CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Theory of elasticity and electric polarization effects in the group-III nitrides
Authors	Caro Bayo, Miguel Ángel
Publication date	2013
Original Citation	Caro Bayo, M. Á. 2013. Theory of elasticity and electric polarization effects in the group-III nitrides. PhD Thesis, University College Cork.
Type of publication	Doctoral thesis
Rights	© 2013, Miguel Ángel Caro Bayo - http://creativecommons.org/licenses/by-nc-nd/3.0/
Download date	2026-11-07 09:39:49
Item downloaded from	https://hdl.handle.net/10468/1344

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Theory of elasticity and electric polarization effects in the group-III nitrides

Miguel Ángel Caro Bayo

Submitted for the degree of
Doctor of Philosophy



UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Department of Physics
&
Tyndall National Institute
University College Cork

September 9, 2013

Supervisors:
Eoin P. O'Reilly
Stefan Schulz

Head of Department:
John McInerney

Contents

Preface **i**

Acknowledgments **iii**

List of publications **vii**

1 Introduction **1**

2 Theory and methods **7**
2.1 Strain effects in periodic solids 8
2.1.1 Elasticity 11
2.1.2 Piezoelectricity 13
2.2 Density functional theory 15
2.2.1 Kohn-Sham DFT 16
2.2.2 Generalized Kohn-Sham approaches: the HSE hybrid functional 22
2.2.3 Pseudopotentials and the projector augmented-wave method 24
2.2.4 Plane-wave formalism 28
2.2.5 The VASP package 30
2.3 Modern theory of polarization: the Berry-phase technique 32
2.3.1 Fundamentals of the Berry-phase formalism 33
2.3.2 Practical considerations: removing the uncertainty in **P** . . . 35

3 Built-in field control in nitride nanostructures **43**
3.1 Introduction 43
3.2 Calculation of the built-in field 46
3.2.1 Built-in potential in quantum dots and wells 46
3.2.2 Nitride alloy material parameters 48
3.3 Built-in field reduction 52
3.3.1 GaN/AlInN heterostructures 53
3.3.2 Quaternary structures: Unstrained barriers 55
3.3.3 Pseudomorphic barriers on a GaN substrate 59
3.3.4 Strain and miscibility limitations to built-in field suppression 61
3.3.5 UV devices: importance of the AlInN band gap bowing . . . 64

3.4 Summary 67

4 Elastic and structural properties of solids in the context of plane wave-based DFT 69
4.1 Introduction 69
4.2 Calculation of elastic constants and structural parameters 71
4.2.1 Harmonic elasticity 72
4.2.2 Internal strain in zinc-blende and wurtzite lattices 74
4.2.3 DFT implementation of total energy and stress tensor 79
4.3 Effects of basis set incompleteness and k-point sampling on results . 80
4.3.1 AlN as a test case material 81
4.3.2 Effect of the plane wave basis set on the calculation 82
4.3.3 Effect of the number of $\mathbf{k}$ points on the calculation 85
4.3.4 Effect of reduced crystal symmetry on the robustness of the total energy and stress methods 87
4.3.5 Additional results for selected materials 96
4.3.6 Correcting the total energy method: strain-adapted basis set 98
4.4 Elastic constants and structural parameters of group-III nitrides using the HSE hybrid functional scheme 99
4.5 Accuracy of theoretical calculations in the context of experimental results 103
4.6 Summary 104

5 Theory of local electric polarization and its relation to internal strain in group-III nitrides 107
5.1 Introduction 107
5.2 Theory of local electric polarization 109
5.2.1 Formal definition of the local polarization 111
5.2.2 Validity of the model 114
5.2.3 Testing the theory for group-III nitrides 117
5.3 Point dipole method for the calculation of the polarization potential 127
5.3.1 From polarization to dipole moment 127
5.3.2 Solution for materials with different dielectric constant . . . 129
5.3.3 Comparison to the solution of Poisson's equation for simple structures 130
5.3.4 Computational aspects: method of layers and application to quantum wells 133
5.4 Selected results for InGaN quantum wells 134
5.5 Summary 136

6 Staggered-grid formulation of a continuum elasticity method for efficient structural relaxation of zinc-blende and wurtzite crystal

lattices **139**
6.1 Introduction . . . 139
6.2 General form of the strain on a discrete mesh . . . 140
6.2.1 Forward and backward finite differences . . . 140
6.2.2 Formulation of the strain in a staggered grid . . . 143
6.3 Elastic energy minimization . . . 147
6.3.1 Form of the elastic energy in discrete continuum elasticity theory . . . 147
6.3.2 Analytical reformulation of the condition of minimum . . . 150
6.3.3 Retrieving the strain from the displacement field . . . 152
6.3.4 Implementation in rotated (111)-oriented zinc-blende systems 153
6.4 Testing the method . . . 154
6.4.1 Performance . . . 155
6.5 Summary . . . 157

7 Conclusions **161**
7.1 Summary . . . 161
7.2 Ongoing and future work . . . 163

8 Bibliography **167**

Appendices

Appendix A Additional material on local polarization and the point dipole method **187**
A.1 Example of the calculation of a local piezoelectric coefficient . . . 187
A.2 Point dipole solution for the three-dielectric problem . . . 188
A.3 Gaussian smearing of point dipoles . . . 191
A.4 Computational tricks . . . 193

Appendix B Expression for the derivative of the total energy in the staggered grid **197**

Declaration

I hereby declare that, unless otherwise stated, this work is my own, and that it has not been submitted for another degree, either at University College Cork or elsewhere.

Cork, ...,

Miguel A. Caro

Summary

In this work, the properties of strained tetrahedrally bonded materials are explored theoretically, with special focus on group-III nitrides. In order to do so, a multiscale approach is taken: accurate quantitative calculations of material properties are carried out in a quantum first-principles frame, for small systems. These properties are then extrapolated and empirical methods are employed to make predictions for larger systems, such as alloys or nanostructures. We focus our attention on elasticity and electric polarization in semiconductors. These quantities serve as input for the calculation of the optoelectronic properties of these systems.

Regarding the methods employed, our first-principles calculations use highly-accurate density functional theory (DFT) within both standard Kohn-Sham and generalized (hybrid functional) Kohn-Sham approaches. We have developed our own empirical methods, including valence force field (VFF) and a point-dipole model for the calculation of local polarization and local polarization potential. Our local polarization model gives insight for the first time to local fluctuations of the electric polarization at an atomistic level.

At the continuum level, we have studied composition-engineering optimization of nitride nanostructures for built-in electrostatic field reduction, and have developed a highly efficient hybrid analytical-numerical staggered-grid computational implementation of continuum elasticity theory, that is used to treat larger systems, such as quantum dots.

A mi madre,
y a la memoria de mi padre.

Why is it that when one man builds a wall,
the next man immediately needs to know
what's on the other side?

GEORGE R. R. MARTIN

Preface

This thesis consists, for the most part, of a collection of published and unpublished manuscripts that have been the subject of my work as a PhD student in the Tyndall National Institute from April 2009 until June 2013, under the supervision of Prof. Eoin O'Reilly and Dr. Stefan Schulz. The published papers have been revised to be appropriate for this thesis, making small corrections with respect to the journal versions, replotting some of the figures, restructuring the sections, etc., but remain basically the same. Chapter 3 on built-in field control consists of Refs. [1, 2] and Chapter 4 on elastic properties of semiconductors has been compiled from the work contained in Refs. [3, 4], as well as some later additions that were presented at the Total Energy and Force Methods workshop in ICTP, Trieste in January 2013.

The unpublished manuscripts have been included in a format more suitable for a work of this kind than for final journal form, and will probably undergo further modifications that are normal to the peer-reviewing process. Chapter 5 on local polarization was motivated by the initial work on alloy fluctuations in InGaN alloys presented in Glasgow at the 9$^{\text{th}}$ International Conference on Nitride Semiconductors (ICNS9) in July 2011, summarized in a short paper later in 2012 [5]. The work now included here is however much deeper in scope, comprehensive and structurally formal, with very few inclusions taken directly from Ref. [5]. The manuscript to be submitted for journal publication will eventually include tight-binding calculations performed by Stefan Schulz on the electronic structure of InGaN alloys, particularly the band gap bowing, when the local strain and polarization potential are taken into account. Chapter 6 contains the theoretical foundations and some technical details behind my code for strain energy minimization in a staggered-grid formalism, HUSH (*Hybrid Utility for Strained Heterostructures*).[1]

[1]The original name I chose was HUSR, or *Hybrid Utility for Strained Relaxation*, which I decided to change after a meeting in which my supervisor misread the correct name for what sounded to me at the time as a much cooler one.

In addition to the research chapters, this thesis also contains the usual Introduction, where I motivate the work, and Conclusions, where I give an idea of its relevance and what are the immediate objectives related to this project. There is also an introductory chapter with details of some of the theory and methods used throughout the rest of the thesis that were not appropriate for the original papers. It contains, together with basic "textbook" theory, a practical guide to calculate polarization in the context of the Berry-phase technique.

I resort to a short abstract at the beginning of each of the chapters, in the fashion of journal papers. In my opinion this improves in readability given the fact that these chapters are, although closely related to each other, rather self-contained, and can be read on their own. An outline of the thesis is given at the end of the Introduction.

Credit is due as collaborating authors to Eoin O'Reilly and Stefan Schulz on all of my papers, to Sorcha Healy (who acted as my co-supervisor for a few months before she decided to get a real job) on the early work on field control, and to Andreas Amann on the staggered grids. Also to Oliver Marquardt with whom I collaborated on a number of articles on quantum dots with Stefan Schulz as the main author, where we employed HUSH for some of the calculations [6–8].

This is my humble contribution to science, my attempt to catch a glimpse of –in the words of George Martin quoted a couple of pages back– what lies *on the other side of the wall*. I feel privileged for having the opportunity to get paid for doing what I like, and can only wish for more exciting projects to come in the future. My hope is that, besides gathering dust on the bookshelf of a proud mother who cannot understand a word of what her son wrote (even if it was written in Spanish), this work will be of use to someone, somewhere. If that is the case, don't forget to cite!

Cork, June 24, 2013

Miguel A. Caro

Acknowledgments

Many thanks are due to many people for many reasons. Since, after all, this aspires to be a scientific work, I will start by thanking that part of the bunch.

As jokingly recommended by himself, I start the acknowledgments by thanking "first and foremost" my supervisor Eoin O'Reilly. He has the (sometimes annoying) habit of being always right, and an extremely keen eye for spotting other people's mistakes, which in my particular case has saved me from self-embarrassment on a number of papers and presentations. But as much as I have to thank him for his guidance as a great scientist, I also have to thank him for being a great boss: for letting me follow my own naive ideas, and for letting me worry only about science.

Next, I thank Stefan Schulz. His constant support, availability, dedication and systematic scrutiny of my work have been a motivation to me. But most of all I want to thank him for, since day one, having treated me as a colleague, rather than the helpless PhD student that I actually was.

I would also like to thank a number of people who have spared some of their expertise and time to assist me with my work. Thanks to Morten Willatzen and Peter Parbrook for agreeing to be my thesis examiners. Thanks again to Peter for helping me get through with the publication of my first paper on field control. To Andreas Amann for help with the staggered grids and a number of mathematical enquiries I have made him throughout my time in Tyndall. My thanks to Martin Friák and Richard Needs for assistance on the convergence issue with DFT total energy elasticity calculations. Thanks are also due to David Vanderbilt and Vincenzo Fiorentini for very helpful email correspondence on the practicalities of Berry-phase calculations.

Thanks as well to some of the guys in the office at Tyndall on many discussions over these four years, both regarding science and pretty much everything else: Conor Coughlan, Chris Broderick, Paudie Scanlon, Patrycja Heinricht, Nicola Brandonisio, Masoud Seifikar, Shijie Deng, John Buckeridge and Ian Math-

ews. Special thanks to Eoin Clerkin, who introduced me to Bash scripting and eventually made a huge positive impact on my productivity.

My gratitude goes to the people running the computer cluster at Tyndall: Mick Nolan for help compiling VASP and John Kissane from IT. Also to Finbarr O'Callaghan and Mark Szepieniec for discussions on cluster stuff.

I also acknowledge Science Foundation Ireland for funding my scholarship and paying the bills, including those at the pub.

Thanks to the people developing some of the tools I have used during my work, particularly `gnuplot`, VASP, `Mathematica`, LaTeX and Ti*k*Z. Special thanks to Martijn Marsman, who wrote the VASP Berry-phase module, for being kind and patient enough to answer a few stupid questions I asked him over the email.

Getting through four years of PhD requires a lot more than good scientific interaction. On the more "social" side of things I have of course to thank my mum, who provided an endless collection of cheese, ham and sausages to keep me well fed, and sent them over from Spain either by plane or post. And to all my family back home. Thanks to the people who came to visit me during my time in Cork, my cousins Javi and Juan Redondo, my sister Cristina, and my friends Silvia Melo, Unai Gallastegui, Benedetta Vulcani, Annina Scherrer and Miguel Navarro.

Thanks to the friends from Tyndall, to the guys in the Sports & Social Committee, to all the guys that showed up for the S&S soccer tournaments I organized, to Mahdi Shirazi, to the Nemo guys for many evening soccer games and to the UCC handball team. To Coco for the good moments together. And of course big thanks to my poker team and dear friends, for many games of Texas Hold'em for which we have actual statistics running over a period of four years. They are David Bitauld, Julien Dufresne, Mark Szepieniec, Shane McDermott, Finbarr O'Callaghan and Nitin Deepak, with whom I also shared an apartment for quite some time. Also thanks to all the guys that played at some stage or joined the poker squad permanently: Alfonso Sánchez (who I knew from college in La Laguna), Shane O'Mahony (the New Shane), Michel Minot, Joonhwa Cho (Korea!!!), Ciaran Murray, Tom Sadler, and many more. To them I dedicate figure "zero" of this thesis which, to tell the truth and not without a bit of shame, required a lot more work and time (and beer!) to compile than any other in this thesis. Thanks also to all the guys that came to the Abbott Ale House or elsewhere for pints after work and I have not mentioned yet: Siobhan O'Halloran, Thomas Kelly, Anna Hauber, Amélie Wahl and Sylwia Klejna. And to everyone I might have forgotten.

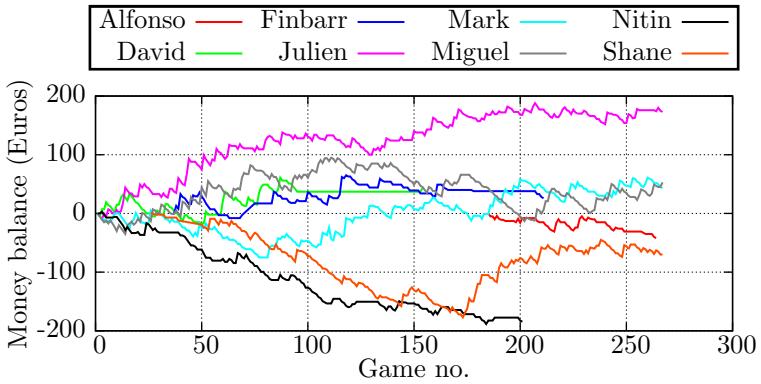


Figure 0: Poker statistics after more than four years of Texas Hold'em in Cork.

Finally but most importantly, a warm thank you to my lovely Veera. For being there to share the good moments and for supporting me during the bad ones. Thank you for everything you give me, which is free but priceless. I came to Cork to get a PhD in Physics, and it now feels like a wooden spoon compared to this other winning prize I am taking home.

List of publications

This is a complete list of all the publications that derived from the work carried out as part of this thesis project, either as main or collaborating author. The list includes all the peer-reviewed and conference papers.

Peer-reviewed articles

- **M. A. Caro, S. Schulz and E. P. O'Reilly**
  Comparison of stress and total energy methods for calculation of elastic properties of semiconductors
  J. Phys.: Condens. Matter 25, 025803 (2013)

- **S. Schulz, M. A. Caro, and E. P. O'Reilly**
  Prediction of strong ground state electron and hole wave function spatial overlap in nonpolar GaN/AlN quantum dots
  Appl. Phys. Lett. 101, 113107 (2012)

- **M. A. Caro, S. Schulz, and E. P. O'Reilly**
  Hybrid functional study of the elastic and structural properties of wurtzite and zinc-blende group-III nitrides
  Phys. Rev. B 86, 014117 (2012)

- **M. A. Caro, S. Schulz, and E. P. O'Reilly**
  Effect of alloy fluctuations on the local polarization in nitride nanostructures
  Phys. Stat. Sol. B 249, 526 (2012)

- **S. Schulz, M. A. Caro, E. P. O'Reilly, and O. Marquardt**
  Piezoelectric properties of zinc blende quantum dots
  Phys. Stat. Sol. B 249, 521 (2012)

- **M. A. Caro, S. Schulz, S. B. Healy, and E. P. O'Reilly**
  Built-in field control in nitride nanostructures operating in the UV
  Phys. Stat. Sol. C 9, 838 (2012)

- **S. Schulz, M. A. Caro, E. P. O'Reilly, and O. Marquardt**
  Symmetry-adapted calculations of strain and polarization fields in (111)-oriented zinc-blende quantum dots
  Phys. Rev. B 84, 125312 (2011)

- **M. A. Caro, S. Schulz, S. B. Healy, and E. P. O'Reilly**
  Built-in field control in alloyed c-plane III-N quantum dots and wells
  J. Appl. Phys. 109, 084110 (2011)

Conference talks

- **M. A. Caro, S. Schulz, and E. P. O'Reilly**
  Direct relation between internal strain and piezoelectric response of ordered and disordered III-N crystal structures
  UK Semiconductors 2013, July 3$^{\text{rd}}$ – 4$^{\text{th}}$ 2013, Sheffield, UK

- **S. Schulz, M. A. Caro, and E. P. O'Reilly**
  Control and Engineering of Spontaneous and Piezoelectric Polarisation in Nitride-based Nanostructures (invited)
  International Conference on Metallurgical Coatings and Thin Films, Apr. 23$^{\text{rd}}$–27$^{\text{th}}$ 2012, San Diego, CA, USA

- **M. A. Caro, S. Schulz, and E. P. O'Reilly**
  Effect of the local polarization in nitride nanostructures
  TMCS III, Jan. 18$^{\text{th}}$ – 20$^{\text{th}}$ 2012, Leeds, UK

- **S. Schulz, M. A. Caro, and E. P. O'Reilly**
  Polarization switching in semipolar InGaN/GaN quantum wells: The role of the deformation potentials
  UKNC Annual Meeting 2012, Jan. 4$^{\text{th}}$ – 5$^{\text{th}}$ 2012, Bath, UK

- **M. A. Caro, S. Schulz, and E. P. O'Reilly**
  Effect of alloy fluctuations on the local polarization in nitride nanostructures
  Photonics Ireland 2011, Sept. 7$^{\text{th}}$ – 9$^{\text{th}}$ 2011, Malahide, Ireland

- **M. A. Caro, S. Schulz, and E. P. O'Reilly**
  Effect of alloy fluctuations on the local polarization in nitride nanostructures
  ICNS9, July 10$^{\text{th}}$ – 15$^{\text{th}}$, Glasgow, UK

- **M. A. Caro, S. Schulz, S. B. Healy and E. P. O'Reilly**
  Nitride alloys: Study of nonlinearities in structural and elastic properties using an interatomic potential model
  UK Semiconductors 2010, July 7$^{\text{th}}$ – 8$^{\text{th}}$ 2010, Sheffield, UK

- **M. A. Caro, S. Schulz, S. B. Healy and E. P. O'Reilly**
  Built-in field control in wurtzite III-N QDs and QWs
  UKNC Annual Meeting 2010, Jan. 12$^{\text{th}}$ – 13$^{\text{th}}$ 2010, Cork, Ireland

Conference posters

- **M. A. Caro, S. Schulz, and E. P. O'Reilly**
  Role of Internal Strain on the Electromechanical Properties of Wurtzite Nitrides: The Local Piezoelectric Tensor
  ICNS10, Aug. 25$^{\text{th}}$ – 30$^{\text{th}}$ 2013, Washington DC, USA

- **M. A. Caro, S. Schulz, and E. P. O'Reilly**
  Comparison of stress and total energy methods for the calculation of elastic properties of semiconductors
  16$^{\text{th}}$ International Workshop on Computational Physics and Materials Science: Total Energy and Force Methods, Jan. 10$^{\text{th}}$ – 12$^{\text{th}}$ 2013, Trieste, Italy

- **S. Schulz, O. Marquardt, M. A. Caro, and E. P. O'Reilly**
  Electronic properties of site-controlled (111)-oriented InGaAs quantum dots: A symmetry adjusted approach
  Quantum Dot 2012 Conference, May 13$^{\text{th}}$ – 18$^{\text{th}}$ 2012, Santa Fe, NM, USA

- **M. A. Caro, S. Schulz, S. B. Healy and E. P. O'Reilly**
  Built-in field control in nitride nanostructures operating in the UV
  ICNS9, July 10$^{\text{th}}$ – 15$^{\text{th}}$, Glasgow, UK

- **M. A. Caro, S. Schulz, S. B. Healy and E. P. O'Reilly**
  Nitride alloys: Study of nonlinearities in structural and elastic properties

using an interatomic potential
Fabrication to Application School, Oct. 5^{th} – 7^{th} 2010, Cork, Ireland

- **M. A. Caro, S. Schulz, S. B. Healy and E. P. O'Reilly**
  An interatomic potential model for the study of the elastic and structural properties of nitride alloys
  Empirical methods in semiconductor nano-structures design and modelling, June 21^{st} – 25^{th} 2010, Dublin, Ireland & Manchester, UK

- **M. A. Caro, S. Schulz, S. B. Healy and E. P. O'Reilly**
  Built-in field control in wurtzite III-N QDs and QWs
  TMCS II, Jan. 13^{th} – 15^{th} 2010, York, UK

- **M. A. Caro, S. Schulz, S. B. Healy, and E. P. O'Reilly**
  Built-in field control in c-plane nitride systems
  Photonics Ireland 2009, Sept. 24^{th} – 26^{th} 2009, Kinsale, Ireland

1 Introduction

The group-III nitrides, or more correctly group-IIIB nitrides, comprise the compounds of nitrogen (N) and the elements in the first column of the *p*-element block of the periodic table, boron (B), aluminium (Al), gallium (Ga), indium (In) and thallium (Tl).[1] The most technologically important of these compounds are GaN, InN, AlN and their alloys. BN presents structural and electronic properties that do not follow the trends of GaN, InN and AlN, and is most interesting in solid-state physics for nanotube applications [15, 16]. There is a possibility that TlN could be used alloyed with the other nitrides given some similarities in their electronic properties [17]. However, perhaps partly because Tl is extremely toxic and its use in industry or research raises health issues [18], there exist very few studies on the properties and potential uses of TlN. Therefore, whenever we use the terminology *group-III nitrides* throughout this thesis, we implicitly refer to the binaries GaN, InN and AlN, the ternaries InGaN, AlGaN and AlInN, and the quaternary AlInGaN.

The use of nitrides as versatile semiconductor materials has driven significant advances over the last two decades in the field of optoelectronic applications. This is mainly due to the fact that the wurtzite (WZ) III-N materials are all direct band gap semiconductors. By alloying GaN ($E_g = 3.43$ eV), InN ($E_g = 0.64$ eV) and AlN ($E_g = 6.14$ eV) one gains potential access to the whole visible spectrum, as well as near infrared (IR) and ultraviolet (UV) [21], as we show in Fig. 1.1 in a direct comparison with other III-V materials. Although zinc-blende (ZB) nitrides might in principle appear as an alternative to their WZ counterparts, the truth is that growth of ZB nitrides is complicated and single-phase crystals are

[1] The *other* group-III nitrides are compounds of N and the transition metals in group-IIIA, mainly scandium (Sc) and yttrium (Y). Other transition metals in group-IIIA are rather more exotic and include some radioactive elements: lanthanum (La), lutetium (Lu), actinium (Ac) and lawrencium (Lr). There are ongoing investigations on the properties of Sc-containing nitrides, that present interesting features in the context of *functional materials* research [9–14].

