Phys. Rev. Lett. **92**, 086102 (2004).

[139] R. T. Jacobsen, R. B. Stewart, and M. Jahangiri, J Phys Chem Ref Data **15**, 735 (1986), URL `http://webbook.nist.gov`.

[140] J. E. Northrup and J. Neugebauer, Appl. Phys. Lett. **85**, 3429 (2004).

[141] J. E. Northrup, R. DiFelice, and J. Neugebauer, Phys. Rev. B **56**, R4325 (1997).

[142] H. H. Farrell and C. J. Palmstrom, J. Vac. Sci. Technol. B **8**, 903 (1990).

[143] W. G. Schmidt, S. Mirbt, and F. Bechstedt, Phys. Rev. B **62**, 8087 (2000).

[144] P. Laukkanen, M. Kuzmin, R. E. Perala, M. Ahola, S. Mattila, I. J. Vayrynen, J. Sadowski, J. Konttinen, T. Jouhti, C. S. Peng, et al., Phys. Rev. B **72**, 045321 (2005).

[145] A. Ohtake, M. Ozeki, T. Yasuda, and T. Hanada, Phys. Rev. B **65**, 165315 (2002).

[146] A. Ohtake, Phys. Rev. B **74**, 165322 (2006).

[147] A. Ohtake, Surf. Sci. Rep. **63**, 295 (2008).

[148] M. Takahasi, Y. Yoneda, and J. Mizuki, Appl. Surf. Sci. **237**, 219 (2004).

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

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

[151] G. R. Bell, J. G. Belk, C. F. McConville, and T. S. Jones, Phys. Rev. B **59**, 2947 (1999).

[152] J. E. Northrup and S. Froyen, Phys. Rev. Lett. **71**, 2276 (1993).

[153] E. Penev, P. Kratzer, and M. Scheffler, Phys. Rev. Lett. **93**, 146102 (2004).

[154] S. H. Lee, W. Moritz, and M. Scheffler, Phys. Rev. Lett. **85**, 3890 (2000).

[155] A. Ohtake, S. Tsukamoto, M. Pristovsek, and N. Koguchi, Appl. Surf. Sci. **212**, 146 (2003).

[156] A. Ohtake, P. Kocan, K. Seino, W. G. Schmidt, and N. Koguchi, Phys. Rev. Lett. **93**, 266101 (2004).

[157] K. Seino, W. G. Schmidt, and A. Ohtake, Phys. Rev. B **73**, 035317 (2006).

[158] F. Grosse, W. Barvosa-Carter, J. J. Zinck, and M. F. Gyure, J. Vac. Sci. Technol. B **20**, 1178 (2002).

[159] C. Ratsch, W. Barvosa-Carter, F. Grosse, J. H. G. Owen, and J. J. Zinck, Phys. Rev. B **62**, R7719 (2000).

[160] B. P. Tinkham, W. Braun, K. H. Ploog, M. Takahasi, J. Mizuki, and F. Grosse, J. Vac. Sci. Technol. B **26**, 1516 (2008).

[161] R. H. Miwa and G. P. Srivastava, Phys. Rev. B **62**, 15778 (2000).

[162] C. Ohler, C. Daniels, A. Forster, and H. Luth, J. Vac. Sci. Technol. B **15**, 702 (1997).

[163] R. H. Miwa, R. Miotto, and A. C. Ferraz, Surf. Sci. **542**, 101 (2003).

[164] A. Jenichen, C. Engler, G. Leibiger, and V. Gottschalch, Phys. Status Solidi B **242**, 2820 (2005).

[165] R. Z. Bakhtizin, T. Sakurai, Q. K. Xue, and T. Hashizume, Sov. Phys.–Usp. **167**, 1227 (1997).

[166] W. G. Schmidt, Appl. Phys. A: Mater. Sci. Process. **65**, 581 (1997).

[167] M. Itoh and T. Ohno, Phys. Rev. B **62**, 7219 (2000).

[168] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).

[169] J. S. Wang, A. R. Kovsh, L. Wei, J. Y. Chi, Y. T. Wu, P. Y. Wang, and V. M. Ustinov, Nanotechnology **12**, 430 (2001).

[170] T. D. Veal, L. F. J. Piper, P. H. Jefferson, I. Mahboob, C. F. McConville, M. Merrick, T. J. C. Hosea, B. N. Murdin, and M. Hopkinson, Appl. Phys. Lett. **87**, 182114 (2005).

[171] S. Nishio, A. Nishikawa, R. Katayama, K. Onabe, and Y. Shiraki, J. Cryst. Growth **251**, 422 (2003).

[172] Q. Zhuang, A. Godenir, and A. Krier, J. Phys. D: Appl. Phys. **41**, 132002 (2008).

[173] J. Wagner, K. Kohler, P. Ganser, and M. Maier, Appl. Phys. Lett. **87**, 051913 (2005).