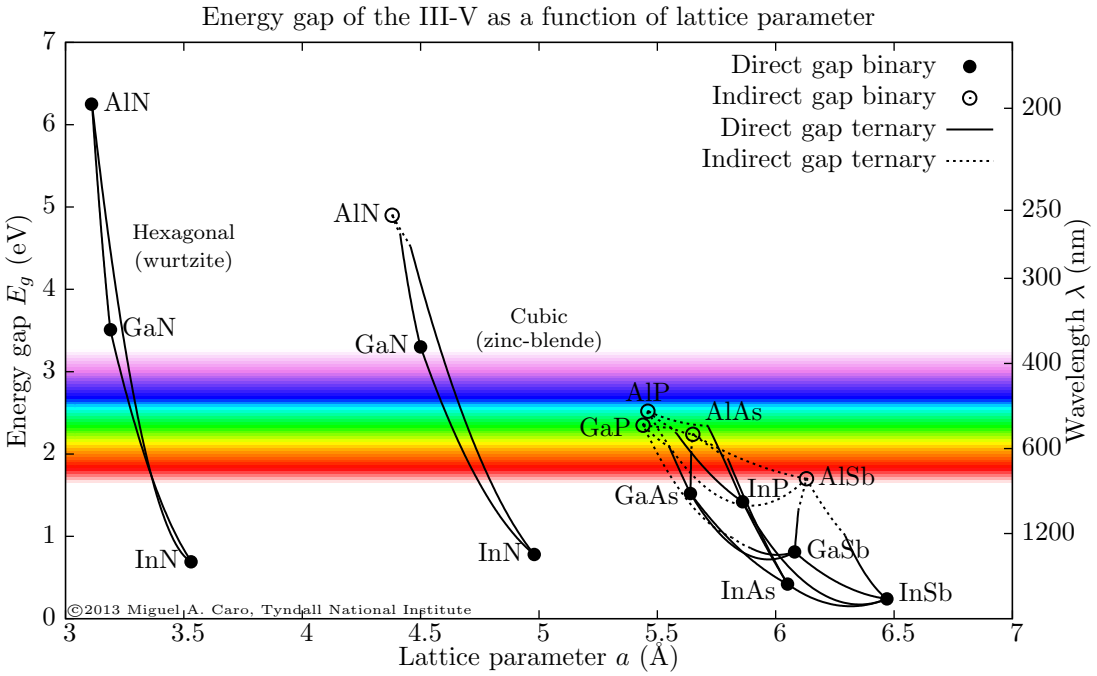


Figure 1.1: Band gap E_g of the technologically important III-V binaries as a function of lattice parameter (in-plane lattice parameter in the case of WZ nitrides). Solid circles indicate a direct gap material while empty circles indicate an indirect gap. The curves show the band gap for some of the ternaries, where solid lines indicate direct gap and dashed lines indicate indirect gap. The data for the cubic III-Vs has been taken from Ref. [19], except for ZB nitrides, which is from Ref. [20]. The data for the WZ nitrides is from Wu's review paper [21]. The band gap of the ZB ternaries has been calculated as the minimum of the gaps at Γ, X and L from the respective band gap bowing parameters provided in Refs. [19, 20], using Eq. (3.8). The variation of E_g with composition for the WZ nitride ternaries has been calculated using the same equation from the bowing parameters recommended by Wu [21]. See Chapter 3 for a discussion on III-N bowing parameters, AlInN in particular. A linear interpolation (Végard's law, see Ch. 3) is assumed for the lattice parameters of the ternaries. All data is for the zero-temperature limit $T = 0$.

very difficult to achieve [22]. There is no other material system that allows light emission over such a wide spectral range, while at the same time maintaining a direct gap. GaN-based heterostructures have been successfully employed in the commercial fabrication of violet and blue light-emitting diodes (LEDs) [23,24] and laser diodes (LDs) [25]. Other emerging applications of the nitrides include the extension of LEDs towards longer [26] and shorter [27] wavelengths, and their use in high electron mobility transistors (HEMTs) [28] and as multijunction high-efficiency solar cells [29]. Group-III nitrides emerged during the last decade of the twentieth century as a solution to the production of blue and green light. Shuji Nakamura, at Nichia Corporation in Japan, worked on the development of GaN-based light-emitting diode (LED) and laser diode (LD) technology from 1989 until 1999 [25]. This allowed Nichia to produce and commercialize blue LEDs (1993), green LEDs (1995) and violet LDs (1999) [25].

However, and in spite of their evident success, there are still dark areas in the field of nitride research, and the booklet of properties of these materials is ever expanding. This is well illustrated by the fact that Vurgaftmann and Meyer published a revised version of their review, for the nitrides only [20], scarcely two years after the publication of their almost-biblical anthology on band parameters of the III-Vs [19]. A quantity as fundamental as the energy gap of InN was revised between the two versions by more than a factor of two, from 1.99 eV to 0.78 eV. It seems that its zero-temperature value has been finally established at around 0.69 eV [21]. In this context, the (eventually enormous) success of the Nichia/Nakamura tandem is very striking: the most suitable substrate for GaN, sapphire (Al_2O_3), presents a lattice mismatch of 16%, meaning that GaN grown on sapphire exhibits a large dislocation density. Dislocation densities of the order of 10^9 cm^{-2} are typical in InGaN quantum wells (QWs) [30], with lower –but still relatively high– values achievable by techniques such as epitaxial lateral overgrowth (ELOG) [31]. However, dislocations act, in principle, as non-radiative recombination centres and dislocation densities as low as 10^3 cm^{-2} are known to quench light emission in other III-Vs, such as GaAs [24, 26, 32]. To date, the reason or reasons behind this unexpectedly high efficiency of GaN-based devices remain a matter of debate. Some researchers have attributed it to the presence of alloy fluctuations, In-clustering and the associated local variations in the carrier confinement potential [33,34], while others have suggested QW width fluctuations as the origin of the observed high efficiency [32].

Other important issues regarding growth of nitrides with high In content (including InN itself) are strain and miscibility. High In content alloys are needed in order to extend device operation energies towards longer wavelengths. While the ternary AlGaN can be grown practically at any desired composition, there are severe limitations as to how much In can be incorporated into InGaN and AlInN alloys [35,36]. Nitrides are usually grown by either molecular beam epitaxy (MBE) or metal-organic vapour-phase epitaxy (MOVPE). MOVPE was the technique originally used by Nakamura at Nichia, and has since remained the method of choice because of advantages regarding device efficiency and growth rate, compared to MBE [37, 38].[2] InN and GaN have very different optimal growth temperatures, and InGaN alloys are grown on GaN templates, which means, given the lattice mismatch between InN and GaN (see Fig. 1.1), that the strain in the InGaN layer will significantly increase as the In molar fraction increases (we give a brief account of these limitations and the current capabilities regarding In incorporation in Section 3.3.4).

One of the main problems that are encountered when dealing with nitride heterostructures, and one that severely affects device efficiency, is the existence of built-in electrostatic fields. Although the other III-V materials also present piezoelectric properties due to their cubic ZB structure, the WZ crystal structure in which III-N materials crystallize is compatible with the existence of both spontaneous and piezoelectric polarization. Additionally, the increased ionicity of the III-N compared to the other III-V accounts for a much stronger effect. The discontinuity of the polarization vector between two nitride materials that differ in composition and/or strain state leads to a strong polarization potential, with effects on the electronic properties of these systems.

The study of strain effects in III-N materials, including electric polarization in particular, is the central theme of this thesis. The first studies that we carried out, and that eventually motivated the rest of this work, were on polarization field control in nitride QWs and quantum dots (QDs). They are presented in Chapter 3, where we propose composition engineering as a route towards suppression (or reduction) of the strong built-in electrostatic fields present in III-N QWs and QDs. During the time in which that work was carried out, we encountered several aspects of the description of nitride materials that, in our opinion, needed improvement.

[2]More recently, *plasma-assisted* MBE seems to have taken a leap forward and is improving in quality upon more traditional MBE techniques [38].

Specifically, the description of the electric polarization in nitride alloys previously available (and used in our calculations of Chapter 3) was based on the assumption that its value only depends on macroscopic strain and average composition in the sample. However, we challenge this simplified virtual crystal model and show that the local atomic landscape matters and has a large impact on the computation of average properties.

In this context, we build a comprehensive description of the electric polarization in nitrides from the bottom up. In Chapter 4 we study elastic properties and internal strain of tetrahedrally bonded materials under deformation. Then, in Chapter 5 we construct our theory of local polarization in crystalline solids and link it to macroscopic and internal strain. We have also developed computational methods for the calculation of the corresponding local polarization potential in nanostructures with a large number of atoms. Our model is based on and validated against calculations performed in the frame of first-principles density functional theory (DFT) and the Berry-phase formalism, which are outlined in Chapter 2.

As discussed through Chapters 4 and 5, an atomistic description of material systems is always preferable in terms of capturing the discrete nature of matter. However, it is not always practical to adopt this level of theory because of computational considerations. For example, the optimization procedure that we carry out in Chapter 3 would become prohibitively expensive if performed at the atomic level. Moreover, in some cases, such as nitride systems where alloys are not present (e.g. GaN/AlN QDs [7]), the atomistic effects are of secondary importance and continuum calculations provide additional insight into material or device characteristics. Therefore, the development of efficient strategies to treat strain at the continuum level is vital in the context of simulation of device-sized structures that contain of the order of millions of atoms, such as QDs. In Chapter 6 we present an efficient hybrid (analytical/numerical) implementation of continuum elasticity theory for the calculation of strain fields in device-size crystal structures. Our method shows good performance improvement upon purely numerical methods, without compromising accuracy.

Finally, the conclusions on the present work and prospects for future and ongoing projects are outlined in Chapter 7.

2 Theory and methods

The scope of this chapter is to provide the theoretical background to some of the methods, principles and techniques that are employed throughout the remainder of this thesis. Some of these are, because of practical considerations, not explained or introduced in the pertinent chapters.

This thesis is devoted to the study of the properties of deformed crystals. Therefore, in Section 2.1 we introduce the fundamental formulas and definitions involved with some of the properties of strained crystals: elasticity and piezoelectricity in crystalline solids, and the definition of strain itself.

In Section 2.2 we present a brief overview of density functional theory (DFT), as well as the different approximations commonly in place for realistic calculations: functionals for exchange and correlation, pseudopotentials and plane waves.

Finally, in Section 2.3 we give an introduction to the modern theory of polarization, based on the Berry-phase technique. Since the Berry-phase technique has been developed in the context of *ab initio* calculations, we believe it is more appropriate to introduce it after the DFT section rather than as a part of the general piezoelectricity discussion of Section 2.1.

2.1 Strain effects in periodic solids

A crystalline solid is characterized by a periodic repetition of a set of interacting atoms that adopt a particular spatial configuration. The specific arrangement in which these atoms are positioned is referred to as the crystal lattice [39]. Experimental access to the crystal structure of a particular material can be obtained, for example, using x-ray diffraction techniques [39]. Theoretical determination of crystal structures for a specific mix of chemical elements can be carried out using total energy *ab initio* methods such as different implementations of density functional theory (Section 2.2). The latter are very useful in cases where the experimental techniques fail or where material samples are not available, or even to make structure predictions. Ultimately, the crystal structure for any given ensemble of atoms is given by their energetic interaction. The energy surface as a function of all of the atomic coordinates will present minima corresponding to stable (global minimum) or metastable (local minima) configurations given the physical conditions, e.g. of pressure or temperature. For vanishing external pressure, the global minimum of the energy surface corresponds to the equilibrium lattice structure. In group-III nitrides, as we have already discussed, the global minimum at equilibrium corresponds to the wurtzite (WZ) lattice and there exists also a local minimum that corresponds to the metastable zinc-blende (ZB) lattice.

Deviations from a crystal's equilibrium atomic positions introduce what is known as *strain*. The strain gives the displacement of the atoms from equilibrium (per unit length) as one moves along the different Cartesian directions [39, 40].[1] The strain, as defined in textbooks in terms of a homogeneous (or slowly varying between neighbouring unit cells) displacement field $\mathbf{u}$, is a macroscopic property of the crystal. For convenience, in this introduction we will work under that assumption, but always keeping in mind that this situation is just a simplification only valid for homogeneously deformed crystals. Inhomogeneous strains in a fully atomistic picture and a position-dependent continuum frame are treated in Chapters 5 and 6, respectively. Therefore, we forget about the atomic nature of matter and think about a volume element of our crystal as the solid box given in Fig. 2.1. Before deformation, the box is a perfect cube [Fig. 2.1 (left)]. The most general

[1] This means that the strain components will be in general reference frame dependent, given the different possible orientations of the crystal; see for instance Section 4.3.4 or our work on [111]-oriented ZB systems [6].

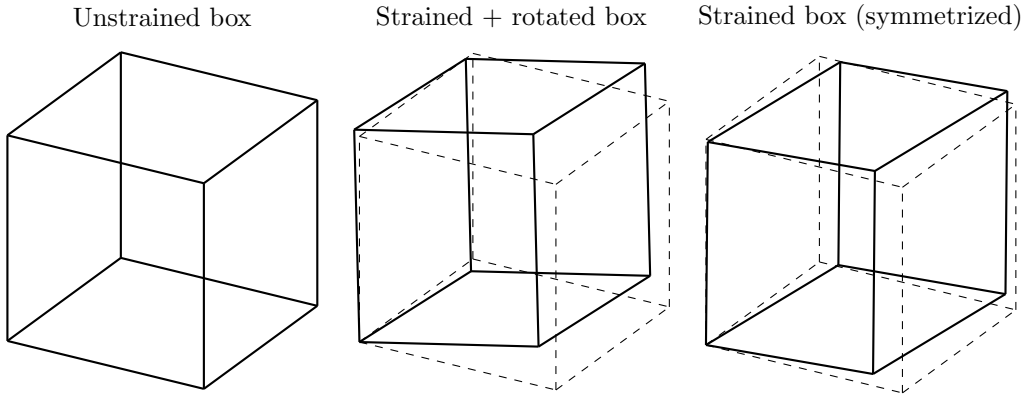


Figure 2.1: Box (cube) before and after random deformation, including also the effect of rotation.

transformation that can be applied to a solid is a change in shape followed by a rotation [Fig. 2.1 (middle)].[2] In such a general situation, the displacement of the material points contained in the box with respect to their original positions (which are represented by the cube of dashed sides) as one moves along the different Cartesian directions is described by the following derivative [39, 41]:

$$\tilde{\epsilon}_{ij} = \frac{\partial u_i}{\partial j}, \tag{2.1}$$

where i and j represent the usual Cartesian directions, characterized by the usual x, y, z. A material point in the box whose position before the transformation was $\mathbf{r} \equiv (r_x, r_y, r_z)$ has a new position $\mathbf{r}'$ given by

$$\begin{pmatrix} r'_x \\ r'_y \\ r'_z \end{pmatrix} = \begin{pmatrix} 1+\tilde{\epsilon}_{xx} & \tilde{\epsilon}_{xy} & \tilde{\epsilon}_{xz} \\ \tilde{\epsilon}_{yx} & 1+\tilde{\epsilon}_{yy} & \tilde{\epsilon}_{yz} \\ \tilde{\epsilon}_{zx} & \tilde{\epsilon}_{zy} & 1+\tilde{\epsilon}_{zz} \end{pmatrix} \begin{pmatrix} r_x \\ r_y \\ r_z \end{pmatrix}. \tag{2.2}$$

Although Eq. (2.2) is the most general transformation (except for a translation), rotations do not affect the *shape* of the box [Fig. 2.1 (right)]. In a crystal, it is the change in shape and volume that determines the distortions of interatomic distances and angles: the physical properties of the crystal are unaffected by rotations. Therefore, in order to define a physically meaningful property of the

[2]One can also apply a trivial translation which we will neglect for reasons that will become clear later on: a translation of the crystal as a whole does not show in the calculated derivatives of the displacement field.

crystal, we proceed to separate the transformation in Eq. (2.2) into its symmetric and antisymmetric components. The symmetric part

$$\epsilon_{ij} = \frac{1}{2}\left(\tilde{\epsilon}_{ij} + \tilde{\epsilon}_{ji}\right) \tag{2.3}$$

gives the change in shape, and it is known as the *strain*, whereas the antisymmetric part

$$\bar{\epsilon}_{ij} = \frac{1}{2}\left(\tilde{\epsilon}_{ij} - \tilde{\epsilon}_{ji}\right) \tag{2.4}$$

gives the rotational part of Eq. (2.2) [39]. We therefore define the strain tensor ϵ as

$$\epsilon = \begin{pmatrix} \epsilon_{xx} & \epsilon_{xy} & \epsilon_{xz} \\ \epsilon_{xy} & \epsilon_{yy} & \epsilon_{yz} \\ \epsilon_{xz} & \epsilon_{yz} & \epsilon_{zz} \end{pmatrix}, \tag{2.5}$$

which has only 6 independent components. Because of this, it is common practice in the context of solid-state physics to contract the indices so that they become $xx \to 1$, $yy \to 2$, $zz \to 3$, $yz \to 4$, $xz \to 5$ and $xy \to 6$. This simplification is known as "Voigt notation" [41] and will be extensively used throughout this thesis. In the case of the strain tensor, the exact correspondence is $\epsilon_1 = \epsilon_{xx}$, $\epsilon_2 = \epsilon_{yy}$, $\epsilon_3 = \epsilon_{zz}$, $\epsilon_4 = 2\epsilon_{yz}$, $\epsilon_5 = 2\epsilon_{xz}$ and $\epsilon_6 = 2\epsilon_{xy}$.

As mentioned, strain affects the crystal properties because it modifies the distances and angles between the interacting atoms in the crystal. In particular, strain will induce an increase in the crystal's energy since, as we have already discussed, the unstrained configuration corresponds to a global or local energy minimum.[3] Understanding the relation between total energy and strain is important because it allows to gain access to the *elastic* properties of the crystal. These determine how difficult it is to compress or expand a material, or perform any arbitrary shape deformation in general. They also permit to know how a material will respond to strain if it is allowed to relax in a particular direction (see for instance the discussion on pseudomorphic growth of nitride quantum wells in Chapter 3). The

[3]This statement is always true for *infinitesimal* strain. However, for finite strain it could happen that a material that undergoes a deformation along a particular path encounters other minima along that path (e.g. stress-induced phase transitions).

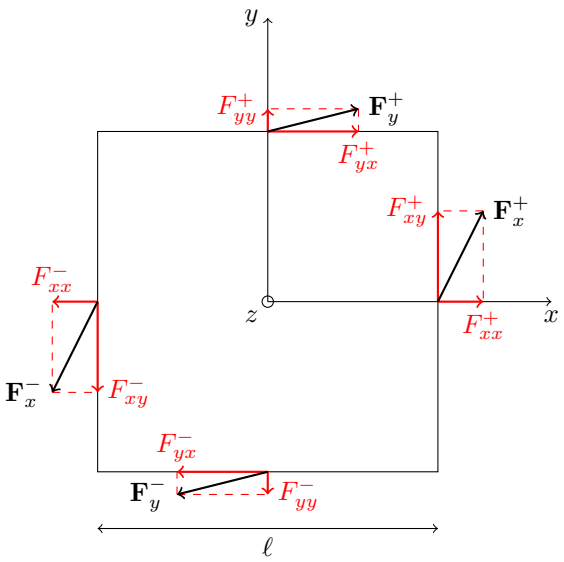


Figure 2.2: Section of the volume element V across the xy plane, indicating the (external) forces acting on it after deformation.

next section is therefore devoted to the elastic properties of crystalline solids.

2.1.1 Elasticity

When a volume element V of a crystal undergoes strain, there must be an internal force acting on it, related to the energetic states of the volume element before and after deformation. If the volume element is in mechanical equilibrium with its surroundings, the external forces acting on the element are equal to the internal ones. For simplicity, we consider in Fig. 2.2 a section of V perpendicular to z and assume that all components of the forces along z vanish (although in the following our equations will be general). We call the forces acting on the different faces $\mathbf{F}^+_x$, $\mathbf{F}^-_x$, $\mathbf{F}^+_y$ and $\mathbf{F}^-_y$, as indicated in the figure. If the volume element is in mechanical equilibrium, the forces acting on opposite faces must be of opposite sign, therefore $\mathbf{F}^+_x = -\mathbf{F}^-_x \equiv \mathbf{F}_x$ and $\mathbf{F}^+_y = -\mathbf{F}^-_y \equiv \mathbf{F}_y$. We define the components of these vectors as

$$\begin{aligned} \mathbf{F}_x &\equiv (F_{xx}, F_{xy}, F_{xz})\,, \\ \mathbf{F}_y &\equiv (F_{yx}, F_{yy}, F_{yz})\,, \\ \mathbf{F}_z &\equiv (F_{zx}, F_{zy}, F_{zz})\,, \end{aligned} \tag{2.6}$$

where in our simplified example $F_{xz} = F_{yz} = 0$ and $\mathbf{F}_z = \mathbf{0}$. Furthermore, the torque due to the forces acting on the centre of the faces $\boldsymbol{\tau} = \ell\,(F_{xy} - F_{yx})\,\hat{\mathbf{z}}$ (where ℓ is the side length of the cube and $\hat{\mathbf{z}}$ is a unit vector along z) must also vanish. Hence, $F_{xy} = F_{yx}$. We now make use of Hooke's law [39], which describes the restoring force[4] for small displacements of an elastic material as $\mathbf{F} = \kappa\,\delta\mathbf{r}$, where κ is a tensor of constants and $\delta\mathbf{r}$ represents a small deformation. As we have discussed, any crystal deformation is perfectly defined by the strain tensor. Therefore if we allow each of the components of $\mathbf{F}_x$, $\mathbf{F}_y$ and $\mathbf{F}_z$ to depend upon all of the components of the strain tensor, the rank of κ is four. This is the most general form of Hooke's law for three-dimensional solids [39, 41]. In this context, we can write the forces in terms of the strains as

$$F_{ij}^{\pm} = \pm \underbrace{\sum_{kl} C_{ijkl}\epsilon_{kl}}_{\sigma_{ij}}\ \ell^2, \tag{2.7}$$

where we have rescaled κ so our new tensor C is given in units of force per unit area. Since the forces act on the *faces* of our volume element (that is, on a surface), this ensures that C does not depend on the size of V and is therefore a well defined bulk property of the crystal. The C_{ijkl} are the components of the *stiffness tensor* or tensor of *elastic constants*, and σ_{ij} is the *stress tensor*. It follows from our discussion of torque that the stress tensor is symmetric. The work done by $F_{ij}^{\pm}$ upon an infinitesimal deformation from the unstrained structure is

$$\mathrm{d}W^{\pm} = \pm \sum_{ij} F_{ij}^{\pm}\,\frac{\ell}{2}\,\mathrm{d}\epsilon_{ij}, \tag{2.8}$$

with the total work, or *elastic energy*, being $\mathrm{d}E = \mathrm{d}W^{+} + \mathrm{d}W^{-}$. This can be readily written in terms of the stress tensor σ_{ij} as

$$\mathrm{d}E = V \sum_{ij} \sigma_{ij}\,\mathrm{d}\epsilon_{ij}, \tag{2.9}$$

[4] In our case, we are actually interested in the force *opposite* to the restoring force.

where $V = \ell^3$ is the volume of the box. Alternatively, Eq. (2.9) can be written in terms of the stiffness tensor C_{ijkl} as

$$\mathrm{d}E = V \sum_{ijkl} C_{ijkl}\epsilon_{kl}\,\mathrm{d}\epsilon_{ij}. \tag{2.10}$$

Differentiating with respect to the strains, we arrive at [41]

$$\frac{1}{V}\frac{\partial}{\partial\epsilon_{kl}}\frac{\partial E}{\partial\epsilon_{ij}} = C_{ijkl}. \tag{2.11}$$

The fact that one can obtain C_{ijkl} as the second derivative of the energy with respect to ϵ_{ij} and ϵ_{kl} means that the stiffness tensor C_{ijkl} is symmetric upon exchange of $ij \leftrightarrow kl$. Integrating Eq. (2.11) making use of that fact, one arrives at the expression for the strain energy for small deformations [41]:[5]

$$E = \frac{V}{2}\sum_{ijkl} C_{ijkl}\epsilon_{ij}\epsilon_{kl}. \tag{2.12}$$

Further expressions for the elastic energy in Voigt notation are given in Chapter 4. In that chapter we deal with how to perform calculations for the C_{ijkl} in the context of density functional theory. Additionally, a detailed description of internal strain, which has not been reviewed in this section, is given.

2.1.2 Piezoelectricity

Piezoelectricity, which originates from the Greek words "*piezen*" (to *squeeze*) and "*elektron*" (*amber*, which has electrostatic properties) refers to the property of certain crystals to exhibit an electric polarization as a response to deformation. As a macroscopic quantity, the piezoelectric vector $\mathbf{P}^{\mathrm{pz}}$ can be related (to first order) to the stress and strain tensors via the piezoelectric moduli d_{ijk} and piezoelectric coefficients e_{ijk}, respectively [41]:

$$\begin{aligned} P_i^{\mathrm{pz}} &= \sum_{jk} d_{ijk}\sigma_{jk} \\ &= \sum_{jk} e_{ijk}\epsilon_{jk}. \end{aligned} \tag{2.13}$$

[5]The integration of Eq. (2.11) is most straightforward when performed in Voigt notation, because the variables in the integrand (the strains) are explicitly independent [41].

It is explicit in Eq. (2.13) that different deformations of the crystal give rise to different piezoelectric responses. The general form of the response can be deduced based on the symmetries of the crystal. For instance, crystals with an inversion centre cannot display this effect [41]. All the non-centrosymmetric crystal structures can in principle exhibit piezoelectricity. Nye [41] obtained all the non-zero elements of the linear piezoelectric tensor and established their equivalences. The most relevant for this work are the piezoelectric tensors of the wurtzite (WZ) and zinc-blende (ZB) crystal lattices. In Voigt notation, they are:

$$e_{ij}^{\mathrm{WZ}} \equiv \begin{pmatrix} 0 & 0 & 0 & 0 & e_{15} & 0 \\ 0 & 0 & 0 & e_{15} & 0 & 0 \\ e_{31} & e_{31} & e_{33} & 0 & 0 & 0 \end{pmatrix},$$
$$e_{ij}^{\mathrm{ZB}} \equiv \begin{pmatrix} 0 & 0 & 0 & e_{14} & 0 & 0 \\ 0 & 0 & 0 & 0 & e_{14} & 0 \\ 0 & 0 & 0 & 0 & 0 & e_{14} \end{pmatrix}. \tag{2.14}$$

More recently, Grimmer [42] has found all the non-zero elements of the second-order piezoelectric tensor. Here, we will only mention the existence of second-order piezoelectricity without going into further detail. Its importance in the context of III-V materials is briefly commented in Chapter 5.

In this section we have given a very basic description of piezoelectricity which mostly consists of introducing the formula linking it to strain. Details regarding the calculation of the polarization vector and the meaning of polarization itself, in the context of the "modern theory of polarization", are given in Section 2.3. Phenomena arising from the piezoelectric properties of crystals manifest themselves strongly in heterostructures, where discontinuities of the polarization vector across surfaces separating different materials lead to polarization charge accumulation and associated electrostatic fields. We extensively deal with these issues for group-III nitrides in Chapter 3.

2.2 Density functional theory

> *The fundamental tenet of density functional theory is that any property of a system of many interacting particles may be viewed as a functional of the ground state density [...]. The existence proofs for such functionals [...] are disarmingly simple. However, [...] no exact functionals are known for any system of more than one electron. DFT would remain a minor curiosity today if it were not for the ansatz made by Kohn and Sham [...].*
>
> R. M. Martin [43]

Density functional theory (DFT) constitutes one of the biggest achievements of physics during all of the twentieth century. Its success is due to the ability to make quantitative predictions about many of the properties of matter *ab initio*, that is, without experimental input. This allows access to properties that are difficult to measure in a laboratory, modelling of materials that have not been grown or for which high-quality samples are not available, and complementing and explaining experimental results. For instance, the description of internal strains and electric polarization conducted in Chapters 4 and 5 of this work can hardly be achieved in the lab with today's technical capabilities. Although current implementations of DFT are approximations, and there are several notable failures such as the *band gap problem* [44], the Hohenberg-Kohn theory is exact at heart, as summarized by Richard Martin in the quote above. In this chapter we will introduce the basic theoretical concepts of DFT that concern some of the successive chapters of this thesis: standard Kohn-Sham DFT and the local density and generalized gradient approximations (LDA and GGA, respectively); generalized Kohn-Sham DFT and hybrid functionals to overcome the band gap problem; and practical implementations of DFT in a plane-wave formalism and its implications for the obtained results. Finally we will give a brief overview of the package we have used to perform our *ab initio* calculations: VASP.

2.2.1 Kohn-Sham DFT

We will present a basic overview of the Kohn-Sham formulation of DFT that is largely based on the text by Martin [43].[6] The starting point for any method to calculate the electronic structure of matter is to write down the total energy of the system. In a non-relativistic system made up of an arbitrary number of interacting electrons and nuclei, the Hamiltonian operator can be written [43]:[7]

$$\begin{aligned}\hat{H} = &-\frac{\hbar^2}{2m_e}\sum_i \nabla_i^2 - \frac{1}{4\pi\epsilon_0}\sum_{i,I}\frac{Z_I e^2}{|\mathbf{r}_i - \mathbf{R}_I|} + \frac{1}{8\pi\epsilon_0}\sum_{i\neq j}\frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} \\ &-\frac{\hbar^2}{2M_I}\sum_I \nabla_I^2 + \frac{1}{8\pi\epsilon_0}\sum_{I\neq J}\frac{Z_I Z_J e^2}{|\mathbf{R}_I - \mathbf{R}_J|},\end{aligned} \tag{2.15}$$

where lower case subscripts denote electrons and upper case denote nuclei. Z_I denotes the nuclear charge in units of the elementary charge e. All the other symbols are standard. In accordance with Martin's notation [43], we rewrite Eq. (2.15) in a more compact form:

$$\hat{H} = \hat{T} + \hat{V}_{\text{ext}} + \hat{V}_{\text{int}} + \hat{T}_{\text{nuc}} + E_{II}. \tag{2.16}$$

The first term in Eq. (2.16) $\hat{T}$ is the kinetic operator for the electrons; $\hat{V}_{\text{ext}}$ is the attractive electrostatic potential acting on the electrons due to the nuclei; $\hat{V}_{\text{int}}$ gives the repulsive electron-electron electrostatic interaction; $\hat{T}_{\text{nuc}}$ is the kinetic operator for the nuclei; and E_{II} is the repulsive nucleus-nucleus interaction. The total energy of the system is thus given after Eq. (2.15) by the expectation value of the Hamiltonian operator:

$$E = \langle\psi|\hat{H}|\psi\rangle \equiv \langle\hat{H}\rangle, \tag{2.17}$$

where $|\psi\rangle$, which we assume to be normalized, is the wave function characterizing the state of the system.

Once our very fundamental equation, Eq. (2.15) has taken an explicit form,

[6] Our treatment will include spin effects only implicitly, as we perform spin-unpolarized calculations in successive chapters. By implicit we mean that the orbital $\psi_i(\mathbf{r})$ is characterized by a subscript $i \equiv [(1,\uparrow),(1,\downarrow),(2,\uparrow),\ldots]$ that denotes also the spin.

[7] Although Martin uses atomic units, he defines the electrostatic factor as $\frac{4\pi}{\epsilon_0}$, whereas the standard electrostatic factor in SI units is $\frac{1}{4\pi\epsilon_0}$ (see, for instance, Jackson [45]). In order to avoid confusion, we will keep an explicit notation in SI units throughout the text.

one can start making approximations in order to simplify the problem of solving it. The first approximation is to consider the movement of electrons and nuclei separately: the more massive nuclei move rather slowly compared with the fast motion of the electrons. In this scenario, the wave function can be separated into parts containing the nuclear and electronic degrees of freedom independently:

$$|\psi\rangle = |\psi\rangle_{\text{elec}} \times |\psi\rangle_{\text{nuc}}. \tag{2.18}$$

This is known as the Born-Oppenheimer approximation [46]. We are mostly interested in $|\psi\rangle_{\text{elec}}$, which for simplicity we denote by $|\psi\rangle_{\text{elec}} \equiv |\psi\rangle$. In this context, we neglect the term in Eq. (2.15) containing $\frac{1}{M_I}$, that is $\hat{T}_{\text{nuc}}$, and the nuclear degrees of freedom become parameters, with the nuclear-nuclear electrostatic interaction E_{II} becoming a constant term, important only for the matter of total energy calculations that concern geometry optimization (see Chapter 4).

The calculation of $\langle \hat{V}_{\text{int}} \rangle$, as will become clear in the remainder of this section, is one of the main difficulties in DFT. While in a classical picture it would be obtained simply as a space integral involving the electronic charge density, in a quantum-mechanical frame one must pay attention to the quantum nature of electrons. To simplify the problem, $\langle \hat{V}_{\text{int}} \rangle$ can therefore be expressed as the sum of the classical electrostatic energy and *the rest*. If we denote the electronic charge density by

$$n\left(\mathbf{r}\right) = |\psi\left(\mathbf{r}\right)|^2, \tag{2.19}$$

then we can write the total energy as a function of explicit integrals of the density:

$$E = \langle \hat{T} \rangle + \int \mathrm{d}\mathbf{r}\, V_{\text{ext}}\left(\mathbf{r}\right) n\left(\mathbf{r}\right) + \underbrace{\frac{1}{8\pi\epsilon_0} \int\int \mathrm{d}\mathbf{r}\, \mathrm{d}\mathbf{r}'\, \frac{n\left(\mathbf{r}\right) n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + E_{\text{xc}}}_{\langle \hat{V}_{\text{int}} \rangle} + E_{II}, \tag{2.20}$$

where the *exchange-correlation* energy E_{xc} (*the rest*) is given as the difference between the energy of the quantum-mechanical electron-electron interaction and that corresponding to the classical electrostatic electron-electron interaction (Hartree energy E_{Hartree}), which in Eq. (2.20) is given by the double integral over the charge density. An initial attempt to solve Eq. (2.20) was made by Thomas [47] and Fermi [48]. In the Thomas-Fermi model, $\langle \hat{T} \rangle$ was assumed to be a functional of

the density $\langle \hat{T} \rangle = T\left[n\left(\mathbf{r}\right)\right]$, with the explicit form given by the kinetic energy of a non-interacting uniform electron gas:

$$T_{\mathrm{TF}} = \frac{3\hbar^2}{10m_e}\left(3\pi^2\right)^{\frac{2}{3}} \int \mathrm{d}\mathbf{r}\, \left[n\left(\mathbf{r}\right)\right]^{\frac{5}{3}}. \tag{2.21}$$

Initially, the exchange-correlation term E_{xc} was neglected, only to be included later by Dirac [49], who proposed a local density approximation (LDA), to which we shall come back further on. As it turned out, the Thomas-Fermi approximation for the kinetic energy led to much larger error than neglecting the exchange-correlation term.[8] Thus, in order to overcome the limitations introduced by the approximation imposed to the kinetic energy functional, Kohn and Sham [51], building upon the Hohenberg-Kohn theorems [52],[9] proposed an independent-particle approach where the density is given by a set of independent wave functions (or *orbitals*):

$$n\left(\mathbf{r}\right) = \sum_{i=1}^{N} |\psi_i\left(\mathbf{r}\right)|^2, \tag{2.22}$$

where i runs over all occupied states $(1, \ldots, N)$. In the Kohn-Sham (KS) formalism, it is assumed that all the many-body effects are contained within the exchange-correlation functional. Eq. (2.22) allows to conveniently express the kinetic energy in terms of the independent-particle orbitals $|\psi_i\rangle$:

$$T_{\mathrm{KS}} = -\frac{\hbar^2}{2m_e} \sum_{i=1}^{N} \langle \psi_i | \nabla_i^2 | \psi_i \rangle, \tag{2.23}$$

so that we can write

$$\begin{aligned} E_{\mathrm{KS}} = &-\frac{\hbar^2}{2m_e} \sum_{i=1}^{N} \langle \psi_i | \nabla_i^2 | \psi_i \rangle + \int \mathrm{d}\mathbf{r}\, V_{\mathrm{ext}}\left(\mathbf{r}\right) n\left(\mathbf{r}\right) + \frac{1}{8\pi\epsilon_0} \int\int \mathrm{d}\mathbf{r}\, \mathrm{d}\mathbf{r}'\, \frac{n\left(\mathbf{r}\right) n\left(\mathbf{r}'\right)}{|\mathbf{r} - \mathbf{r}'|} \\ &+ E_{\mathrm{xc}}\left[n\left(\mathbf{r}\right)\right] + E_{II}. \end{aligned} \tag{2.24}$$

[8] However, and in spite of its limitations, there is renewed interest in the development of kinetic functionals for use in multiscale simulations, in the context of orbital-free density functional theory (OFDFT), given the much reduced computational requirements compared to Kohn-Sham DFT with explicit orbitals [50].

[9] The Hohenberg-Kohn theorems establish that (i) the external potential is uniquely determined, except for a constant, by the density and (ii) that there exists a universal energy functional of the density $E\left[n\left(\mathbf{r}\right)\right]$ such that the density which minimizes the functional is the density of the ground state.

The *only* task left at this point is the non-trivial one of finding an exchange-correlation functional; this will, in practice, limit the accuracy of the method. With the definitions given above, E_{xc} can be directly expressed as the difference between the exact many-body interacting system and the independent-particle approximation:

$$E_{\mathrm{xc}} = \langle \hat{T} \rangle - T_{\mathrm{KS}} + \langle \hat{V}_{\mathrm{int}} \rangle - E_{\mathrm{Hartree}}. \tag{2.25}$$

Applying a variational principle to the energy E_{KS}, for instance that it be invariant with respect to changes in the wave function [43],[10] one arrives at the Kohn-Sham eigenvalue equation:

$$\hat{H}_{\mathrm{KS}} \psi_i (\mathbf{r}) = \epsilon_i \psi_i (\mathbf{r}) , \tag{2.26}$$

where the Kohn-Sham Hamiltonian is given by the kinetic operator and the KS effective potential:

$$\hat{H}_{\mathrm{KS}} = -\frac{\hbar^2}{2m_e} \nabla^2 + \hat{V}_{\mathrm{KS}} (\mathbf{r}) . \tag{2.27}$$

The *effective* KS potential $\hat{V}_{\mathrm{KS}}$ is given by

$$\hat{V}_{\mathrm{KS}} = \hat{V}_{\mathrm{ext}} (\mathbf{r}) + \frac{\delta E_{\mathrm{Hartree}}}{\delta n (\mathbf{r})} + \frac{\delta E_{\mathrm{xc}}}{\delta n (\mathbf{r})}. \tag{2.28}$$

From Eq. (2.25) it is manifestly clear that when solving Eq. (2.26), the success of the Kohn-Sham approach must greatly rely on finding an accurate functional for E_{xc}.

Local density and generalized gradient approximations

The most straight-forward approximation for an exchange-correlation energy, and the first one to come to mind, is to use that corresponding to the homogeneous electron gas (HEG). The local density approximation (LDA) is based on that, and therefore only the local value of the charge density enters the expressions for both

[10] The variational equation, $\delta E_{\mathrm{KS}} / \delta \psi_i^* = 0$ can be solved using the method of Lagrange multipliers subject to the condition of orthonormality for the wave functions $\langle \psi_i | \psi_j \rangle = \delta_{ij}$: $\delta E_{\mathrm{KS}} / \delta \psi_i^* - \delta / \delta \psi_i^* \sum_j \epsilon_j \left(\langle \psi_i | \psi_j \rangle - \delta_{ij} \right) = 0$, which leads to Eq. (2.26).

exchange and correlation energies. For a homogeneous system of density n, we define the Wigner-Seitz radius r_s, which gives the radius of a sphere containing the charge corresponding to one electron:

$$\frac{4\pi}{3}{r_s}^3 n = 1 \qquad \rightarrow \qquad r_s = \left(\frac{3}{4\pi n}\right)^{\frac{1}{3}}. \tag{2.29}$$

The exchange energy of the HEG per electron is then [43]:

$$\frac{E_{\mathrm{x}}^{\mathrm{HEG}}}{N_e} = \epsilon_{\mathrm{x}}^{\mathrm{HEG}} = -\frac{1}{4\pi\epsilon_0}\frac{3e^2}{4\pi}\left(\frac{9\pi}{4}\right)^{\frac{1}{3}}\frac{1}{r_s}. \tag{2.30}$$

The LDA for the exchange part is realized following Eq. (2.30) by substituting the constant charge density n of the HEG by the local charge density $n(\mathbf{r})$ of the KS independent-particle system:

$$E_{\mathrm{x}}^{\mathrm{LDA}}[n(\mathbf{r})] = \int \mathrm{d}\mathbf{r}\, n(\mathbf{r})\, \epsilon_{\mathrm{x}}^{\mathrm{HEG}}(n(\mathbf{r})). \tag{2.31}$$

While the integrand in Eq. (2.31) for the exchange energy has a convenient analytical dependence on $n(\mathbf{r})$, the correlation part $\epsilon_{\mathrm{c}}^{\mathrm{HEG}}$ must be *parametrized*, relying on auxiliary numerical data. One of the most widely used parametrizations, and the one employed for the results obtained in this work within the LDA approximation, is that due to Perdew and Zunger (PZ) [53]. The PZ parametrization is a fit to quantum Monte Carlo (QMC) data for the energy of the HEG, performed by Ceperley and Alder [54]. Within this scheme, the correlation energy per electron $\epsilon_{\mathrm{c}}^{\mathrm{PZ}}$ is given, in terms of $\tilde{r}_s$, by

$$\epsilon_{\mathrm{c}}^{\mathrm{PZ}}(\tilde{r}_s) = \begin{cases} A\ln\tilde{r}_s + B + C\tilde{r}_s\ln\tilde{r}_s + D\tilde{r}_s & \text{for} \quad \tilde{r}_s < 1, \\ \dfrac{\gamma}{1+\beta_1\sqrt{\tilde{r}_s}+\beta_2\tilde{r}_s} & \text{for} \quad \tilde{r}_s \geq 1, \end{cases} \tag{2.32}$$

where $\tilde{r}_s$ is the Wigner-Seitz radius given in units of the Bohr radius, and with $A = 0.846$ eV, $B = -1.31$ eV, $C = 0.054$ eV, $D = 0.316$ eV, $\gamma = 3.872$ eV, $\beta_1 = 28.65$ eV and $\beta_2 = 9.072$ eV [53].[11] The total exchange-correlation energy in

[11] These interpolation coefficients are for the unpolarized case. $\tilde{r}_s = \frac{r_s}{a_0}$, with $a_0 \approx 0.529177$ Å. The coefficients A, B, C, etc. were originally given by Perdew and Zunger [53] in atomic units, we apply the conversion 1 a.u. = 27.211 eV.

the LDA is then expressed as

$$E_{\mathrm{xc}}^{\mathrm{LDA}}\left[n\left(\mathbf{r}\right)\right]=\int\mathrm{d}\mathbf{r}\,n\left(\mathbf{r}\right)\left(\epsilon_{\mathrm{x}}^{\mathrm{HEG}}\left(n\left(\mathbf{r}\right)\right)+\epsilon_{\mathrm{c}}^{\mathrm{HEG}}\left(n\left(\mathbf{r}\right)\right)\right), \tag{2.33}$$

where $\epsilon_{\mathrm{x}}^{\mathrm{HEG}}$ has an exact analytical form, as given by Eq. (2.30), and $\epsilon_{\mathrm{c}}^{\mathrm{HEG}}$ has a non-analytical (but exact) form that can be parametrized into an analytical expression, e.g. the PZ parametrization given in Eq. (2.32).

While Eq. (2.33) gives the exact exchange-correlation energy for a system where the charge density remains constant, in real systems the density changes with position. It is then natural to expect that the value of $\epsilon_{\mathrm{xc}}\left(\mathbf{r}\right)$ depends not only on the density itself $n\left(\mathbf{r}\right)$ but also on the successive spatial derivatives of $n\left(\mathbf{r}\right)$, that is, its gradients. The simplest approximation within this gradient expansion picture is to assume that $\epsilon_{\mathrm{xc}}\left(\mathbf{r}\right)$ is a functional, in addition to $n\left(\mathbf{r}\right)$, of the *magnitude* of the first order gradient $\left|\nabla n\left(\mathbf{r}\right)\right|$:

$$E_{\mathrm{xc}}^{\mathrm{GGA}}\left[n\left(\mathbf{r}\right)\right]=\int\mathrm{d}\mathbf{r}\,n\left(\mathbf{r}\right)\epsilon_{\mathrm{xc}}^{\mathrm{GGA}}\left(n\left(\mathbf{r}\right),\left|\nabla n\left(\mathbf{r}\right)\right|\right). \tag{2.34}$$

The approximation implied by Eq. (2.34) is referred to as the generalized gradient approximation (GGA). A careless implementation of a GGA might lead to some inconsistencies. Therefore, to improve upon LDA, a GGA must take these issues into consideration [43]. Many GGAs exist that improve some of the predictions of the LDA for numerous systems. We will limit ourselves to the parametrization of Perdew, Burke and Ernzerhof (PBE) [55], since that is the GGA employed in the calculations of the present work. The form of the PBE GGA for the exchange-correlation energy is cast in terms of an *enhancement* factor F_{xc} over the HEG local exchange energy, that portrays the non-localicity of E_{xc} as a function of the charge gradient. In particular, the PBE approximation is:

$$E_{\mathrm{xc}}^{\mathrm{PBE}}\left[n\right]=\int\mathrm{d}\mathbf{r}\,n\epsilon_{\mathrm{x}}^{\mathrm{HEG}}\left(n\right)F_{\mathrm{xc}}^{\mathrm{PBE}}\left(r_{s},s\right), \tag{2.35}$$

where $s\propto\frac{\nabla n}{n}r_{s}$ is a dimensionless relative charge gradient, and $F_{\mathrm{xc}}^{\mathrm{PBE}}$ is given by:[12]

$$F_{\mathrm{xc}}^{\mathrm{PBE}}=F_{\mathrm{x}}^{\mathrm{PBE}}\left(s\right)+\frac{\epsilon_{\mathrm{c}}^{\mathrm{HEG}}+H\left(r_{s},t\right)}{\epsilon_{\mathrm{x}}^{\mathrm{HEG}}}, \tag{2.36}$$

[12]For a full spin-polarized treatment, refer to Ref. [55].

where $t \propto \frac{\nabla n}{n}\sqrt{r_s a_0}$ is another dimensionless gradient. Physically meaningful $F_{\mathrm{x}}^{\mathrm{PBE}}$ and $H(r_s, t)$ are realized by requiring them to satisfy several conditions. In particular, for the slowly varying limit ($s, t \to 0$), $F_{\mathrm{x}} \to 1$ and $H \to 0$, recovering the LDA limit. For the rapidly varying limit ($s, t \to \infty$), $H \to -\epsilon_{\mathrm{c}}^{\mathrm{HEG}}$ (strong confinement leads to reduced interaction between electrons) and $F_{\mathrm{x}} \to$ const., which satisfies the Lieb-Oxford bound.[13] Several other requirements establish the precise analytical form of $F_{\mathrm{x}}^{\mathrm{PBE}}$ and $H(r_s, t)$ [55].

Depending on the particular properties of the system at hand, either LDA or GGA can lead to better agreement with experiment. In particular, it is known that LDA tends to *overbind* while GGAs tend to *underbind*. In practice this leads to theoretical bond lengths and lattice parameters that are too short (LDA) or too long (GGA) (see for instance Zoroddu *et al.* [57] for a comparison of both methods for nitrides). Another not-so-well-known difference between methods is that LDA seems to predict piezoelectric coefficients in better agreement with experiment than GGA, at least for the III-Vs (see Beya-Wakata *et al.* [58]). The main issue, which affects both approximations, is related to the prediction of the band gap for semiconductors and insulators. The KS eigenvalues of Eq. (2.26), within the standard KS DFT theory, do *not* give the correct energy difference between the highest occupied and lowest unoccupied states [44]. In particular, both LDA and GGA underestimate band gaps, which leads to the incorrect theoretical prediction of conducting states for materials that are insulating in practice. This is a considerable problem in the case of narrow gap semiconductors, such as the In-containing III-Vs. In order to correct these issues, several approaches to generalize KS theory have been proposed. In the next section we deal with the hybrid functional approach, which we employ later in the text, that leads to improved agreement between theory and experiment for bond lengths and energy gaps.

2.2.2 Generalized Kohn-Sham approaches: the HSE hybrid functional

The fundamental band gap E_g of a semiconductor or insulator is given by energy differences when adding and removing electrons from an N-particle system. If N is the number of electrons in the ground state, then $E_g = E(N+1) + E(N-1) - 2E(N)$ [44, 59]. Naively one would expect that this energy finds a direct

[13] The Lieb-Oxford bound [56] establishes that E_{xc} must be bound from below, $E_{\mathrm{xc}} \geq -1.679e^2 \int \mathrm{d}\mathbf{r}\, n^{\frac{4}{3}}$. Since for the rapidly varying limit the correlation energy vanishes, this sets a direct bound on F_{x} within this limit.

correspondence with the KS eigenvalues of Eq. (2.26) as $E_g = \epsilon_{N+1} - \epsilon_N$, *at least if the exact exchange-correlation functional* E_{xc} *was known.* Surprisingly, Perdew and Levy [44], and Sham and Schlüter [59, 60], showed independently that there is no direct correspondence between both quantities due to a discontinuity of the functional derivative of E_{xc} with respect to the number of electrons. This sets a fundamental limitation to the ability of *exact* KS theory to correctly predict band gaps of semiconductors and insulators.