[174] P. E. Blöchl, C. G. van de Walle, and S. T. Pantelides, Phys. Rev. Lett. **64**, 1401 (1990).

[175] G. Brocks, P. J. Kelly, and R. Car, Phys. Rev. Lett. **66**, 1729 (1991).

[176] G. Brocks, P. J. Kelly, and R. Car, Phys. Rev. Lett. **70**, 2786 (1993).

[177] B. D. Yu and A. Oshiyama, Phys. Rev. Lett. **72**, 3190 (1994).

[178] T. Zywietz, J. Neugebauer, and M. Scheffler, Appl. Phys. Lett. **73**, 487 (1998).

[179] M. A. Salmi, M. Alatalo, T. Ala-Nissila, and R. M. Nieminen, Surf. Sci. **425**, 31 (1999).

[180] J. Neugebauer, T. Zywietz, M. Scheffler, and J. Northrup, Appl. Surf. Sci. **159**, 355 (2000).

[181] K. Seino, A. Ishii, and T. Aisaka, J. Cryst. Growth **237-239**, 121 (2002).

[182] A. Ishii, T. Aisaka, J. W. Oh, M. Yoshita, and H. Akiyama, Appl. Phys. Lett. **83**, 4187 (2003).

[183] J. Neugebauer, T. K. Zywietz, M. Scheffler, J. E. Northrup, H. Chen, and R. M. Feenstra, Phys. Rev. Lett. **90**, 056101 (2003).

[184] P. Kratzer, E. Penev, and M. Scheffler, Appl. Phys. A: Mater. Sci. Process. **75**, 79 (2002).

[185] G. Mills, H. Jonsson, and G. K. Schenter, Surf. Sci. **324**, 305 (1995).

[186] P. Hänggi, P. Talkner, and M. Borkovec, Rev. Mod. Phys. **62**, 251 (1990).

[187] V. Grillo, M. Albrecht, T. Remmele, H. P. Strunk, A. Y. Egorov, and H. Riechert, J. Appl. Phys. **90**, 3792 (2001).

[188] M. Ramsteiner, G. Mussler, P. Kleinert, and K. H. Ploog, Appl. Phys. Lett. **87**, 111907 (2005).

[189] M. Hugues, B. Damilano, J. M. Chauveau, J. Y. Duboz, and J. Massies, Phys. Rev. B **75**, 045313 (2007).

[190] H. Carrere, A. Arnoult, A. Ricard, X. Marie, T. Amand, and E. Bedel-Pereira, Solid-State Electron. **47**, 419 (2003).

[191] H. Carrere, A. Arnoult, A. Ricard, and E. Bedel-Pereira, J. Cryst. Growth **243**, 295 (2002), URL http://www.sciencedirect.com/science/article/B6TJ6-46B6X4G-7/2/eb2b20acd18edc2db944955cd5ba59b7.

[192] M. M. Oye, T. J. Mattord, G. A. Hallock, S. R. Bank, M. A. Wistey, J. M. Reifsnider, A. J. Ptak, H. B. Yuen, J. S. Harris, and A. L. Holmes, Appl. Phys. Lett. **91**, 191903 (2007).

[193] V. A. Grant, R. P. Campion, C. T. Foxon, W. Lu, S. Chao, and E. C. Larkins, Semicond. Sci. Technol. **22**, 15 (2007).

[194] J. Miguel-Sanchez, A. Guzman, U. Jahn, A. Trampert, J. M. Ulloa, E. Munoz, and A. Hierro, J. Appl. Phys. **101**, 103526 (2007).

List of Publications

1. *Kinetics of N adatoms at GaAs(001) surface: The breakdown of conventional surface diffusion.* Hazem Abu-Farsakh and Jörg Neugebauer. To be submitted.

2. *Kinetic mechanisms for subsurface incorporation of N at GaAs(001) surface from first-principles.* Hazem Abu-Farsakh and Jörg Neugebauer. To be submitted.

3. *Theoretical modeling of growth processes, extended defects, and electronic properties of III-nitride semiconductor nanostructures.* Liverios Lymperakis, Hazem Abu-Farsakh, Oliver Marquardt, Tilmann Hickel, and Jörg Neugebauer. To be submitted.

4. *Enhancing nitrogen solubility in GaAs and InAs by surface kinetics: An ab-initio study.* Hazem Abu-Farsakh and Jörg Neugebauer. Physical Review B **79**, 155311 (2009).

5. *Compositional correlation and anticorrelation in quaternary alloys: Competition between bulk thermodynamics and surface kinetics.* M. Albrecht, H. Abu-Farsakh, T. Remmele, L. Geelhaar, H. Riechert, J. Neugebauer. Physical Review Letters **99**, 206103 (2007).

6. *Ionicity scale based on the centers of maximally localized Wannier functions.* Hazem Abu-Farsakh and Abdallah Qteish. Physical Review B **75**, 085201 (2007).