In order to overcome this limitation of the standard KS formalism, so-called generalized Kohn-Sham (GKS) approaches have been proposed [61]. The main idea underlying this generalized formalism is to introduce exact Hartree-Fock-like exchange into the exchange-correlation energy. An example of such a *hybrid* functional is the Heyd-Scuseria-Ernzerhof (HSE) screened-exchange functional [62, 63], that we will employ in Section 4.4. In the HSE scheme, the Coulomb operator for the exchange part of the energy is split into short-range (SR) and long-range (LR) components:

$$\frac{1}{|\mathbf{r}-\mathbf{r}'|} = \underbrace{\frac{\mathrm{erfc}\left(\omega|\mathbf{r}-\mathbf{r}'|\right)}{|\mathbf{r}-\mathbf{r}'|}}_{\mathrm{SR}} + \underbrace{\frac{\mathrm{erf}\left(\omega|\mathbf{r}-\mathbf{r}'|\right)}{|\mathbf{r}-\mathbf{r}'|}}_{\mathrm{LR}}, \tag{2.37}$$

where ω is a *tunable* screening parameter. Within the HSE approximation, only a portion of short-range exchange is Hartree-Fock (HF), the rest of the exchange, including all of the long-range exchange, as well as the correlation energy, is given by the PBE functional. Schematically, $E_{\mathrm{xc}}^{\mathrm{HSE}}$ can be written as

$$E_{\mathrm{xc}}^{\mathrm{HSE}} = aE_{\mathrm{x}}^{\mathrm{HF,SR}}(\omega) + (1-a)\,E_{\mathrm{x}}^{\mathrm{PBE,SR}}(\omega) + E_{\mathrm{x}}^{\mathrm{PBE,LR}}(\omega) + E_{\mathrm{c}}^{\mathrm{PBE}}, \tag{2.38}$$

where a is the *mixing parameter* that determines how much exact exchange is included in the calculation.[14] In order to obtain the HF short-range part, an

[14] The current standard implementation of the HSE functional, called HSE06, employs $a = 0.25$ and $\omega = 0.11{a_0}^{-1}$, where a_0 is the Bohr radius [64]. The screening for the PBE part is introduced through scaling the PBE *exchange hole* [43]. The latter is then integrated to obtain an enhancement factor, as given in Eq. (2.36), that includes the Coulomb screening [62].

explicit double integral over the orbitals has to be calculated:

$$E_{\mathrm{x}}^{\mathrm{HF,SR}}(\omega) = -\frac{1}{4\pi\epsilon_0}\sum_{j>i}\int\int \mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{r}\,'\psi_i^*(\mathbf{r})\,\psi_j^*(\mathbf{r}')\,\frac{\mathrm{erfc}(\omega|\mathbf{r}-\mathbf{r}'|)}{|\mathbf{r}-\mathbf{r}'|}\psi_i(\mathbf{r}')\,\psi_j(\mathbf{r})\,. \tag{2.39}$$

Although the screening effects are included in the HSE functional in order to minimize the computational requirements of the calculation, it can be anticipated from Eq. (2.39) that the inclusion of the HF exchange will add to the computational load compared to PBE. As for the results, HSE leads to the desired improved agreement with experiment for the band gap *and* structural and elastic parameters for a wide range of semiconductors [3,64–67]. In Section 4.4 we present an accurate calculation of structural and elastic properties of group-III nitrides using the HSE functional.

So far we have studied the different approximations made to simplify the calculation of the electron-electron interaction in DFT. We turn now to approximations to the nucleus-electron interaction or, more precisely, *core*-electron interaction, which are in place in the context of frozen-core and pseudopotential theory.

2.2.3 Pseudopotentials and the projector augmented-wave method

In this section we will briefly introduce pseudopotentials and the related projector augmented-wave method (PAW) implemented in VASP and used in the calculations of Chapter 4. For a more exhaustive review, the reader is referred to the book by Martin [43] and the paper by Blöchl [68].

A pseudopotential is an effective potential designed to substitute the all-electron potential of Eq. (2.15), and is constructed to reproduce the effects of the all-electron system on the valence states beyond a certain cutoff distance from the ionic core. In a pseudopotential calculation, the "core" consists of the nucleus and the inner electrons, which are highly localized around the nucleus. Outside the core region (that is, beyond the cutoff), the potential matches the Coulomb interaction between the core (whose ionic charge Z_{core} equals that of the nucleus minus the inner electrons) and each of the valence electrons. Inside the core region, the Coulomb potential is replaced by a smooth function which is more easily representable, for instance, by plane waves. Pseudopotentials are constructed so that the wave functions of the valence electrons outside the core region match those of an all-electron

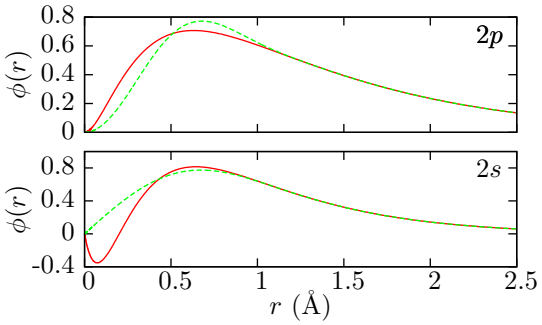


Figure 2.3: Valence orbitals of C ($2s$ and $2p$) calculated in an LDA-DFT all-electron scheme (red solid curves) and the corresponding pseudo wave functions (green dashed lines). The (norm-conserving) pseudopotentials were generated with the Quantum Espresso pseudopotential generator tool [69, 70], with the constraint that atomic all-electron and pseudo wave functions match at and beyond the cutoff radius $r_c = 2.5a_0$ (~ 1.322 Å).

calculation, in a calculation for the isolated atom. To illustrate this, in Fig. 2.3 we show the radial part $\phi(r)$ of the $2s$ and $2p$ orbitals of atomic C together with the corresponding pseudo wave functions obtained with pseudopotentials we have designed to make them match at $r_c = 2.5a_0$ (~ 1.322 Å).

The main argument underlying the use and success of pseudopotentials is the fact that most chemical properties of atoms can be accurately modelled taking account of the interacting valence states alone. The core states, much lower in energy, have little interaction with valence states of the same or surrounding atoms. A notable exception to this are elements with "shallow core" or "semicore" states, which are high enough in energy so that they influence the chemical properties of those elements. These states, for instance $3d$ and $4d$ orbitals in Ga and In, respectively, are close in energy to the valence orbitals, and must be included in accurate calculations [71]. As an example, we show a band structure calculation for ZB GaN in Fig. 2.4, using both $3d$ states explicitly as valence states and without them. The semicore states can be identified as the curves with flat dispersion at the bottom of the figure. It can be observed that these states interact with the lowest valence band that originates from the original atomic valence orbitals. The effect of this interaction is a shift of all the higher bands and a poor description of the lowest valence band if the Ga $3d$ orbitals are not explicitly included in the calculation.

The most common types of *ab initio*[15] pseudopotentials are "norm-conserving" [43]

[15]There exist also "empirical" pseudopotentials, which are fitted to experiment.

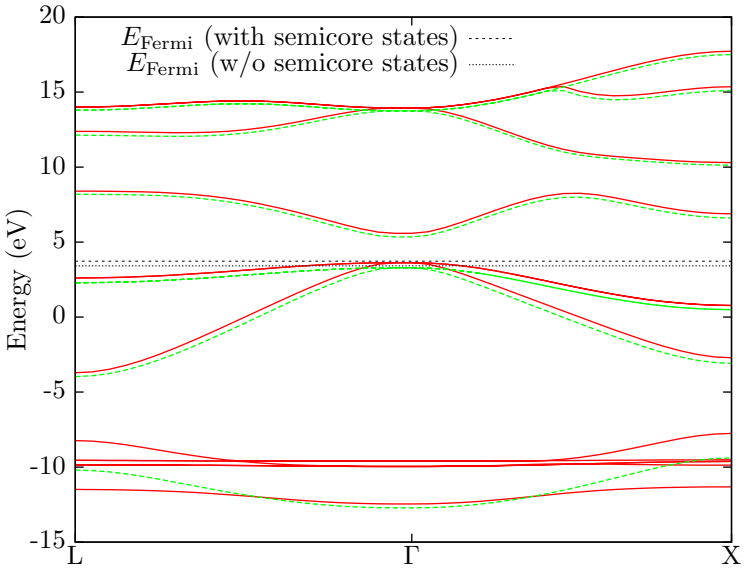


Figure 2.4: Band structure along high symmetry directions for ZB GaN for two different cases: including the semicore $3d$ states as valence states (red solid lines) and omitting them in the calculation (green dashed lines). Results obtained with VASP using the PAW-LDA library [72].

and "ultrasoft" pseudopotentials [73]. The advantage of norm-conserving pseudopotentials is that the orthonormality condition for the pseudo wave functions is preserved, leading to an eigenvalue equation analogous to Eq. (2.26):

$$\hat{H}_{\text{KS}}^{\text{PS}}\psi_i^{\text{PS}}(\mathbf{r}) = \epsilon_i\psi_i^{\text{PS}}(\mathbf{r}), \tag{2.40}$$

where PS indicates that the system's Hamiltonian and the corresponding wave functions have been obtained by the introduction of a pseudopotential. There are a series of conditions that a norm-conserving pseudopotential must comply with [43]. These conditions improve the transferability of norm-conserving pseudopotentials from one problem to another. On the other hand, ultrasoft pseudopotentials introduced by Vanderbilt [73] offer better computational efficiency at the cost of complicating the formulas. The "hardness" of a pseudopotential refers to how smooth or *soft* the potential looks in real space, in particular how rapidly the corresponding pseudo wave functions vary in the core region. In the context of plane wave calculations, functions that are smooth in real space can be represented in reciprocal space by a lower number of plane waves, compared to more rapidly varying functions.

In more recent times, the projector augmented-wave method (PAW) [68, 74]

has gained popularity due to its providing further computational advantages compared to the norm-conserving and ultrasoft pseudopotential methods. The general aspects of the PAW method will be introduced next.

The projector augmented-wave method

Here we give a general description of the projector augmented-wave (PAW) method based on Blöchl's original paper [68]. The starting point for the PAW method is to consider a linear transformation $\mathcal{T}$ that relates the valence all-electron wave function $|\psi\rangle$ and the smooth pseudo wave function $|\psi^{\mathrm{PS}}\rangle$:

$$|\psi\rangle = \mathcal{T}|\psi^{\mathrm{PS}}\rangle. \tag{2.41}$$

$\mathcal{T}$ is chosen to be unity outside the *augmentation region* Ω_R, which in the context of pseudopotential theory corresponds to the core region. That is, $|\psi\rangle$ and $|\psi^{\mathrm{PS}}\rangle$ are identical outside Ω_R, with $\mathcal{T} = \mathbb{1} + \hat{\mathcal{T}}_R$ defining a transformation operator $\hat{\mathcal{T}}_R$ that acts only within Ω_R. The operator $\hat{\mathcal{T}}_R$ is determined by two sets of *partial waves* $|\phi_k\rangle$ (all-electron partial waves) and $|\phi_k^{\mathrm{PS}}\rangle$ (pseudo partial waves) that give $|\psi\rangle$ and $|\psi^{\mathrm{PS}}\rangle$ within Ω_R:

$$\begin{aligned} |\psi\rangle &= \sum_k c_k|\phi_k\rangle \quad \text{within} \quad \Omega_R, \\ |\psi^{\mathrm{PS}}\rangle &= \sum_k c_k|\phi_k^{\mathrm{PS}}\rangle \quad \text{within} \quad \Omega_R, \end{aligned} \tag{2.42}$$

where the c_k are coefficients determined by the *projector functions* $\langle p_k|$:

$$c_k = \langle p_k|\phi_k^{\mathrm{PS}}\rangle, \tag{2.43}$$

with $\langle p_k|$ satisfying the condition

$$\sum_k |\phi_k^{\mathrm{PS}}\rangle\langle p_k| = \mathbb{1} \quad \text{within} \quad \Omega_R. \tag{2.44}$$

Finally, the transformation $\mathcal{T}$ can be written as

$$\mathcal{T} = \mathbb{1} + \sum_k \left(|\phi_k\rangle - |\phi_k^{\mathrm{PS}}\rangle\right)\langle p_k|, \tag{2.45}$$

and the all-electron and pseudo wave functions can be related as

$$|\psi\rangle = |\psi^{\mathrm{PS}}\rangle + \sum_k \left(|\phi_k\rangle - |\phi_k^{\mathrm{PS}}\rangle\right) \langle p_k|\psi^{\mathrm{PS}}\rangle. \tag{2.46}$$

That is, within the PAW method, the all-electron wave function $|\psi\rangle$ can be retrieved from the pseudo wave function as given by Eq. (2.46). When using the pseudo wave functions as variational parameters in an actual calculation, the all-electron operators need to be transformed into pseudo operators. Guidelines for this as well as explicit expressions for the charge density and total energy of the system have been given by Blöchl [68].

Technical details regarding the practical implementation of the method and how to generate PAW data sets (partial waves, projector functions, etc), in particular in the context of plane waves and the VASP package, can be found in the paper by Kresse and Joubert [74].

2.2.4 Plane-wave formalism

The convenience of formulating the electronic structure problem in a plane wave frame stems from the periodic nature of crystals. For any function $f(\mathbf{r})$ with the same periodicity as the lattice,[16] its Fourier transform vanishes at all points $\mathbf{k}$ in reciprocal space except for those that are given by the reciprocal lattice vectors $\mathbf{G}$. The corresponding integral needs to be performed only over the unit cell volume:

$$\tilde{f}(\mathbf{G}) = \frac{1}{V_{\mathrm{cell}}} \int\limits_{V_{\mathrm{cell}}} \mathrm{d}\mathbf{r}\, f(\mathbf{r})\, \mathrm{e}^{-\mathrm{i}\mathbf{G}\cdot\mathbf{r}}, \tag{2.47}$$

where the tilde indicates the Fourier transform.

In a system where the Hamiltonian is given by the kinetic energy operator and an effective potential $V_{\mathrm{eff}}(\mathbf{r})$, the independent-particle Schrödinger equation can be conveniently written as [43]:

$$\sum_{\mathbf{G}'} H_{\mathbf{G},\mathbf{G}'}(\mathbf{k})\, c_{i,\mathbf{G}'}(\mathbf{k}) = \epsilon_i(\mathbf{k})\, c_{i,\mathbf{G}}(\mathbf{k}), \tag{2.48}$$

where the matrix elements of $\hat{H}_{\mathrm{eff}}$ in the momentum basis $|\mathbf{q}\rangle = |\mathbf{k}+\mathbf{G}\rangle$, $|\mathbf{q}'\rangle =$

[16] This means $f(\mathbf{r}) = f(\mathbf{r}+\mathbf{R})$, with $\mathbf{R}$ any lattice vector.

$|\mathbf{k}+\mathbf{G}'\rangle$ are given by

$$H_{\mathbf{G},\mathbf{G}'}(\mathbf{k}) = \frac{\hbar^2}{2m}|\mathbf{k}+\mathbf{G}|^2\delta_{\mathbf{G},\mathbf{G}'} + \tilde{V}_{\text{eff}}(\mathbf{G}-\mathbf{G}'), \tag{2.49}$$

and the $c_{i,\mathbf{G}}(\mathbf{k})$ are the Fourier expansion coefficients of the single-particle states $\psi_{i,\mathbf{k}}(\mathbf{r})$. That is, there is one eigenvalue $\epsilon_i(\mathbf{k})$ per $\mathbf{k}$ point and per band i. To obtain properties such as total energy accurately, a summation in k-space (restricted to the first Brillouin zone) must be performed. Since for macroscopic crystals the allowed values for $\mathbf{k}$ form a quasi-continuum [39,43], one can replace the summation by an integral. For a parameter $\xi_i(\mathbf{k})$, the corresponding integral giving its average is

$$\bar{\xi}_i = \frac{V_{\text{cell}}}{(2\pi)^3}\int_{\text{BZ}} \mathrm{d}\mathbf{k}\,\xi_i(\mathbf{k}). \tag{2.50}$$

Additionally, the summation over bands for parameter ξ involves occupancies $f_{i,\mathbf{k}}$, which in the zero-temperature limit are 1 (occupied band) or 0 (unoccupied band):[17]

$$\bar{\xi} = \frac{V_{\text{cell}}}{(2\pi)^3}\sum_i\int_{\text{BZ}} \mathrm{d}\mathbf{k}\,f_{i,\mathbf{k}}\xi_i(\mathbf{k}). \tag{2.51}$$

Because the Schrödinger equation [Eq. (2.48)] must be solved at each $\mathbf{k}$ point, evaluating Eq. (2.50) very accurately might become prohibitive in realistic calculations, and k-space sampling must be performed for a limited amount of $\mathbf{k}$ points. In Chapter 4 we present an overview of how some of the calculated crystal properties, in particular elastic properties, behave depending on how accurately the integral in Eq. (2.50) [or Eq. (2.51)] is approximated.

In addition to sampling in k-space, the number of plane waves (corresponding to reciprocal lattice translations $\mathbf{G}$) used to approximate the different Fourier expansions will also limit the accuracy of calculations. This effect is also reviewed in Chapter 4. Expressions for the total energy and stresses in the plane-wave formalism are given there too.

[17] In the case of non spin-polarized calculations, all the bands are degenerate and the occupancies are 0 or 2. Also, for the finite-temperature case, the occupancies can be *partial*, and vary continuously from 0 to 1 (or 0 to 2, for spin-unpolarized); see Section 2.2.5.

2.2.5 The VASP package

VASP, or the Vienna *Ab initio* Simulation Package [72, 75], is a powerful computational tool for total energy calculations using a plane-wave basis set. In VASP, the Schrödinger equation is solved self-consistently by iteratively optimizing the charge density that determines the KS Hamiltonian that then, in turn, determines the single-particle eigenstates. These eigenstates are used to calculate the new charge density that will serve as an input for the following iteration. The total energy difference between one iteration and the following will progressively decrease as convergence (and hence self-consistency) is achieved. Once the energy difference between two consecutive iterations falls within the desired precision range, the iterative process comes to an end: the single-particle eigenstates determine a charge density which gives rise to those same eigenstates (within the convergence requirements).

For the calculation of total energy, charge density and, in general, any parameter whose calculation involves an integration over $\mathbf{k}$ points [Eqs. 2.50 and 2.51], the contribution of each band i needs to be weighed with the occupancy of that band at each $\mathbf{k}$ point $f_{i,\mathbf{k}}$. For energies close to the Fermi level, the occupancies might vary for the same band at different $\mathbf{k}$ points. This can be particularly problematic in the case of metals and narrow-gap semiconductors, for which a large number of $\mathbf{k}$ points might be needed to reconstruct the Fermi surface [76]. To accelerate convergence with the number of $\mathbf{k}$ points, one of the possible strategies is to use a smearing method, introduced originally by Fu and Ho [76]. In this scheme, the energies of each of the eigenstates are smeared using, for instance, a Gaussian distribution, and the Fermi energy is calculated from the density of states (DOS) given by the smeared energies. Then, the partial occupancy of each band equals the portion of its smeared value lying below the calculated Fermi energy [76]. In this context, the total energy E is replaced by a generalized free energy F, whose extra term resembles the entropy term of the finite-temperature case. Substituting the integral over the Brillouin zone by a summation over $\mathbf{k}$ points we have [72, 75]:

$$F = E - \sum_{i,\mathbf{k}} w_{\mathbf{k}} \sigma S\left(f_{i,\mathbf{k}}\right), \tag{2.52}$$

where $w_{\mathbf{k}}$ is a weight arising from having replaced the integral by a summation (which can also take into account which $\mathbf{k}$ points are equivalent depending on the

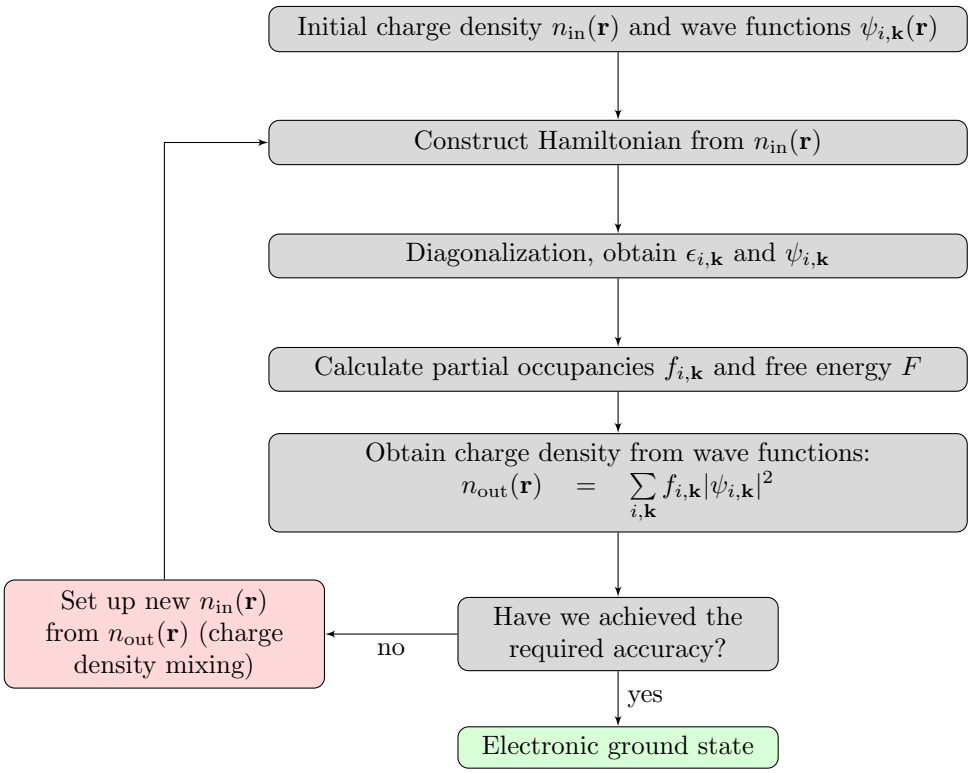


Figure 2.5: Flux diagram showing the basic procedure that VASP performs towards obtaining self-consistently the electronic ground state. See the VASP Manual online for further information [72].

crystal symmetries), σ is the smearing parameter and S is an entropy-like term that depends on the partial occupancies [75].

Taking all this into account, we show in Fig. 2.5 a simplified flux diagram highlighting the essentials of VASP's self-consistency loop for convergence to the electronic ground state [72]. The global ground state also involves geometry optimization, and therefore the calculation of forces acting on ions and stresses acting on the cell. These are calculated by VASP as the derivatives of the free energy with respect to the ionic degrees of freedom, once the electronic (instantaneous) ground state for the particular ionic arrangement has been obtained.

2.3 Modern theory of polarization: the Berry-phase technique

A non-vanishing dipole moment for a given charge distribution arises from the lack of radial symmetry of the distribution [45]. In the context of the description of periodic crystals, it is more convenient to work in terms of the dipole moment per unit volume, that is, the density of dipole moment, or polarization. Crystals whose symmetry allows an inversion centre cannot present a net dipole moment [41]. For crystals without an inversion centre, except point group 432,[18] certain deformations of the crystal lattice give origin to net dipole moments, known as the *piezoelectric effect*. In addition to this, the subset of those crystals that present an anisotropic direction in the lattice, called *polar*, are compatible with the existence of net dipoles even in the unstrained state, which is referred to as *spontaneous polarization*. The WZ crystal structure belongs to the latter class and therefore WZ nitrides present both piezoelectric and spontaneous polarization [41].

Calculating the polarization of a periodic crystal might seem at first a trivial problem, with a possible intuitive definition being given by the charge density of the unit cell. However, there is no way of unambiguously defining the polarization vector using such a method, with an array of possible values arising from different choices of origin [77]. Consider the case of Si, for instance. In Fig. 2.6 we show the charge density of the Si unit cell. The arrows show the direction and (relative) magnitude of the electronic polarization calculated from the charge density using the different points shown on the plane as the origin from where the unit cell vectors are spanned. The polarization calculated with such a method is clearly dependent on the reference frame used, which means that the charge density does *not* lead to a useful definition of polarization.

A rigorous frame for the computation of polarization in periodic solids was not available until as recently as the 1990s. The main developments were presented in the seminal papers by Vanderbilt and King-Smith [78, 79], building up on an idea originally suggested by Resta [80], where the foundations of the *Berry-phase* theory of polarization, or *modern theory of polarization* [81], where laid. This theory allows a calculation of the dipole moment of the unit cell of a periodic insulating

[18] 432 is the Hermann-Mauguin symbol; using the Schoenflies system, the equivalent point group is O (orthorhombic symmetry). While point group 432 does not present *linear* piezoelectricity, Grimmer has shown that it is compatible with second-order piezoelectricity [42].

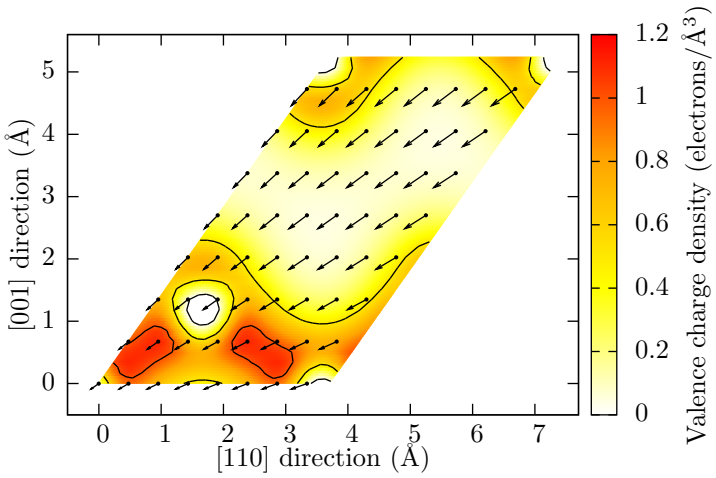


Figure 2.6: Charge density along plane $[1\bar{1}0]$ of a Si unit cell, calculated with VASP [72]. The arrows indicate the direction and relative magnitude (in arbitrary units) of the electronic polarization of the unit cell calculated as an integral of the charge density over the unit cell, when the lattice vectors are spanned from each of the points indicated by black dots.

system, which is well defined modulo $e\mathbf{R}$ (where e is the elementary charge and $\mathbf{R}$ is a lattice vector). The latter ambiguity can be removed in different ways, such that a meaningful value for the polarization can be obtained [78, 79, 82]. In this section we will introduce the fundamentals of this approach. We will pay special attention to the practicalities of applying the corrections to the electronic (Berry-phase) and ionic parts of the polarization. This is a fundamental part of Berry-phase calculations especially when the lattice vectors change, with the corrections being typically larger than the change in polarization itself. Our impression is that the subtleties concerning these corrections have been only vaguely mentioned in the literature [79, 82], and that it might be helpful to give here a practical guide to actual Berry-phase calculations based on our experience.

2.3.1 Fundamentals of the Berry-phase formalism

In the Berry-phase theory of polarization, the concept of absolute polarization is abandoned, and only *differences* in polarization between two states of an insulating material are meaningful [77]. These two states could be, for instance, a centrosymmetric (CS) structure and its symmetry-broken counterpart. In the specific case of zinc-blende III-Vs, the former state corresponds to the unstrained structure while the latter state corresponds to strain-induced shear deformation.

The corresponding change in polarization $\Delta\mathbf{P} = \mathbf{P}_{\text{non-CS}} - \mathbf{P}_{\text{CS}}$, and not the end points $\mathbf{P}_{\text{non-CS}}$ or $\mathbf{P}_{\text{CS}}$ by themselves, is the quantity under consideration. In the context of the modern theory, the polarization of the CS system $\mathbf{P}_{\text{CS}}$ can be assumed to vanish only based on symmetry arguments, but not as a result of an actual calculation [77]. If the adiabatic evolution of the system is parametrized through λ, which varies between 0 and 1, then $\Delta\mathbf{P}$ can be expressed as [77]

$$\Delta\mathbf{P} = \int_0^1 \mathrm{d}\lambda \, \frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\lambda}. \tag{2.53}$$

As explained earlier, if $\lambda = 0$ corresponds to the CS structure, then it might be justified to take the liberty to *define* the absolute value of the polarization through the integral in Eq. (2.53). However, this would mean in practice that one must devise a path in parameter space that brings the system from the CS structure to the structure of interest *adiabatically* [79, 82]. We will come back to this in Section 2.3.2 when dealing with the example of the spontaneous polarization of GaN. We will focus here on the electronic part of the polarization $\mathbf{P}_{\text{ele}}$, since the ionic part $\mathbf{P}_{\text{ion}}$ (given by the dipole moment of the nuclei) can be dealt with in a classical frame. How to compute the ionic part will be explained in Section 2.3.2.

In an experimental context, $\Delta\mathbf{P}$ would be determined by measuring the total charge flow associated with a transient electrical current, as schematically shown in Fig. 2.7, inspired by Resta's example [77]: a piezoelectric sample that has been shorted gives rise to a finite transient current flow after it is strained. If the current is given by $\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}t} = \mathbf{j}(t)$ and the current flows during a time interval Δt, then the difference in polarization in the sample between the unstrained and strained situation is $\Delta\mathbf{P} = \int_0^{\Delta t} \mathrm{d}t\, \mathbf{j}(t)$ [77].[19] $\mathbf{P}_{\text{ele}}(\lambda)$ is therefore obtained as an integrated current [Eq. (2.53)] which, in the context of quantum mechanics, bears a connection to the phase of the wavefunction. Following the work by King-Smith and Vanderbilt [78,79], we express $\mathbf{P}_{\text{ele}}(\lambda)$ as an explicit sum over occupied bands:

$$\mathbf{P}_{\text{ele}}(\lambda) = -\frac{\mathrm{i}e}{(2\pi)^3} \sum_{i=1}^{N} \int_{\text{BZ}} \mathrm{d}\mathbf{k} \, \langle u_{i\mathbf{k}}^{(\lambda)} | \nabla_{\mathbf{k}} | u_{i\mathbf{k}}^{(\lambda)} \rangle, \tag{2.54}$$

where $u_{i\mathbf{k}}^{(\lambda)}$ are the cell-periodic functions given by the Bloch theorem. The quantity

[19] As pointed out by Resta and Vanderbilt [77], in the adiabatic limit $\Delta t \to \infty$ and $\mathbf{j}(t) \to 0$, while the integral remains finite.

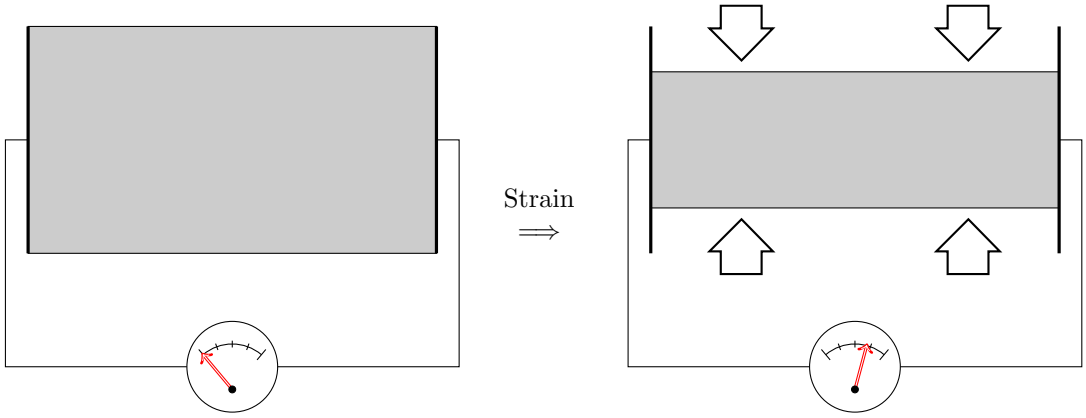


Figure 2.7: Example of experimental measurement of piezoelectric polarization: a sample of piezoelectric material is placed between two electrodes. When the strain state of the material is changed, a pulse of electric current can be measured. The current will stop flowing after the charges accumulated at the electrodes counteract the induced polarization. Based on Fig. 3 of Resta and Vanderbilt [77].

$\mathrm{i}\langle u_{i\mathbf{k}}^{(\lambda)}|\nabla_{\mathbf{k}}|u_{i\mathbf{k}}^{(\lambda)}\rangle$ is known as a Berry connection and its integral over the Brillouin zone is the Berry phase [77, 83]. For an extension of Eq. (2.54) to discrete space and an alternative formulation in terms of Wannier functions the reader is referred to Refs. [77, 79].

It is vital to note that a closed-loop integral of the form of Eq. (2.53) need not vanish [79]:

$$\oint \mathrm{d}\lambda \, \frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\lambda} = \frac{e\mathbf{R}}{V_{\mathrm{cell}}}, \tag{2.55}$$

where $\mathbf{R}$ is a lattice vector. This is related to the fact that two quantities whose phases are equal modulo 2π portray the exact same physical information. Consequently, within the Berry-phase formalism the polarization is only well defined modulo $e\mathbf{R}/V_{\mathrm{cell}}$. This is sometimes referred to as the "quantum of polarization" [77]. We will explain its practical implications and how to deal with them in the following section.

2.3.2 Practical considerations: removing the uncertainty in P

As we have previously mentioned, $\mathbf{P}_{\mathrm{ele}}$ is well defined only modulo $e\mathbf{R}/V_{\mathrm{cell}}$. Therefore, when calculating differences in polarization using Eq. (2.53), one can in prin-

ciple only calculate

$$\Delta\tilde{\mathbf{P}}_{\text{ele}} = \underbrace{\mathbf{P}^1_{\text{ele}} - \mathbf{P}^0_{\text{ele}}}_{\Delta\mathbf{P}_{\text{ele}}} + e\left(\frac{\mathbf{R}_1}{V^1_{\text{cell}}} - \frac{\mathbf{R}_0}{V^0_{\text{cell}}}\right), \tag{2.56}$$

where $\mathbf{R}_1$ and V^1_{cell}, and $\mathbf{R}_0$ and V^0_{cell} are a lattice vector and the unit cell volume in states 1 and 0, respectively. To access the quantity of interest, namely $\Delta\mathbf{P}_{\text{ele}}$, one must subtract the last term in Eq. (2.56). Note that in principle not only do we have to correct the results if the primitive lattice vectors change between calculations (e.g. for applied strain): $\mathbf{R}_1$ and $\mathbf{R}_0$ are lattice vectors that could also correspond to two different lattice translations generated with the *same* set of primitive lattice vectors. The question now is, how do we know how many lattice translations to subtract from the result of our calculation? Although $\mathbf{P}_{\text{ele}}$ is much smaller than $e\mathbf{R}/V_{\text{cell}}$, the truth is that for materials with a strong piezoelectric response, $\Delta\mathbf{P}_{\text{ele}}$ and $e\left(\mathbf{R}_1/V^1_{\text{cell}} - \mathbf{R}_0 V^0_{\text{cell}}\right)$ are comparable.

To deal with this issue one must figure out the branch in which the polarization is being computed, that is, which specific lattice vectors $\mathbf{R}_1$ and $\mathbf{R}_0$ must be subtracted to achieve the final results. In order to do this, one makes use of the fact that $\mathbf{P}_{\text{ele}} << e\mathbf{R}/V_{\text{cell}}$ upon direct inspection of the results, provided that a fortunate choice of origin has been made. The meaning of "fortunate choice of origin" will be made clear at the end of this section.

We now turn our attention towards the ionic polarization. In the context of the frozen-core approximation, the *ionic* dipole moment of the unit cell can be computed classically as that of an ensemble of point charges (the cores) at specific distances from the origin (see e.g. Jackson [45]). The charge of each of the cores is equal to the number of valence electrons of the corresponding atom. For the ionic polarization, there is an uncertainty similar to the one existing for the electronic part: the value of the ionic dipole moment will depend on which set of atoms are taken as the crystal's asymmetric unit and how the origin is chosen (see Vanderbilt and King-Smith's discussion on "ionic basis" [79]). In the most general case, we

can write

$$\frac{e}{V^1_{\text{cell}}}\sum_i Z_i\boldsymbol{\tau}_i^1 - \frac{e}{V^0_{\text{cell}}}\sum_i Z_i\boldsymbol{\tau}_i^0 = \underbrace{\mathbf{P}^1_{\text{ion}} - \mathbf{P}^0_{\text{ion}}}_{\Delta\mathbf{P}_{\text{ion}}} + e\left(\frac{q_{\mathbf{a}_1}\mathbf{a}_1 + q_{\mathbf{b}_1}\mathbf{b}_1 + q_{\mathbf{c}_1}\mathbf{c}_1}{V^1_{\text{cell}}} - \frac{q_{\mathbf{a}_0}\mathbf{a}_0 + q_{\mathbf{b}_0}\mathbf{b}_0 + q_{\mathbf{c}_0}\mathbf{c}_0}{V^0_{\text{cell}}}\right), \tag{2.57}$$

where Z_i are the ionic (frozen core) charges and $\boldsymbol{\tau}_i$ are the position vectors of the ions within the unit cell. The different q are the coefficients accounting for the amount that must be subtracted, and $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ are the lattice vectors for states 0 and 1, as indicated by their respective subscripts.

For consistency between calculations, one must give all the core distances relative to the position of one of the cores, which has to remain the same for both 0 and 1. The simplest way to do this is by placing that core at the origin. Additionally, one must ensure that the summation is carried out over the *same set* of atoms between calculations, that is, no atom should be replaced by any of its periodic replicas. If these conditions are fulfilled, then $q_{\mathbf{a}_1} = q_{\mathbf{a}_0}$, $q_{\mathbf{b}_1} = q_{\mathbf{b}_0}$ and $q_{\mathbf{c}_1} = q_{\mathbf{c}_0}$. However, in order to determine any one of them we must rely on either 0 or 1 corresponding to the CS structure. As a convention, we have assumed 0 is the CS structure. Therefore, for $\lambda = 0$ the (corrected) ionic polarization must vanish:

$$\sum_i Z_i\boldsymbol{\tau}_i^0 - q_{\mathbf{a}_0}\mathbf{a}_0 - q_{\mathbf{b}_0}\mathbf{b}_0 - q_{\mathbf{c}_0}\mathbf{c}_0 = \mathbf{0} \quad \Longleftrightarrow \quad \lambda = 0 \rightarrow \text{CS}. \tag{2.58}$$

Equation (2.58) leads to three additional equations, each allowing to determine one of the q. For all the CS structures that we considered, we always found $q_{\mathbf{a}_0}$, $q_{\mathbf{b}_0}$ and $q_{\mathbf{c}_0}$ to be fractional numbers.

Since all this might seem a bit cumbersome, we will resort to a practical example: the spontaneous polarization of WZ GaN.

Spontaneous polarization of GaN

In the spirit of Eq. (2.53), we need a CS reference to give sense to the spontaneous polarization calculated for WZ GaN. Therefore we choose a 12-atom [111]-oriented ZB GaN unit cell as the CS structure. This allows the construction of the smallest

reference cell which is commensurate to the WZ one: both have hexagonal symmetry and, for ideal WZ, share the same primitive lattice vectors. If a_{ZB} is the lattice constant of ZB GaN, and a_{WZ} and c_{WZ} are the lattice constants of WZ GaN, then the three unit cells used in this example are the ones given in Fig. 2.8.

The calculation of the Berry phase contribution to the electronic part of the polarization has been implemented by Martijn Marsman in VASP [72].[20] VASP's output for this PAW-LDA calculation is given in terms of the dipole moment of the unit cell, and is the following:

```
[111]-oriented ZB:

  Lattice vectors:
a_ZB = 3.1524234518858663    0.0000000000000000    0.0000000000000000
b_ZB = 1.5762117259429331    2.7300787928189911    0.0000000000000000
c_ZB = 0.0000000000000000    0.0000000000000000   15.4436578206071378

  p[elc]=( -0.00985 -0.00569 15.44113 ) electrons Angst


Ideal WZ:

  Lattice vectors:
a_idWZ = 3.1524234518858663    0.0000000000000000    0.0000000000000000
b_idWZ = 1.5762117259429331    2.7300787928189911    0.0000000000000000
c_idWZ = 0.0000000000000000    0.0000000000000000   15.4436578206071378

  p[elc]=( 0.00000 0.00000 -15.30572 ) electrons Angst


Real WZ:

  Lattice vectors:
a_reWZ = 3.1539999999999999    0.0000000000000000    0.0000000000000000
```

[20] We have used this module to obtain *ab initio* values of spontaneous polarization, piezoelectric coefficients and Born effective charges of group-III nitrides, as well as to test our theory of local polarization. The results of these calculations will be presented in Chapter 5. For simplicity, we will overlook for this example the technical details of the calculations, which can be found in Chapters 4 and 5 for the structural optimization and polarization, respectively.

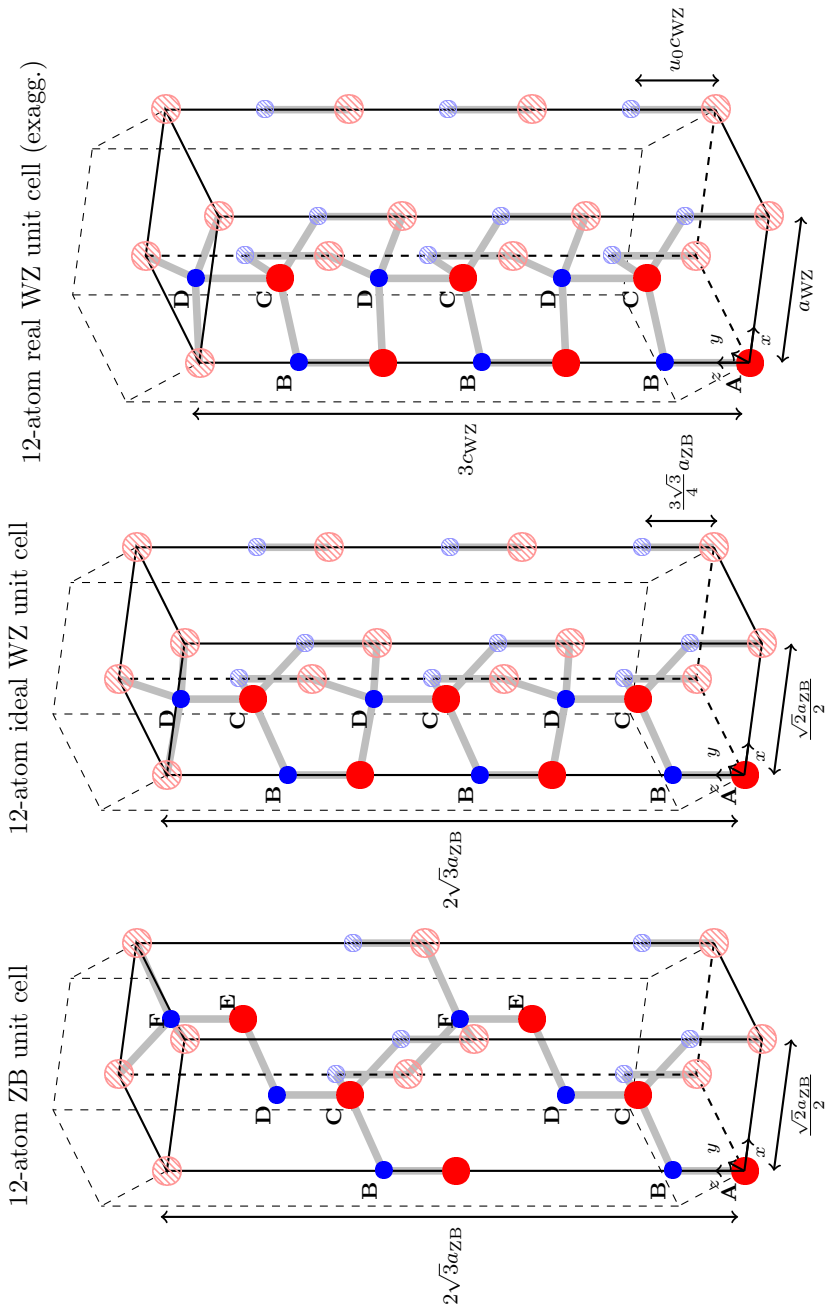


Figure 2.8: 12-atom commensurate cells used to calculate the spontaneous polarization of GaN. The ZB and ideal WZ cells share the exact same lattice vectors. The relative proportions of the real WZ cell have been exaggerated to dramatize the effect of the deviation of the c/a ratio and the internal parameter u_0 with respect to their ideal values ($\sqrt{8/3}$ and 3/8, respectively).

```
b_reWZ = 1.5770000000000000    2.7314441235361193    0.0000000000000000
c_reWZ = 0.0000000000000000    0.0000000000000000   15.4230000000000000

  p[elc]=( 0.00000 0.00000 -14.94106 ) electrons Angst
```

Attending to the fact that $\mathbf{P}_{\rm ele} << e\mathbf{R}/V_{\rm cell}$, it is straight-forward to see that the electronic part of the dipole moment is offset by minus the third lattice vector, for the ZB cell, and plus the third lattice vector, for the two WZ cells.[21] After calculating the cell volumes from the lattice vectors, the electronic contributions to the total polarization along the polar c-axis (z-axis in the convention of Fig. 2.8) are:[22]

$$\begin{aligned} P_{\rm ele}^{\rm ZB} &= 0.0003\,{\rm C/m^2}, \\ P_{\rm ele}^{\rm idWZ} &= -0.0168\,{\rm C/m^2}, \\ P_{\rm ele}^{\rm reWZ} &= -0.0581\,{\rm C/m^2}. \end{aligned} \tag{2.59}$$

Note that, without the correction, the difference between the real and ideal WZ values would have been -0.0573 C/m^2 rather than the calculated value of -0.0413 C/m^2.

For the ionic part, we place atom A (Ga) at the origin and sum over distances for the other atoms [Eq. (2.57)], with $Z_{\rm Ga} = 13$ and $Z_{\rm N} = 5$ given by our choice to treat the d semicore electrons of Ga explicitly. Following the LDA structural parameters of GaN (Chapter 5) and the positions of the atoms within the different unit cells (Chapter 4), the ionic contributions $\tilde{P}_{\rm ion}$ *without correction* are:

$$\begin{aligned} \tilde{P}_{\rm ion}^{\rm ZB} &= 91.4444\,{\rm C/m^2}, \\ \tilde{P}_{\rm ion}^{\rm idWZ} &= 91.4444\,{\rm C/m^2}, \\ \tilde{P}_{\rm ion}^{\rm reWZ} &= 90.6902\,{\rm C/m^2}. \end{aligned} \tag{2.60}$$

The corrected values $P_{\rm ion}^{\rm ZB}$ and $P_{\rm ion}^{\rm idWZ}$ should vanish by symmetry, and one can verify that the uncorrected values above correspond to exactly $195e/4V_{\rm cell}$ times the third lattice vector, which is the quantity that must be subtracted. Therefore we

[21] Note that the units used by VASP are *electrons* rather than elementary charges, and hence the sign of the results must be reversed.

[22] Note that the in-plane components of the polarization of the ZB cell do not vanish because of the finite k-sampling and the fact that the cell used to represent the crystal (hexagonal) does not have the crystal symmetry (cubic). Increasing the number of $\mathbf{k}$ points allows to systematically reduce these numerical artefacts.

apply the correction $-195e\,\mathbf{c}_{\mathrm{ZB}}/4V_{\mathrm{ZB}}$. The same quantity (with the corresponding volume and lattice vector) must also be subtracted from $\tilde{P}_{\mathrm{ion}}^{\mathrm{reWZ}}$, which has been computed using the same convention for the sum. The corrected values are:

$$\begin{aligned} P_{\mathrm{ion}}^{\mathrm{ZB}} &= 0, \\ P_{\mathrm{ion}}^{\mathrm{idWZ}} &= 0, \\ P_{\mathrm{ion}}^{\mathrm{reWZ}} &= 0.0279\,\mathrm{C/m^2}. \end{aligned} \tag{2.61}$$

The uncorrected difference in this case between ideal and real WZ would have been huge (and of opposite sign) compared to the actual difference. Finally, the spontaneous polarization of ideal and real WZ GaN can be obtained as the difference between those structures and the reference CS structure:

$$\begin{aligned} P_{\mathrm{sp}}^{\mathrm{idWZ}} &= P_{\mathrm{ele+ion}}^{\mathrm{idWZ}} - P_{\mathrm{ele+ion}}^{\mathrm{ZB}} = -0.0171\,\mathrm{C/m^2}, \\ P_{\mathrm{sp}}^{\mathrm{reWZ}} &= P_{\mathrm{ele+ion}}^{\mathrm{reWZ}} - P_{\mathrm{ele+ion}}^{\mathrm{ZB}} = -0.0305\,\mathrm{C/m^2}. \end{aligned} \tag{2.62}$$

In this case the corrections were relatively simple to apply. There is a pertinent question that yet remains unanswered: why is the CS reference cell needed if it is clear from the results in Eq. (2.59) (see also footnote 22 in the same page) that its polarization vanishes? The answer is: it vanishes only because we made a "fortunate choice of origin".

A fortunate choice of origin

The results of Eq. (2.59) where obtained from VASP for calculations where the atoms were positioned exactly as in Fig. 2.8, with the origin placed at atomic site A. Obviously, any translation from the origin of all 12 atoms contained within the unit cell by an arbitrary vector (even one that places them outside the unit cell) leads to the same physical system. Would such a transformation affect the results of a Berry phase calculation? The answer is yes, and not by a small amount. In Fig. 2.9 we show how for a standard 4-atom ideal WZ GaN unit cell the (electronic) spontaneous polarization varies linearly as the positions of all the atoms are shifted along the c-axis. The variation is cyclic and brings the polarization from the branch corresponding to $\mathbf{R} = (0, 0, +c)$ (where c is the lattice constant along the c-axis) to the branch corresponding to $\mathbf{R} = (0, 0, -c)$. We do not fully understand the reason for this behaviour, although it is clear that the period of a cycle equals c/N_{bands},

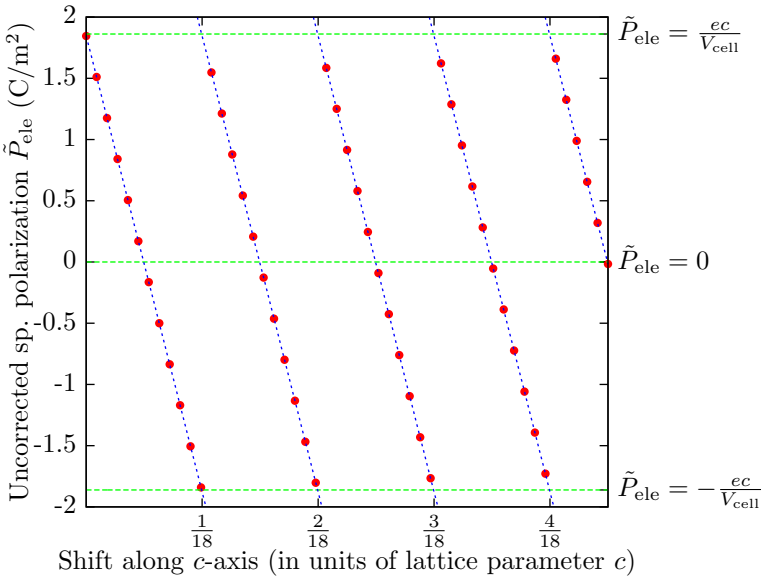


Figure 2.9: Uncorrected Berry-phase results for the electronic part of the polarization of a 4-atom ideal WZ GaN unit cell in which the atoms are shifted from the origin along the c-axis (z direction in Fig. 2.8). Blue dashed lines are linear fits to each of the data series, and green dashed lines indicate the quantum of polarization of the three branches spanned in this example.

where N_{bands} is the number of occupied bands (18 in the case of this GaN spin-unpolarized calculation). It is clear from the figure that an "unfortunate" choice of origin, for example corresponding to 1/4 or 3/4 of a cycle, leads to the result being placed halfway between the $\mathbf{R} = (0, 0, 0)$ branch and branches $\mathbf{R} = (0, 0, +c)$ or $\mathbf{R} = (0, 0, -c)$, respectively. Removing the uncertainty in $\mathbf{P}$ then becomes a serious issue, since there is no obvious way to assign the result to a particular branch. Therefore, in actual Berry-phase calculations one should focus on a first stage on finding a "nice" origin for the particular system at hand.

3 Built-in field control in nitride nanostructures

Abstract

In this chapter we investigate the degree to which the built-in electric field can be suppressed by employing polarization-matched barriers in III-N quantum well and dot structures grown along the *c*-axis. Our results show that it is possible to take advantage of the opposite contributions to the built-in potential arising from the different possible combinations of wurtzite GaN, InN and AlN when alloying the materials. We show that, for a fixed dot/well band gap, optimal alloy compositions can be found that minimize the built-in field across the structure. We discuss and study the impact of different material parameters on the results, including the influence of nonlinear effects in the piezoelectric polarization. Structures grown with unstrained barriers and on GaN epilayers are considered, including discussion of the effects of constraints such as strain limits and alloy miscibility. We apply the principle underlying built-in field control to study in more detail ultraviolet structures. For that spectral region, we assess the large impact on the results of the value of the AlInN band gap bowing parameter.

3.1 Introduction

Group III nitrides have attracted great attention over the last few years due to their potential for optoelectronic applications given the wide energy range over which their band gap E_g spans [26]. Alloying InN, GaN and AlN can in principle allow access to a wide spectral range, from near infrared (IR) to the ultraviolet (UV), spanning from an energy gap of 0.64 eV in InN to 6.16 eV in AlN [21]. InGaN-based devices have been demonstrated and proven to be a reliable solution for lighting applications in the blue part of the spectrum ($E_g \sim 2.7$ eV) [23, 24]. However, technical challenges still remain present, including for example the growth of high-

quality In-rich InGaN systems, which would allow the implementation of efficient solutions for green, yellow and amber light emitting diodes (LEDs) and lasers (E_g ∼2.0–2.5 eV) [24, 35, 84]. UV radiation sources ($E_g > 3.4$ eV) [26] are of interest for various applications such as optical storage, medical diagnostics and treatment, and sterilization processes [27,85]. In this spectral range, nitride-based structures might be either a good candidate for the replacement of traditional low-efficiency devices, or the only available solution in environments where it is not possible to implement any alternative approach [27, 85, 86]. Although there have been great improvements in the quality of nitride-based materials and structures during the past two decades, there is still need for further improvement towards the goal of high efficiency devices [24, 37].

Even though it is possible under certain conditions to grow III-N materials in the zinc-blende phase [22,87], good crystal quality is only achieved in the wurtzite structure. One of the particularities of this structure for group-III nitrides is the existence of spontaneous electric dipoles along the [0001] direction (c-axis) of the lattice, due to the highly ionic character of the bonds and the lack of an inversion plane perpendicular to the c-axis [88]. Along with this *spontaneous polarization*, additional dipoles are created when the material undergoes strain, generating the so-called *piezoelectric polarization* [89]. Piezoelectric polarization is typically large in nitride nanostructures due to the high lattice mismatch between the binaries (14% between InN and AlN). The total polarization differs for different nitrides, thus giving rise to the accumulation of interfacial electric charges in nitride-based heterostructures. This accumulation is particularly important in the cases where the crystal growth axis coincides with the c-axis. Because it is more difficult to achieve crystals of high quality for nitride systems grown on non-polar substrates [90], c-plane nitrides remain the usual choice when trying to construct semiconductor nanostructures such as quantum wells (QWs) and quantum dots (QDs). Therefore, the charge accumulation remains a problem.

The main consequence of this interfacial charge accumulation is the large built-in electric field present in heterostructures where wurtzite nitrides are employed [91]. Among other explanations, including Auger recombination [92] and defect-related delocalization of carriers [93], it has been suggested that these built-in fields are the origin of carrier leakage that leads to the dramatic *efficiency droop* observed as the drive current is increased in GaN-based multiple quantum well LEDs [94, 95]. The built-in field leads to spatial separation of the electrons and

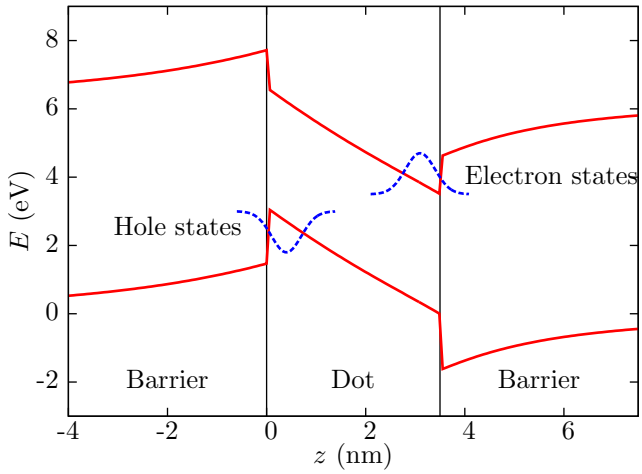


Figure 3.1: Schematic band diagram representing the electron and hole spatial separation in a GaN/AlN QD when a strong built-in field is present. z is assumed parallel to the c-axis.

holes in heterostructures, and to a dramatic reduction in the optical recombination rate [91, 96, 97], as schematically shown in Fig. 3.1. In order to overcome this charge carrier separation, Kim *et al.* [94] proposed the use of polarization-matched barriers and InGaN QWs. Later on, the same group of researchers reported an enhanced performance of those structures by using partially polarization-matched barriers [98–100]. We present here a generalization of that concept and investigate the potential for polarization-matching and minimization of built-in fields in III-N QW and QD structures, based on the results given in our previous papers [1, 2].

We start in Section 3.2.1 with a brief introduction to the methods we use to calculate the built-in fields in QW and QD structures. A discussion of different interpolation methods for the relevant material parameters in nitride alloys (including the elastic and piezoelectric properties) is presented in Section 3.2.2. Section 3.3 overviews the principles behind the concept of *built-in field control* and then presents the field suppression results for nitride systems grown both with unstrained barriers and also with barriers strained to the in-plane lattice constant of an underlying GaN substrate. This analysis is extended in Section 3.3.4 to consider experimentally observed constraints, including strain relaxation and alloy miscibility issues. A detailed study of the UV spectral region and the impact of the AlInN band gap bowing is given in Section 3.3.5. Finally we summarize and present our conclusions in Section 3.4.

3.2 Calculation of the built-in field

3.2.1 Built-in potential in quantum dots and wells

In the case of a QW, simple arguments can be employed to obtain the polarization potential $\varphi^{\mathrm{QW}}(z)$ along the growth direction ([0001] direction or c-axis) as a function of both spontaneous and strain-related piezoelectric polarization vectors. Assuming that stress is applied in the basal plane only, both the spontaneous $\mathbf{P}_{\mathrm{sp}} = P_{\mathrm{sp}}\hat{\mathbf{z}}$ and piezoelectric polarizations $\mathbf{P}_{\mathrm{pz}} = P_{\mathrm{pz}}\hat{\mathbf{z}}$ are constant vectors along the [0001] direction within the QW and within the barrier [89]. In the following, the z axis will be always considered to be parallel to the [0001] direction. The total built-in potential $\varphi^{\mathrm{QW}}(z)$ within the QW is given by:

$$\begin{aligned}\varphi^{\mathrm{QW}}(z) &= \varphi_{\mathrm{sp}}^{\mathrm{QW}}(z) + \varphi_{\mathrm{pz}}^{\mathrm{QW}}(z) \\ &= \left\{ \frac{(P_{\mathrm{sp}}^{\mathrm{W}} - P_{\mathrm{sp}}^{\mathrm{B}}) + P_{\mathrm{pz}}^{\mathrm{W}}}{2\epsilon_0 \epsilon_r^{\mathrm{W}}} \right\} \big(|z| - |z-h| \big),\end{aligned} \tag{3.1}$$

where the QW is of height h, with interfaces at $z = 0$ and $z = h$, ϵ_r^{W} is the QW dielectric constant, and the piezoelectric polarization vector $P_{\mathrm{pz}}^{\mathrm{W}}$ in the QW is given by [89]

$$P_{\mathrm{pz}}^{\mathrm{W}} = 2\epsilon_{xx} e_{31}^{\mathrm{W}} + \epsilon_{zz} e_{33}^{\mathrm{W}}. \tag{3.2}$$

with

$$\epsilon_{zz} = -2 \frac{C_{13}^{\mathrm{W}}}{C_{33}^{\mathrm{W}}} \epsilon_{xx}. \tag{3.3}$$

The well and the barrier are denoted with the indices W and B, respectively, $\epsilon_{xx} = (a_{\mathrm{B}} - a_{\mathrm{W}})/a_{\mathrm{W}}$ is the basal strain, ϵ_{zz} the strain along the c-axis, e_{ij} are the piezoelectric coefficients and C_{ij} are the elastic constants.

The calculation of the total built-in potential φ in a QD is more complicated because the piezoelectric polarization vector is no longer a constant vector along the [0001] direction within the QD and within the barrier. To calculate the polarization potential in nitride-based QDs we apply a real-space surface integral approach developed by Williams *et al.* [101]. This method admits analytical solutions in certain cases and provides an extremely useful insight into the parameters that influence the magnitude and the shape of the polarization potential. In this work, we use the solutions obtained in Ref. [101] for a linescan of the poten-

tial $\varphi^{\mathrm{QD}}(z)$ along the z direction through the center of a truncated-cone shaped QD. Using the notation of Ref. [101], $\varphi^{\mathrm{QD}}(z)$ can be found along the central axis ($x = y = 0$) by evaluating the following surface integrals:

$$\varphi^{\mathrm{QD}}(z) = JI_1 + \left[\frac{P_{\mathrm{sp}}^{\mathrm{D}} - P_{\mathrm{sp}}^{\mathrm{B}}}{4\pi\epsilon_0\epsilon_r} + K\right] I_2, \tag{3.4}$$

where

$$\begin{aligned} I_1 &= \int\limits_{\mathrm{QD}} \frac{(z-z')^2}{\left[x'^2 + y'^2 + (z-z')^2\right]^{\frac{3}{2}}} \hat{\mathbf{z}} \cdot \mathrm{d}\mathbf{s}', \\ I_2 &= \int\limits_{\mathrm{QD}} \frac{1}{\left[x'^2 + y'^2 + (z-z')^2\right]^{\frac{1}{2}}} \hat{\mathbf{z}} \cdot \mathrm{d}\mathbf{s}', \end{aligned} \tag{3.5}$$

and where D refers to the dot and B to the barrier. The analytical solution for these integrals, as well as an expression for the material- and strain-dependent coefficients J and K, are given in Ref. [101].

Finally, the built-in field $E(z)$ corresponding to the potentials given in Eqs. (3.1) and (3.4) is calculated as the derivative of φ with respect to the variable z:

$$E(z) = -\frac{\mathrm{d}\varphi(z)}{\mathrm{d}z}. \tag{3.6}$$

Because the piezoelectric potential arises from the strain-induced polarization, it is crucial to correctly identify which are the *straining* and *strained* layers. In Sections 3.3.1 and 3.3.2 we consider the ideal case in which only a well (or dot) and barrier are present in the structure. Since the dot is completely embedded in the barrier, the relaxation mechanisms in QDs and their consequences for the piezoelectric potential are complex and have to be modeled using Eq. (3.4). In contrast, the description of strain relaxation and its consequences is much simpler for a QW, with the well relaxing along the c-axis as in Eq. (3.3) and with the barrier remaining strain-free. In Section 3.3.3 we extend the analysis to consider the case where the structures are grown on a GaN buffer layer and, therefore, both barrier and well (or dot) experience a modified strain to account for the lattice mismatch with respect to the substrate.

3.2.2 Nitride alloy material parameters

Although the calculation of the potentials in Eqs. (3.1) and (3.4) is straightforward, the value of the material-dependent constants involved in those equations is not so well established. In their review paper from 2003, Vurgaftman and Meyer [20] give a range of parameter values found in the literature for the nitride binaries and propose a *best choice*. Even though their compilation has probably been the set of parameters most widely used by the community over the last few years, the progress in the knowledge of the properties of nitride materials make some of these values inevitably obsolete and a certain degree of revision is desirable.

Some controversy surrounds the sign of the piezoelectric coefficient e_{15}, which is of relative importance in the calculation of the piezoelectric potential in QDs. Further discussion on this issue as well as experimental evidence that supports the use of a negative value for e_{15} has been given by Schulz *et al.* [6, 8, 102] and by Shen *et al.* [103]. Here, we apply the piezoelectric coefficients of Shimada [104], who derives a negative sign for e_{15} and obtains values of e_{31} and e_{33} in line with those reported by Vurgaftman and Meyer [20]. The spontaneous polarization values are also taken from Ref. [20]. The dielectric constants are taken from Wagner and Bechstedt [105] for GaN and AlN, and Furthmüller *et al.* [106] for InN. The parameter values which we use for nitride binaries are listed in Table 3.1. Further comment on the AlInN band gap bowing parameter will follow in Section 3.3.5.

The parameters for ternary and quaternary alloys are in general less well known. In the absence of knowledge on the dependence with composition of a particular parameter, it is usual to take a linear interpolation of the binary values:

$$\xi(\mathrm{Al}_x\mathrm{In}_y\mathrm{Ga}_{1-x-y}\mathrm{N}) = x\xi(\mathrm{AlN}) + y\xi(\mathrm{InN}) + (1-x-y)\xi(\mathrm{GaN}). \tag{3.7}$$

Equation (3.7) is known as Végard's law. It is widely used for the calculation of general nitride parameters, such as the lattice constants [20,107], and has also been used to estimate built-in fields in A_xB_{1-x}N QDs [101, 108]. Where more detailed information is known, it can be useful for some A_xB_{1-x}N parameters to use a quadratic interpolation between the values of the binaries AN and BN via the introduction of the so called *bowing parameter* $b_{AB\mathrm{N}}$ [20, 89, 109], which describes

Table 3.1: If not indicated otherwise, all material parameters for the wurtzite nitride binaries are taken from Ref. [20]. The energy gap of GaN and AlN at $T = 300$ K is estimated by using the Varshni formula and the parameters reported in Ref. [20].

	GaN	AlN	InN
E_g (eV) at T=300 K	3.44	6.16	0.64[1]
a (Å) at T=300 K	3.189	3.112	3.545
c (Å) at T=300 K	5.185	4.982	5.703
$P_{\rm sp}$ (C/m^2)	-0.034	-0.090	-0.042
e_{15} (C/m^2)	-0.38[2]	-0.41[2]	-0.44[2]
e_{31} (C/m^2)	-0.45[2]	-0.45[2]	-0.52[2]
e_{33} (C/m^2)	0.83[2]	1.54[2]	0.95[2]
C_{12} (GPa)	145	137	115
C_{13} (GPa)	106	108	92
C_{33} (GPa)	398	373	224
C_{44} (GPa)	105	116	48
ϵ_r (F/m)	9.6[3]	8.5[3]	7.2[4]

[1]Ref. [21]. [2]Ref. [104]. [3]Ref. [105]. [4]Ref. [106].

Table 3.2: Bowing parameters for wurtzite III-N ternaries used in this paper, taken from Ref. [20] unless otherwise specified.

	InGaN	AlGaN	AlInN
E_g (eV)	1.4[1]	0.7	5.0[1]
$P_{\rm sp}$ (C/m^2)	-0.037	-0.021	-0.070

[1]Ref. [21].

the variation of the parameter ξ as

$$\xi(A_xB_{1-x}\mathrm{N}) = x\xi(A\mathrm{N}) + (1-x)\xi(B\mathrm{N}) - x(1-x)b_{AB\mathrm{N}}. \tag{3.8}$$

The key bowing parameters which have been proposed for wurtzite ternary III-N compounds are listed in Table 3.2.

For quaternary alloys, Glisson *et al.* [110] proposed an expression to interpolate the values of the ternaries, including bowing, in order to obtain a better approximation than the one provided by a simple Végard's approach. This interpolation formula has been commonly used in previous works to deal with the

quaternary [19, 110, 111]. We use the notation introduced by Chen [111]:

$$\xi(\mathrm{Al}_x\mathrm{In}_y\mathrm{Ga}_z\mathrm{N}) = \frac{xy\xi^u(\mathrm{AlInN}) + yz\xi^v(\mathrm{InGaN}) + xz\xi^w(\mathrm{AlGaN})}{xy + yz + xz}, \tag{3.9}$$

with

$$\begin{aligned}
&\xi^u = u\xi(\mathrm{InN}) + (1-u)\xi(\mathrm{AlN}) - u(1-u)b_{\mathrm{AlInN}},\\
&\xi^v = v\xi(\mathrm{GaN}) + (1-v)\xi(\mathrm{InN}) - v(1-v)b_{\mathrm{InGaN}},\\
&\xi^w = w\xi(\mathrm{GaN}) + (1-w)\xi(\mathrm{AlN}) - w(1-w)b_{\mathrm{AlGaN}},\\
&u = \frac{1-x+y}{2}, \qquad v = \frac{1-y+z}{2}, \qquad w = \frac{1-x+z}{2} \qquad \text{and} \qquad x+y+z = 1.
\end{aligned}$$

There is no data available in the literature that provides a reliable way to include any nonlinear dependence of the elastic constants C_{ij} or the piezoelectric coefficients e_{ij} into the interpolation formulas. Fiorentini *et al.* [89, 112] presented a model for ternary compounds that provides in principle a more accurate description of the piezoelectric polarization in a QW as a second order function of the applied basal strain ϵ_{xx}. They provide quadratic formulas to obtain the piezoelectric response for the binaries, and propose the application of Végard's law to those formulas in order to calculate the piezoelectric polarization in the ternaries. As discussed in further detail in Ref. [89], their results for GaN and InN are only valid for compressive strain ($\epsilon_{xx} < 0$), whereas the results for AlN are applicable also for tensile strain.

Figure 3.2 compares the piezoelectric polarization derived from the formulas given in Ref. [112] by Fiorentini *et al.* to the one expected from Eq. (3.2) for a QW using the piezoelectric coefficients calculated by Shimada [104]. The elastic constants reported in Ref. [20] are used. From these results one can conclude that, up to relatively high strain ($|\epsilon_{xx}| \lesssim 6\%$), the two approaches do not differ significantly. Therefore, given the large uncertainties concerning the calculation of accurate polarization-related parameters in III-N materials, and the fact that the expressions derived by Fiorentini *et al.* [112] cannot be used for QD systems, we apply here Végard's law to calculate the piezoelectric constants e_{ij} of the alloys using the values of the binaries reported in Ref. [104] by Shimada.

To calculate the band gap of the materials, we assume $\epsilon_{xx} = \epsilon_{yy} \neq \epsilon_{zz}$ and

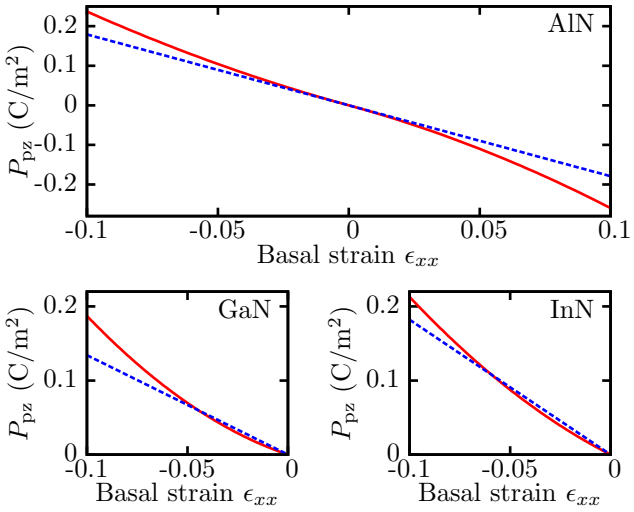


Figure 3.2: Piezoelectric polarization in nitride binaries. Solid lines indicate the quadratic formulas given in Ref. [112] and dashed lines are Eq. (3.2) using the piezoelectric coefficients and elastic constants found in Table 3.1.

choose the smaller of the following transition energies [113]:

$$\begin{aligned} E_{A/B} =& E_{A/B}(0) + (a_{cz} - D_1 - D_3)\epsilon_{zz} + 2(a_{ct} - D_2 - D_4)\epsilon_{xx}, \\ E_C =& E_C(0) + (a_{cz} - D_1)\epsilon_{zz} + 2(a_{ct} - D_2)\epsilon_{xx}. \end{aligned} \tag{3.10}$$

Here $E_{A/B}$ denotes the transition energy between the conduction band and the A (heavy hole) and B (light hole) valence bands. Since we have neglected the weak spin-orbit coupling, in the unstrained system, the A- and B-valence bands are degenerate at the Brillouin zone centre [$E_A(0) = E_B(0)$] [114]. For the unstrained material, the transition energy between the conduction band and the crystal-field split-off band (C valence band) is given by $E_C(0) = E_{A/B}(0) + \Delta_{\mathrm{cr}}$, where Δ_{cr} denotes the crystal-field splitting energy. The valence band deformation potentials are denoted by D_i, while the conduction band deformation potentials are given by a_{cz} and a_{ct}. The values are taken from Yan *et al.* [113]. For the crystal-field splitting energy Δ_{cr} we use values of Vurgaftman and Meyer [20] (cf. Table 3.3). Végard's law [Eq. (3.7)] is applied to obtain the deformation potentials of the ternary and quaternary compounds.

As previously mentioned, in the case of a QD the strain components vary throughout the structure. In line with the surface integral approach [101] employed for the calculation of the piezoelectric potential, we can also use a surface integral

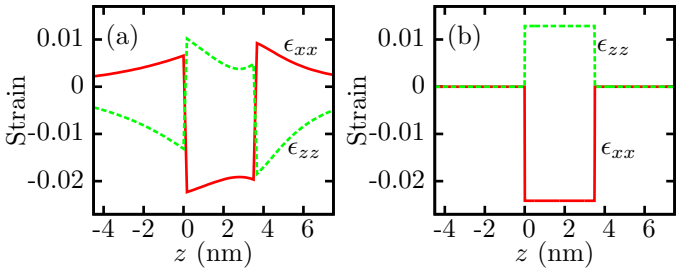


Figure 3.3: (a) Strain components ϵ_{xx} (solid line) and ϵ_{zz} (dashed line) through the centre of a QD along the z direction (parallel to c-axis) for a GaN/AlN truncated-cone shaped dot with height $h = 3.5$ nm, base radius $R_b = 8$ nm (located at $z = 0$) and top radius $R_t = 2$ nm (located at $z = h$) and (b) for a QW of the same height and composition.

Table 3.3: Deformation potentials for nitride binaries, from Yan *et al.* [113]. Crystal-field splitting values taken from Vurgaftman and Meyer [20].

	GaN	AlN	InN
$a_{cz} - D_1$ (eV)	-5.81	-4.31	-3.62
$a_{ct} - D_2$ (eV)	-8.92	-12.11	-4.60
D_3 (eV)	5.47	9.12	2.68
D_4 (eV)	-2.98	-3.79	-1.74
Δ_{cr} (eV)	0.010	-0.169	0.040

approach to obtain the strain in a QD [115, 116]. For a truncated-cone shaped dot which is under compressive strain (e.g. GaN/AlN[1]), ϵ_{xx} and ϵ_{zz} follow the trends shown in Fig. 3.3(a), where a substantial difference can be observed between the strain at the bottom and top interfaces, in contrast to the constant strain across a QW [Fig. 3.3(b)]. To account for this effect, the values of the strain components ϵ_{xx} and ϵ_{zz} are taken as the value at a half of the dot's height when calculating the band gap in QD structures. For the barrier, the band gap value for the "relaxed" barrier (i.e. far enough from the dot, so that $\epsilon_{xx} = \epsilon_{zz} \approx 0$ is valid) is used.

3.3 Built-in field reduction

In this section we present a detailed study of the built-in field reduction which can be obtained in ternary and quaternary structures by applying polarization-

[1]The convention here is to refer to the dot or well before the barrier, i.e. "GaN/AlN" refers to either a GaN dot or well embedded in an AlN barrier.

matched barriers. In a first step, Sec. 3.3.1, we introduce the principles behind the concept of polarization matching. In Section 3.3.2 we show the field suppression results for QW and QD systems grown on ideally unstrained barriers, while results for systems with barriers strained to the in-plane lattice constant of an underlying GaN substrate are discussed in Sec. 3.3.3. All calculations are performed at room temperature. Quantum size effects, which will decrease the effective band offsets and increase the effective transition energies, are not included in this analysis. The exact calculation of these energy shifts is not straightforward as they show a strong dependence on the carrier effective masses as well as the conduction band to valence band offset ratio. The conduction band to valence band offset ratio is still a matter of discussion, with a broad range of different values reported in the literature [20,21,111,117–120]. In addition to this broad spectrum of different values, the interpolation procedure is not well established for quaternary alloys, including the effect of bowing. In view of these uncertainties, we therefore do not attempt to estimate band offsets in this thesis. Conventional InGaN/GaN and GaN/AlGaN systems display Type-I band alignment [121, 122]. Hums *et al.* [123] report a transition from Type-I to Type-II alignment in GaN/AlInN systems above 25% In content. Since we will be focusing mostly on nearly lattice- and polarization-matched structures, transitivity holds and the band alignment of the presented InGaN/AlInGaN structures is expected to be Type-I up to at least 25% In in the barrier, following Ref. [123]. For InGaN/AlInGaN systems above this limit, and AlGaN/AlIn(Ga)N systems (which have not been sufficiently studied to date), Type-II transitions cannot be excluded *a priori*. Taking all these issues into account, further theoretical and experimental investigations on the band offsets in nitride systems are required. This is beyond the scope of the work in this thesis. Therefore, we present here the energy gap E_g of the QW (QD) material and the difference in the energy gap ΔE_g between the QW (QD) and the surrounding barrier material as a guide to a particular spectral region, and focus on the relation between built-in field and alloy composition, which is independent of quantum size effects.

3.3.1 GaN/AlInN heterostructures

Figure 3.4 shows the built-in field for a GaN QW (QD) embedded in an AlN (a) or InN (b) matrix, calculated using Eqs. (3.1) and (3.4). It can be seen that the

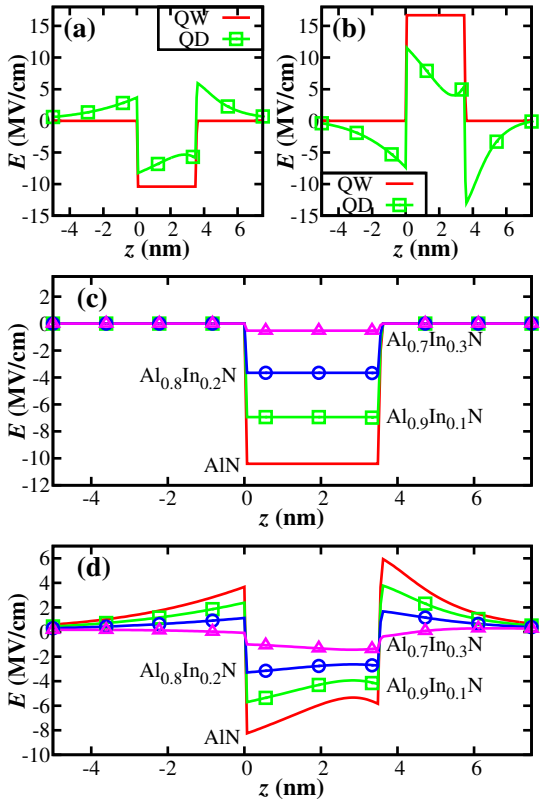


Figure 3.4: Built-in field along the c-axis for (a) GaN/AlN and (b) GaN/InN QD and QW. (c) Built-in field along the c-axis for GaN/$Al_xIn_{1-x}N$ QW and (d) QD, as the barrier composition is varied from $x = 1$ to $x = 0.7$. QW and QD of typical dimensions (well: height $h = 3.5$ nm; truncated-cone dot: base radius $R_b = 8$ nm, top radius $R_t = 2$ nm and height $h = 3.5$ nm).

resultant built-in field is of opposite sign in the GaN/InN structure compared to the GaN/AlN case, leading to the idea that an *ad hoc* superposition of the curves shown in Figs. 3.4(a) and (b) could lead to effective cancelation of the built-in field [94]. We illustrate this principle here by considering GaN as the well or dot material and $Al_xIn_{1-x}N$ as the barrier material. It can be seen in Figs. 3.4(a) and (b) that the calculated built-in field is in both cases reduced in the QD structure compared to the QW case. This occurs both because of the reduced surface area of a dot compared to a QW (reduced area of induced surface charge) and also because of the reduction in the magnitude of ϵ_{xx} and ϵ_{zz} in the QD structure [124, 125], as seen in Fig. 3.3. The growth of QDs can therefore provide an additional route to field reduction in c-plane III-N heterostructures [125, 126]. A recent experimental work reported reduced efficiency droop employing InGaN/GaN QDs [127], although it should be noted that the QDs studied presented a very large aspect

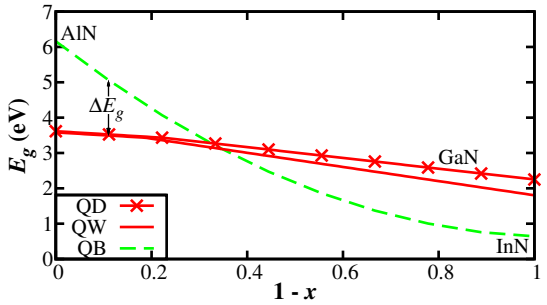


Figure 3.5: Variation of the band gap difference ΔE_g (at room temperature) between the GaN QD or QW and the $Al_xIn_{1-x}N$ barrier (QB) as x decreases in the GaN/$Al_xIn_{1-x}N$ system. The same dot dimensions are assumed as in Fig. 3.4.

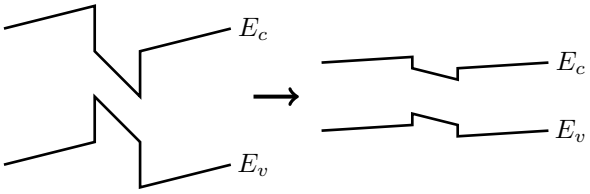


Figure 3.6: Schematic band diagram of a "traditional" nitride QW with a strong built-in field (left) and its counterpart, optimized in terms of built-in field reduction, with the same band gap within the active layer but reduced band offsets.

ratio. Figs. 3.4(c) and (d) depict the calculated built-in field in GaN/$Al_xIn_{1-x}N$ QW and QD structures, showing that a complete suppression of the built-in field can be achieved for $x \approx 0.7$ in both QW and QD structures, at the expense of reducing the initial energy gap difference between dot and barrier from ~ 2.5 eV to ~ 0.3 eV (Fig. 3.5). Figure 3.6 schematically highlights the principle behind built-in field control, which is achieved at the expense of reducing the band offsets. This reduction of the band offsets is a drawback to polarization matching, and a consequence of the fact that there is a close relation between band gap and lattice constant in nitride materials. Because of this relation, two nitride alloys that present a low lattice mismatch (necessary in order to achieve a reduced piezoelectric component of the total polarization) have also similar band gaps. However, Ref. [100] has shown how the loss of confinement is not critical compared to the benefits obtained from field reduction.

3.3.2 Quaternary structures: Unstrained barriers

We turn now to show how the use of unstrained quaternary barrier structures can enable field reduction for a very wide range of QW and QD band gap en-

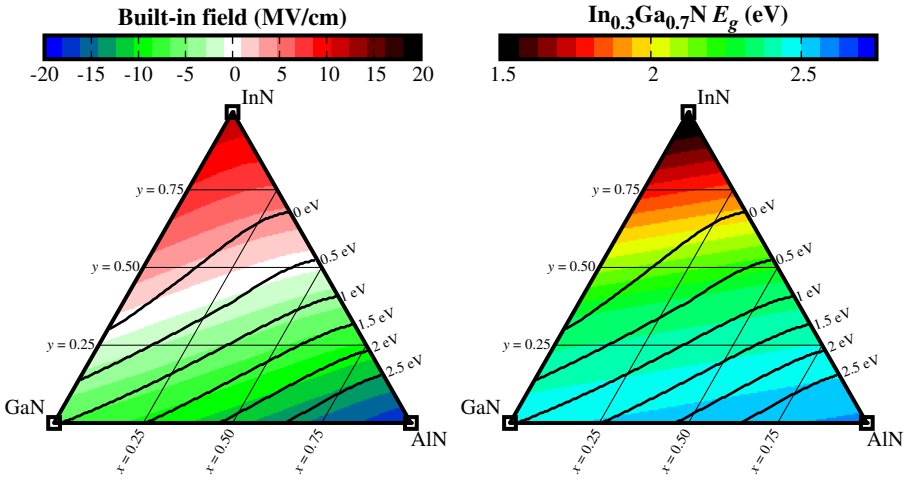


Figure 3.7: (Left) Calculated built-in field for an $In_{0.3}Ga_{0.7}N/Al_xIn_yGa_{1-x-y}N$ QW system and (right) energy gap E_g, at room temperature, of the $In_{0.3}Ga_{0.7}N$ layer as the barrier composition (and hence strain) varies. The thick solid lines indicate the band gap difference ΔE_g between well and barrier.

ergies, supporting efficient optical emission across a wide wavelength range. We use the interpolation formulas from Section 3.2.2, and illustrate the method by considering the range of polarization-matched barriers which can be achieved with an $In_{0.3}Ga_{0.7}N/Al_xIn_yGa_{1-x-y}N$ QW system. The coloured contours in the left panel in Fig. 3.7 show the calculated built-in field across an $In_{0.3}Ga_{0.7}N$ QW as a function of unstrained barrier quaternary composition, while the contours in the right panel show the calculated energy gap of the $In_{0.3}Ga_{0.7}N$ QW strained to the barrier lattice constant. The solid lines in both panels show the difference in energy gap, ΔE_g between the QW and surrounding barrier layer. A whole range of barrier compositions (characterized by the white area in the left panel) is found to effectively suppress the field, and the energy gap of $In_{0.3}Ga_{0.7}N$ is observed to be fixed around 2.25 eV for that range (right panel).

In the most general case, a quaternary alloy can be employed in both well (dot) and barrier. Doing so, the system gains additional degrees of freedom that allow, for a given band gap E_g, to minimize the built-in field value and maximize the band gap difference ΔE_g between well and barrier. Figure 3.8 (top) shows a schematic illustration of the optimization procedure used to calculate the system composition, compatible with certain E_g and ΔE_g, for which the built-in field is minimum. As an example, the calculated minimum built-in field as a function of energy gap difference, ΔE_g, between a strained QW or QD with 2.25 eV energy

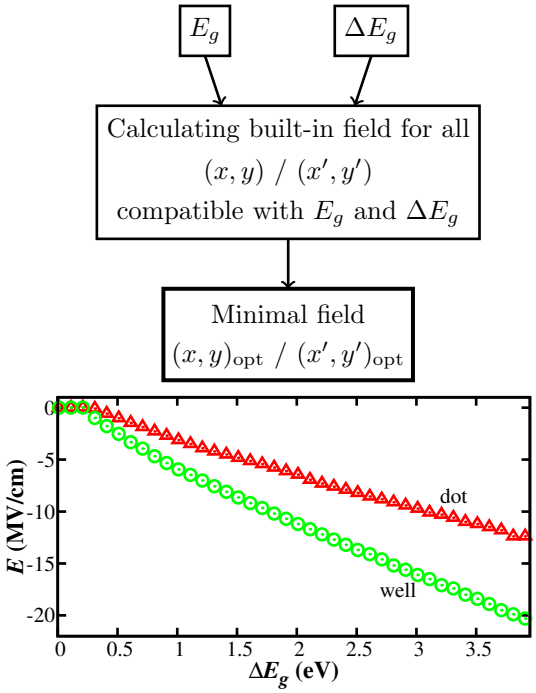


Figure 3.8: (top) Flux diagram explaining how, for a certain combination of E_g and ΔE_g at room temperature, the compositions of the $Al_xIn_yGa_{1-x-y}N/Al_{x'}In_{y'}Ga_{1-x'-y'}N$ system that minimize the built-in field are obtained and (bottom) smallest built-in field that can be obtained for a QW and a QD system where $E_g = 2.25$ eV as the band gap difference ΔE_g between the barrier and well (or dot) increases.

gap and an unstrained barrier layer is shown in Fig. 3.8 (bottom). In this case, it is possible to achieve a maximum offset close to 0.3 eV in both the QD and QW cases. We note for larger offsets that the absolute value of the minimum built-in field achievable is generally smaller in the QD than the QW case, as expected from the earlier discussion of Figs. 3.4(a) and (b). Table 3.4 shows the calculated QW/barrier and QD/barrier compositions of built-in field-free systems which maximize the offset ΔE_g between the strained QW (QD) structure and unstrained barrier layers, as a function of increasing strained QW (QD) energy gap. In the case of $E_g = 2.25$ eV this corresponds to the last zero-field point in Fig. 3.8 (bottom). The table also shows the lattice mismatch between the QW (QD) and barrier materials. In the case of the QW, the lattice mismatch equals ϵ_{xx} and in the case of the QD it equals the in-plane initial misfit strain $\epsilon_{0,a}$ [101].

Table 3.4: List of QW and QD $Al_xIn_yGa_{1-x-y}N/Al_{x'}In_{y'}Ga_{1-x'-y'}N$ systems compatible with zero built-in field that, for a given E_g at room temperature, maximize the band gap difference ΔE_g between well (dot) and barrier. The lattice mismatch is $(a_B - a_{D/W})/a_{D/W}$ for the unstrained barrier case, and $(a_{GaN} - a_{D/W})/a_{D/W}$ for the dot or well and $(a_{GaN} - a_B)/a_B$ for the barrier in the case of pseudomorphic growth on a GaN buffer layer.

	QW				QD			
E_g (eV)	ΔE_g (eV)	$(x,y)_{opt}/(x',y')_{opt}$	Lattice mismatch (%)		ΔE_g (eV)	$(x,y)_{opt}/(x',y')_{opt}$	Lattice mismatch (%)	
			Unstrained barriers					
2.00	0.24	(0.46,0.54)/(0.25,0.46)	-0.4		0.32	(0.43,0.57)/(0.03,0.32)	-1.7	
2.25	0.22	(0.51,0.49)/(0.32,0.42)	-0.3		0.31	(0.50,0.50)/(0.27,0.37)	-0.9	
2.50	0.20	(0.55,0.45)/(0.36,0.38)	-0.3		0.29	(0.54,0.46)/(0.32,0.34)	-0.8	
2.75	0.21	(0.00,0.13)/(0.43,0.34)	1.3		0.27	(0.58,0.42)/(0.36,0.30)	-0.8	
3.00	0.27	(0.01,0.06)/(0.51,0.30)	1.5		0.25	(0.62,0.38)/(0.42,0.27)	-0.7	
3.25	0.32	(0.02,0.00)/(0.52,0.24)	1.5		0.24	(0.66,0.34)/(0.48,0.24)	-0.7	
3.50	0.31	(0.13,0.00)/(0.58,0.22)	1.4		0.22	(0.69,0.31)/(0.51,0.21)	-0.7	
3.75	0.30	(0.23,0.00)/(0.64,0.20)	1.2		0.20	(0.72,0.28)/(0.53,0.17)	-0.8	
4.00	0.28	(0.33,0.00)/(0.70,0.18)	1.1		0.18	(0.76,0.24)/(0.62,0.16)	-0.6	
			Pseudomorphic growth on GaN					
			Well	Barrier			Dot	Barrier
2.00	0.31	(0.42,0.58)/(0.32,0.52)	-5.2	-4.8	0.18	(0.42,0.58)/(0.24,0.53)	-5.2	-5.1
2.25	0.29	(0.00,0.36)/(0.37,0.47)	-3.9	-4.2	0.19	(0.47,0.53)/(0.31,0.48)	-4.6	-4.4
2.50	0.30	(0.00,0.28)/(0.41,0.42)	-3.0	-3.6	0.19	(0.51,0.49)/(0.28,0.40)	-4.1	-3.7
2.75	0.32	(0.00,0.20)/(0.46,0.37)	-2.2	-2.9	0.19	(0.56,0.44)/(0.38,0.37)	-3.4	-3.1
3.00	0.34	(0.01,0.13)/(0.49,0.32)	-1.4	-2.3	0.19	(0.60,0.40)/(0.43,0.33)	-2.9	-2.6
3.25	0.37	(0.00,0.05)/(0.58,0.29)	-0.6	-1.8	0.19	(0.63,0.37)/(0.36,0.23)	-2.5	-1.7
3.50	0.39	(0.04,0.00)/(0.62,0.25)	0.1	-1.3	0.20	(0.67,0.33)/(0.49,0.24)	-2.0	-1.5
3.75	0.36	(0.17,0.00)/(0.65,0.22)	0.4	-0.9	0.20	(0.71,0.29)/(0.51,0.18)	-1.5	-0.8
4.00	0.31	(0.30,0.00)/(0.68,0.18)	0.7	-0.4	0.18	(0.75,0.25)/(0.56,0.14)	-1.0	-0.2

3.3.3 Pseudomorphic barriers on a GaN substrate

In addition to the structures with unstrained barriers, we consider here also systems where both the barrier and the QW (QD) materials are grown pseudomorphically on a substrate with a different lattice constant. Therefore, this analysis gives an insight into the behavior of the built-in fields when the barrier exhibits residual strain due to an underlying substrate. Since nitride-based heterostructures are commonly grown on a GaN buffer layer, we investigate here the compositions and strains required for field-matched QW/barrier and QD/barrier structures grown on an unstrained GaN epilayer. A similar approach was also used in Refs. [94] and [111]. In general, both the barrier and QW (QD) layers will be strained in such structures, so that the piezoelectric charges and hence the built-in field will be modified compared to the previous case.

Since the barrier material is strained to the GaN buffer layer, the total built-in polarization of the barrier material is no longer given by the spontaneous polarization only. Here, the (spontaneous) polarization of the barrier material in Eqs. (3.1) and (3.4) has to be replaced by

$$P^{\mathrm{B}}_{\mathrm{tot}} = P^{\mathrm{B}}_{\mathrm{sp}} + P^{\mathrm{B}}_{\mathrm{pz}}. \tag{3.11}$$

$P^{\mathrm{B}}_{\mathrm{pz}}$ can be obtained from Eq. (3.2), assuming that the basal strain applied to the barrier is given by

$$\epsilon^{\mathrm{B}}_{xx} = \frac{a^{\mathrm{GaN}} - a^{\mathrm{B}}_0}{a^{\mathrm{B}}_0}, \tag{3.12}$$

where a^{B}_0 and a^{GaN} refer to the a lattice constant of the fully relaxed barrier material and of GaN, respectively. To obtain the piezoelectric polarization in the well, only the consideration that it is strained in the basal plane to the GaN lattice constant a^{GaN} instead of a^{B}_0 has to be made:

$$\epsilon^{\mathrm{W}}_{xx} = \frac{a^{\mathrm{GaN}} - a^{\mathrm{W}}_0}{a^{\mathrm{W}}_0}. \tag{3.13}$$

In both cases, the ϵ_{zz} component of the strain is then given by Eq. (3.3).

As pointed out previously, when basal strain alone is applied to a material layer, a relaxation takes place in the [0001] direction (perpendicular to that plane)

as given by Eq. (3.3). When dealing with the QD in the strained barrier case, this relaxation will imply a change in the *effective* lattice constants of the barrier. Therefore, to calculate the strain and the built-in field in such a system, we proceed in the following way. First, the dot embedded in an unstrained barrier will be considered and the surface integral method is used to obtain the built-in field profile as described in Section 3.3.2. Second, the basal strain corresponding to the introduction of the GaN buffer layer will be added and the previous structure will be assumed to follow the relaxation mechanisms of the barrier alone, as expected from Eqs. (3.12) and (3.3). Therefore, the additional contribution to the QD built-in potential φ^{QD} in Eq. (3.4) will be

$$\Delta\varphi^{\mathrm{QD}} = \frac{\Delta P^{\mathrm{D}}_{\mathrm{pz}} - \Delta P^{\mathrm{B}}_{\mathrm{pz}}}{4\pi\epsilon_0\epsilon_r} I_2, \tag{3.14}$$

with the additional introduced piezoelectric polarization being

$$\Delta P^{\mathrm{D}}_{\mathrm{pz}} = 2e^{\mathrm{D}}_{31}\Delta\epsilon^{\mathrm{D}}_{xx} + e^{\mathrm{D}}_{33}\Delta\epsilon^{\mathrm{D}}_{zz}, \tag{3.15}$$

and

$$\Delta P^{\mathrm{B}}_{\mathrm{pz}} = 2e^{\mathrm{B}}_{31}\Delta\epsilon^{\mathrm{B}}_{xx} + e^{\mathrm{B}}_{33}\Delta\epsilon^{\mathrm{B}}_{zz}, \tag{3.16}$$

where we have assumed

$$\Delta\epsilon^{\mathrm{D}}_{xx} = \Delta\epsilon^{\mathrm{B}}_{xx}, \qquad \Delta\epsilon^{\mathrm{D}}_{zz} = \Delta\epsilon^{\mathrm{B}}_{zz}, \tag{3.17}$$

and

$$\Delta\epsilon^{\mathrm{B}}_{xx} = \frac{a^{\mathrm{GaN}} - a^{\mathrm{B}}_0}{a^{\mathrm{B}}_0}, \qquad \Delta\epsilon^{\mathrm{B}}_{zz} = -2\frac{C^{\mathrm{B}}_{13}}{C^{\mathrm{B}}_{33}}\Delta\epsilon^{\mathrm{B}}_{xx}. \tag{3.18}$$

Compared to the previous case with unstrained barriers, pseudomorphic growth on a GaN epilayer allows similar results, as can be seen in Fig. 3.9, which shows the calculated minimum field achievable as a function of band gap difference, ΔE_g, between the QW (QD) and barrier layer for $E_g = 2.25$ eV in the QW (QD) region. Table 3.4 shows the calculated QW/barrier and QD/barrier compositions as a function of increasing QW (QD) energy gap to maximize the offset ΔE_g between the strained QW (QD) and strained barrier layers. The table also shows the lattice

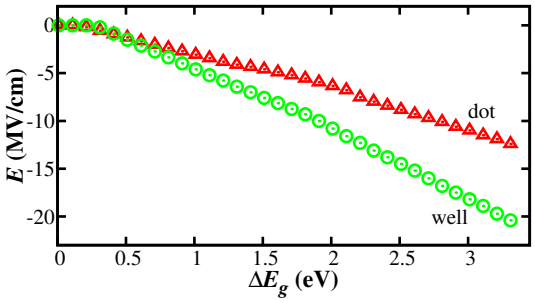


Figure 3.9: Lowest built-in field achievable, as a function of ΔE_g, for a QD and a QW system grown on a GaN substrate for which E_g=2.25 eV.

mismatch in the QW (QD) and the barrier layers. The additional contribution to the strain assumed in Eq. (3.17) is found to be of a higher order of magnitude than the main contribution to the strain field modelled via Eq. (3.5) for the low energy gap structures in the right hand side of Table 3.4. For the sake of completeness, and to be able to compare these results with those obtained in the previous section, we have included the results for the lower-gap structures ($E_g \lesssim 3.5$ eV), even though these systems are impractical to grow, given the large lattice mismatch between the GaN substrate and the barrier layers. For $E_g = 3.5$ eV and above, the mismatch is within the range which could probably be accommodated to allow pseudomorphic growth on GaN. In this range, the contributions to the field from Eqs. (3.5) and (3.17) are found to be of the same order and therefore our approximation should be more accurate within that range. As discussed in the introduction to this chapter, a built-in field reduction in systems with $E_g \geq 3.5$ eV, is of strong interest for applications operating in the UV spectral range [128, 129]. This range will be studied in more detail in Section 3.3.5.

3.3.4 Strain and miscibility limitations to built-in field suppression

The structures studied in Sections 3.3.2 and 3.3.3, and listed in Table 3.4, depict a best-case scenario for the growth of III-N heterostructures where the intrinsic built-in field has been calculated to vanish, while maximizing the energy gap difference between the barrier and well or dot layers. However, these results do not consider any constraints encountered during the epitaxial growth of nitride-based heterostructures. Particularly, the amount of In that can be incorporated into InGaN and AlInN alloys while maintaining high crystal quality is currently for layered structures about 30% in both cases [35, 36, 130–133]. In the case of QDs,

higher In contents of up to ∼35% have been reported in InGaN QDs [134, 135], whereas AlInN QDs have not been studied significantly in the literature. Regarding strain, devices with as much as ∼1.1% compressive in-plane strain in the barriers along the whole active layer have been reported [100]. For InGaN/GaN QWs and QDs, strain (lattice mismatch in the case of QD) of up to 3% and 3.8%, respectively, can be found in the literature [35, 100, 130, 134, 135]. Hence, the calculated strains in the top half of Table 3.4 for QW growth between unstrained barriers are all within the limits achieved experimentally. Turning to the structures grown on GaN in the bottom half of Table 3.4, built-in field reduction in these cases requires similar values for the strain in both the QW and the barrier. Therefore, the strain state of the barrier material is the one that sets a limit to the possible structures. In both halves of Table 3.4, the built-in field minimization procedure yields AlInN QDs as the optimal structures. This initially surprising result is due to the large band gap bowing of AlInN compared to InGaN. For high In contents, the band gap of AlInN is very similar to the band gap of InGaN, as shown in Ref. [21] and dealt with in detail in the next section. Therefore, in terms of built-in field reduction, AlInN QD systems could be more suitable than InGaN QDs. However, as previously mentioned, these AlInN systems have not been explored in detail and significant challenges remain for their experimental realization.

To realistically model the possible built-in field reduction in QW and QD systems, the optimization procedure of Sections 3.3.2 and 3.3.3 can be repeated to take all the constraints discussed above into account. The resultant structures, optimized in terms of built-in field reduction and maximization of band gap difference ΔE_g are shown in Table 3.5. The calculations have been made imposing a minimum value of $\Delta E_g = 0.25$ eV for QWs and $\Delta E_g = 0.30$ eV for QDs, because of the higher confinement energy expected in QDs compared to QWs, that would reduce the effective band gap offset. As can be seen from the data in Table 3.5, the imposition of such constraints sets a limit to the built-in field suppression, especially for those structures with low energy gaps (typically around and below 3 eV). However, it is always possible to achieve at least a partial field reduction in those structures, which has already been proven to have a significant impact on the efficiency of LED devices [100]. For the rest of the structures, a complete built-in field suppression is possible and, in line with the results reported in Ref. [100], an improved performance is expected by incorporating them into devices that span a wavelength range which extends well into the UV part of the spectrum.

Table 3.5: List of $Al_xIn_yGa_{1-x-y}N/Al_{x'}In_{y'}Ga_{1-x'-y'}N$ systems compatible with the constraints discussed in section 3.3.4: 30% maximum In content for QWs and QBs; 35% maximum In content in QDs; 3% maximum strain in QWs; 3.8% maximum strain (lattice mismatch) in QDs; 1.1% maximum strain in QBs; 0.25 eV minimum ΔE_g for QWs; and 0.3 eV minimum ΔE_g for QDs. Note that in the case of a QW system an energy gap of 2.25 and 2.5 eV, for unstrained barriers and pseudomorphic growth on GaN respectively, is not compatible with the given constraints.

	QW					QD				
E_g (eV)	$E_{\text{b-i}}$ (MV/cm)	ΔE_g (eV)	$(x,y)_{\text{opt}}/(x',y')_{\text{opt}}$	Lattice mismatch (%)		$E_{\text{b-i}}$ (MV/cm)	ΔE_g (eV)	$(x,y)_{\text{opt}}/(x',y')_{\text{opt}}$	Lattice mismatch (%)	
	Unstrained barriers									
2.25	n/a	n/a	n/a	n/a		-1.1	0.31	(0.00,0.35)/(0.01,0.23)	-1.3	
2.50	-1.4	0.26	(0.00,0.24)/(0.19,0.27)	-0.1		-1.0	0.30	(0.02,0.29)/(0.00,0.16)	-1.4	
2.75	-0.7	0.25	(0.00,0.15)/(0.33,0.28)	0.6		-1.0	0.30	(0.01,0.21)/(0.03,0.11)	-1.1	
3.00	0.0	0.27	(0.00,0.06)/(0.48,0.29)	1.4		-1.0	0.30	(0.03,0.13)/(0.23,0.15)	-0.3	
3.25	0.0	0.32	(0.00,0.00)/(0.49,0.23)	1.4		-0.8	0.30	(0.03,0.02)/(0.47,0.22)	1.2	
3.50	0.0	0.31	(0.11,0.00)/(0.53,0.20)	1.2		-0.8	0.30	(0.10,0.00)/(0.51,0.19)	1.1	
3.75	0.0	0.30	(0.23,0.00)/(0.64,0.20)	1.2		-0.8	0.30	(0.19,0.00)/(0.52,0.14)	0.8	
4.00	0.0	0.28	(0.33,0.00)/(0.70,0.18)	1.1		-0.8	0.30	(0.29,0.00)/(0.57,0.11)	0.6	
	Pseudomorphic growth on GaN									
				Well	Barrier				Dot	Barrier
2.50	n/a	n/a	n/a	n/a	n/a	-2.1	0.59	(0.02,0.33)/(0.01,0.10)	-3.5	-1.1
2.75	-2.2	0.35	(0.00,0.20)/(0.01,0.10)	-2.2	-1.1	-1.2	0.34	(0.00,0.22)/(0.01,0.10)	-2.4	-1.1
3.00	-1.2	0.25	(0.01,0.13)/(0.12,0.12)	-1.4	-1.0	-0.9	0.30	(0.03,0.15)/(0.16,0.13)	-1.6	-1.1
3.25	-0.4	0.27	(0.00,0.05)/(0.30,0.16)	-0.6	-1.1	-0.6	0.30	(0.01,0.05)/(0.33,0.17)	-0.5	-1.1
3.50	0.0	0.36	(0.04,0.00)/(0.52,0.21)	0.1	-1.1	-0.5	0.30	(0.05,0.00)/(0.47,0.20)	0.1	-1.1
3.75	0.0	0.36	(0.17,0.00)/(0.63,0.21)	0.4	-0.8	-0.6	0.31	(0.16,0.00)/(0.56,0.19)	0.4	-0.8
4.00	0.0	0.31	(0.30,0.00)/(0.68,0.18)	0.7	-0.4	-0.8	0.30	(0.29,0.00)/(0.64,0.16)	0.7	-0.2

3.3.5 UV devices: importance of the AlInN band gap bowing

Our calculations so far show how full polarization-matching and hence complete removal of the built-in field is only possible for a restricted range of energy gaps E_g, if the constraints set by the current epitaxial growth capabilities are taken into account. Particularly, the maximum amount of In that can be incorporated, as well as the maximum strain that barrier layers can withstand, are the main limiting factors. Fig. 3.10 (top) shows the minimum field that can be achieved by composition engineering as a function of the energy gap when the constraints discussed in Section 3.3.4 are taken into account, as well as the material composition needed for that purpose in both active layer and barrier (bottom panels). It is observed that for most of the UV (for $E_g > 3.35$ eV) there is always a combination of well and barrier materials that yields effective zero field, corresponding to AlGaN/AlInGaN. Note that, in terms of built-in field reduction, QDs appear as a promising solution for lower band gaps but do not represent an improvement with respect to QWs above 3.2 eV. Given the possibility of effective built-in field suppression for $E_g > 3.35$ eV, UV nitride devices benefit most from polarization-matching, subjected to the ability to grow the relevant materials. The bottom panel of Fig. 3.10 highlights the importance of the quaternary AlInGaN alloy for the purpose of built-in field control, particularly used as QB material. In contrast, the calculated optimum active layer materials are the conventional ternaries, InGaN for the low band gap end and AlGaN for higher band gaps. Results in Fig. 3.10, obtained using the parameters from Tables 3.1, 3.2 and 3.3, include the revised band gap bowing parameter for AlInN of $b_{\mathrm{AlInN}} = 5$ eV, given by Wu [21]. However, b_{AlInN} is surrounded by controversy as there is a wide range of values reported in the literature going from 2.5 to 10.3 eV [20, 21, 136]. The following analysis shows that the results are strongly influenced by the choice of b_{AlInN}.

Influence of the AlInN band gap bowing

As mentioned above, there is disagreement in the literature regarding the band gap bowing of AlInN. This is illustrated in Fig. 3.11. Our area of interest is the left hand side of the plot ($a < 3.25$ Å), corresponding to the UV part of the spectrum. The most striking characteristic is that, for a fixed lattice constant, AlGaN and AlInN are very similar alloys, from the band gap point of view, when the value of

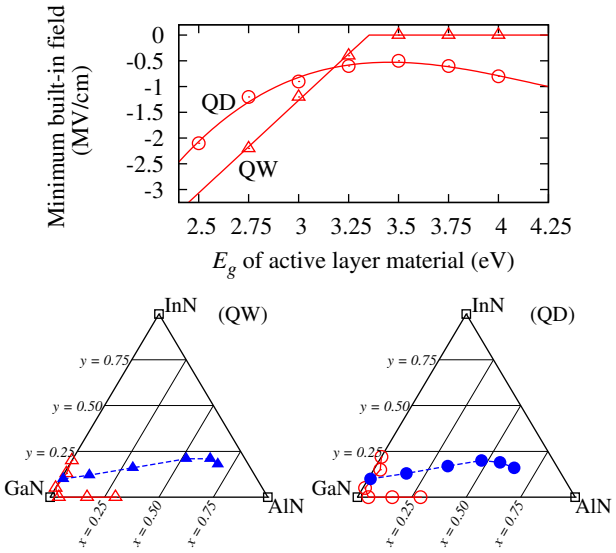


Figure 3.10: (top) Calculated lowest built-in field that can be achieved using partially or completely polarization-matched barriers in nitride QW and QD systems, obtained by composition engineering, and (bottom) the material compositions $Al_xIn_yGa_{1-x-y}N$ of active material (open symbols) and barrier (solid symbols) that correspond to those optimised systems. The band gap bowing parameter of AlInN is chosen to be $b_{\mathrm{AlInN}} = 5$ eV [21].

$b_{\mathrm{AlInN}} = 10.3$ eV reported in Ref. [136] is used.[2]

It is interesting how the change in these band gap bowing parameters affects the polarization-matching results. From Fig. 3.12 one can infer that employing in the calculation the formerly used (and currently widely discarded by the community) value of $b_{\mathrm{AlInN}} = 2.5$ eV given by Vurgaftman and Meyer [20] leads to an improvement of the built-in field values in the case of QDs and does not greatly affect the optimised material compositions, when compared to the results obtained with $b_{\mathrm{AlInN}} = 5$ eV (Fig. 3.10). However, when the value of $b_{\mathrm{AlInN}} = 10.3$ eV presented by Aschenbrenner *et al.* [136] is used in the calculation, great differences can be observed in both built-in field results and material compositions with respect to the previous cases (Fig. 3.13). In the case of the built-in field, effective suppression is found in both QW and QD cases for the whole studied UV range and perhaps part of the blue ($E_g \gtrsim 2.75$ eV). This is due to the replacement of both InGaN

[2]There are indications that the AlInN band gap bowing is composition-dependent, based on the data for the different compositional regimes in which the different bowing parameters are obtained. We have recently found theoretical confirmation of this using AlInN supercell calculations (unpublished). Our preliminary data suggests that Aschenbrenner's bowing parameter of $b_{\mathrm{AlInN}} = 10.3$ eV might indeed be valid for the low-In regime. For higher In content, the bowing is observed to go down rapidly with composition.

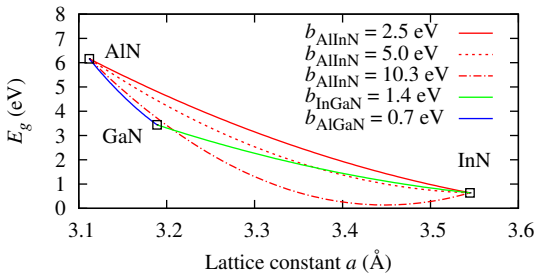


Figure 3.11: Band gap values of the nitride ternaries for the different bowing parameters available in the literature. It is worth noting that for the large $b_{\mathrm{AlInN}} = 10.3$ eV given in Ref. [136]) AlInN and AlGaN have similar band gaps at equal lattice constant, for energies $E_g > E_g(\mathrm{GaN})$.

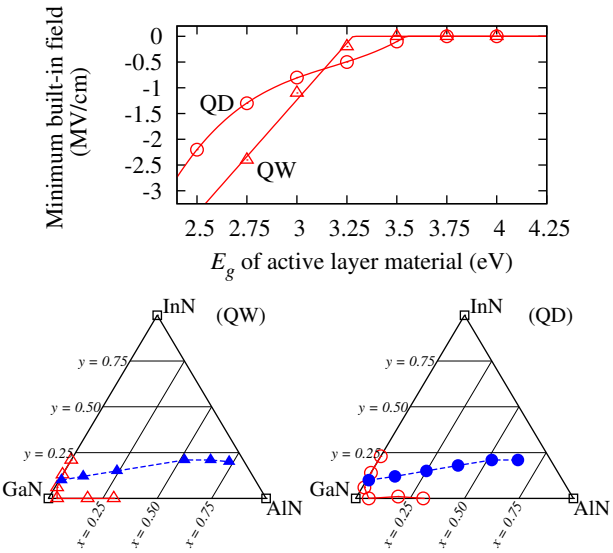


Figure 3.12: (Colour online) Same as Fig. 3.10 for $b_{\mathrm{AlInN}} = 2.5$ eV (Ref. [20]).

(except for very low E_g) and AlGaN by AlInN as active layer material. The reason for this is most clearly seen in Fig. 3.11, where AlInN for $b_{\mathrm{AlInN}} = 10.3$ eV is observed to span lattice constant/band gap areas that are otherwise only covered by InGaN and AlGaN when different bowing parameters are used. $b_{\mathrm{AlInN}} = 10.3$ eV was reported in Ref. [136] for samples grown between 13% and 24% In content and the authors state that the extrapolation of their result beyond this composition range is not supported by the experiment [136]. We set a maximum limit of 30% In content in our calculations, close enough to the 13–24% range for the extrapolation of $b_{\mathrm{AlInN}} = 10.3$ eV to hold.

Two particularities make AlInN extremely interesting for built-in field control purposes: the wide range of lattice constants that it spans and its large band gap

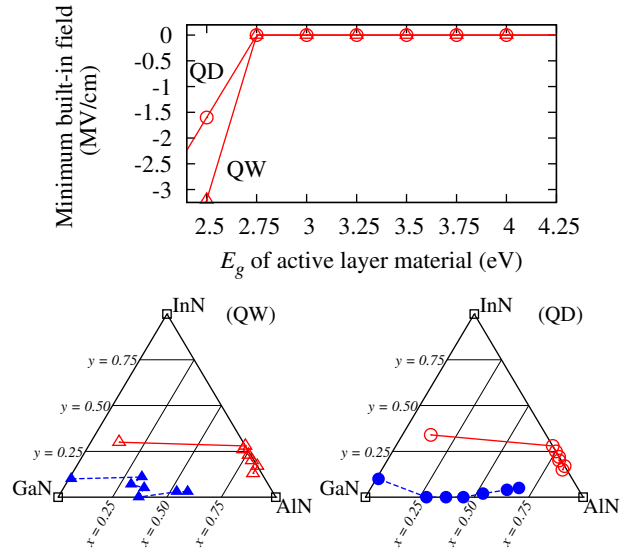


Figure 3.13: (Colour online) Same as Fig. 3.10 for $b_{\mathrm{AlInN}} = 10.3$ eV (Ref. [136]).

bowing, that makes it similar to InGaN and AlGaN for a certain range of compositions. Both properties combined make AlInN a versatile alloy and an excellent candidate for the improvement of devices incorporating nitride technology, via elimination of the strong built-in fields commonly present in these structures. The large uncertainty in its bowing parameter suggests that the energy gap of AlInN might be strongly dependent on the growth conditions, potentially providing a further degree of freedom to enable polarization-matching. The experimental investigation of AlInN alloys towards higher In incorporation and band gap tunability is therefore of great interest. On the other hand, the possible local fluctuation of the carrier confinement potential in AlInN [137] means that effective carrier localization might occur in barrier layers made of that material. Overall, more detailed studies of the properties of AlInN ternary alloys are required.

3.4 Summary

In summary, we have shown that it is in principle possible for a very wide range of emission energies to reduce or even completely suppress the high built-in fields intrinsic to wurtzite III-N heterostructures. This is achieved by taking advantage that the built-in field is of opposite sign across the GaN layer in InN/GaN/InN and in AlN/GaN/AlN structures. The built-in field can then be minimized and a significant energy gap difference ΔE_g can be maintained by a careful choice of

the alloy composition in the QW (QD) and barrier layers. We have presented the calculated QW (QD) and barrier systems to maintain zero field and maximize ΔE_g across a large range of QW (QD) energy gap values. Results were presented both for unstrained barriers and also for QW/barrier and QD/barrier structures grown pseudomorphically on a GaN substrate. Furthermore, we have discussed and considered in our calculations several constraints (miscibility issues, strain limitations) set by the epitaxial growth of III-N semiconductors. The built-in strain in each of the structures considered is also presented. Interpolation between the QW (QD) and barrier compositions can then be used to estimate the maximum offset for a given energy gap as a function of built-in strain in the QW (QD) and barrier layers. We have also taken a detailed look at structures operating in the UV part of the spectrum, where our analysis predicts that built-in field control could have the largest impact. For those structures, we have assessed the importance of the knowledge of how the AlInN band gap behaves with composition. We conclude that growth of suitable ternary and quaternary alloy combinations, in particular AlInN with high In content, has the potential to significantly reduce the built-in electric field in *c*-plane III-N heterostructures, of considerable benefit for a wide range of optoelectronic applications.

Overall, we have noted uncertainties in several critical material parameters, including in particular the strain and composition dependence of piezoelectric coefficients, and the composition-dependent value of the bowing parameter in AlInN. Further theoretical and experimental effort is required to reduce the uncertainty in these values. In some of the subsequent chapters, we therefore tackle the task of achieving a deeper understanding of the elastic and piezoelectric properties of nitride materials. In the case of nitride alloys, an accurate understanding of internal strain effects, in particular, is vital given the large lattice mismatch between nitride binaries that translates into large local distortions of the crystal lattice.

4 Elastic and structural properties of solids in the context of plane wave-based DFT

Abstract

We explore the calculation of the elastic and structural properties of zinc-blende and wurtzite semiconductors in the context of plane wave-based density functional theory. It is shown that the two different approaches commonly used, one based on stress and the other on total energy as a function of strain, do not in general yield the same results. Differences are found in particular in convergence rate between the two methods, with a much slower convergence for total energy calculations with respect to the number of plane waves and **k** points used. The stress method is observed to be more robust than the total energy method with respect to the residual error in the elastic constants calculated for different strain branches, a trend that has until now been surprisingly overlooked by the condensed-matter community. We explain these differences in terms of plane wave basis set inconsistency, for a fixed plane wave cutoff energy, as the lattice vectors change. Finally, the Heyd-Scuseria-Ernzerhof (HSE) screened-exchange hybrid functional approach is employed for an accurate calculation of the elastic constants and structural parameters of zinc-blende and wurtzite group-III nitrides. The calculations are carried out with the plane wave-based package VASP.

4.1 Introduction

The *ab initio* determination of elastic properties of crystalline materials is widely accepted across the material science community as a safe route to determine these quantities for compounds that are difficult to grow or for which the measurement of elastic constants is not a straight-forward procedure [19]. Elastic constants are of crucial importance, for instance, in understanding strain relaxation mechanisms in

nanostructures, such as quantum wells (QWs) or dots (QDs) [1]. There are many studies on the elastic properties of crystalline solids available in the literature that employ first-principles calculations, in particular plane wave approaches to density functional theory (DFT). Two different methods are most commonly used in order to obtain elastic constants of solids: based either on the calculation of stresses [57, 138, 139], or of total energy [140, 141], each calculated as a function of applied strain. However, there seems to be no conclusive comparative work (or general awareness on the issue, for that matter) on which of the two yields more reliable results. A brief mention to this problem is given at the end of section 3 of one of the early papers on the matter of *ab initio* determination of elastic properties of solids by Nielsen and Martin, published in 1985 [138]. Łopuszyński and Majewski [142] have also looked recently at this issue. They calculated the elastic constants of cubic nitrides, reporting differing values of the elastic constants obtained using "strain-energy" and "strain-stress" methods. These differences are significant ($\gtrsim 10\%$) for the elastic constant C_{12}, related to strains that distort the unit cell. However, they do not provide a discussion on whether one of the methods is to be preferred. In fact, most of the available papers tend to choose one or the other method without stating the justification for their particular choice, perhaps assuming the two of them are equivalent. Indeed, because the stress tensor is necessarily obtained as a derivative of the energy, there is no obvious reason to believe one method to be better than the other. However, things get more complicated when one takes a closer look at how the two different approaches are implemented in practice.

This chapter is largely based on our recent papers, Refs. [3, 4]. The first few sections will be devoted to the practicalities of calculating the elastic constants and structural parameters of solids with modern plane wave-based DFT packages. We argue that stress calculations, instead of total energy, should be preferred for the determination of the elastic properties of crystalline solids in the context of plane wave-based DFT. This preference is due to the improved robustness and consistency of the elastic coefficients obtained from the stresses compared to the total energy, leading to a significantly reduced computational cost to obtain converged results [4]. Then, in Section 4.4 we employ the methodology discussed in the previous sections for an accurate study of elastic and structural properties of zinc-blende (ZB) and wurtzite (WZ) group-III nitrides using the HSE hybrid functional approach [3, 62, 63]. In order to test convergence, very high cutoff energies

and large k-point samplings are used in some of the calculations in Section 4.3, that would lead to exceedingly large computational costs in the case of using hybrid functionals. Density functional theory within the local density approximation (LDA-DFT) on the other hand, offers a good compromise between accuracy and computational cost for the calculation of elastic and structural parameters. Therefore, the calculations in Section 4.3 have been performed employing LDA-DFT, rather than the more expensive HSE functional employed in Section 4.4. In both cases we use the projector augmented-wave (PAW) method [68,74] as implemented in VASP [72, 75].

4.2 Calculation of elastic constants and structural parameters

In this chapter we deal with two different kinds of parameters. The first type are those quantities that are given for a material in its equilibrium configuration, that is, equilibrium lattice parameters a_0 and c_0, and the internal parameter u_0 for WZ materials, and the lattice parameter a_0 for ZB materials. The second type consists of those parameters that come into play when strain is applied to the crystal, that is, elastic constants C_{ij} and the internal strain parameters ζ_i (for ZB, the one independent internal parameter ζ is known as the Kleinman parameter). The usual approach in an *ab initio* calculation for the first type of parameter is to undertake a minimization of the total crystal energy with respect to the lattice parameters and the internal degrees of freedom. In doing so, the total energy surface as a function of the lattice parameters (or equivalently, strain) has a local minimum at the equilibrium configuration point.[1] The calculation of the elastic constants can be done either via the stresses acting on the cell, or from the variation of the total energy when strain is applied [72,104,141,143]. In principle, both methods should give the same results for both kind of parameters (however, we shall see that this is not necessarily the case for realistic DFT calculations).

In the first part of this section, we introduce the expressions used in harmonic elasticity that relate energy and stress to strain. Then, we deal with internal strain in ZB and WZ lattices. Finally, we discuss how the plane wave representation of these quantities is carried out in the context of DFT, and what are the implications

[1] Furthermore, if the crystal system is the most stable one for a given material, e.g., WZ instead of ZB for GaN, the local minimum is also global.

of this for the calculation of elastic constants.

4.2.1 Harmonic elasticity

The lattice arrangement of any crystalline solid is determined by the energetics of a large number of interacting atoms. Stable or metastable configurations constitute the different polymorphs of a specific stoichiometric mix of chemical elements, and are given by an absolute or local minimum of the crystal energy as a function of the positions of the atoms. Consequently, any deformation of such materials implies that a particular amount of energy needs to be provided for the material to be brought from equilibrium to a higher energy state. Sufficiently close to the equilibrium, the energy can be well approximated by a quadratic function of all the independent parameters that characterize a crystal deformation, that is, the strains. The elastic constants are then the coefficients of proportionality between the energy density of the material and the macroscopic strain. This quadratic approximation is known as harmonic elasticity, and is valid for sufficiently small strains [144].

The elastic energy of a material with arbitrary crystal structure is given, in the context of harmonic elasticity [144], using Voigt notation,[2] by

$$E = \frac{V_0}{2} \sum_{i=1}^{6} \sum_{j=1}^{6} C_{ij} \epsilon_i \epsilon_j, \tag{4.1}$$

where V_0 is the volume of the sample (typically the unit cell volume), C_{ij} are the elastic constants, and ϵ_i are the components of the strain tensor. Following Eq. (4.1), a certain amount of energy needs to be employed to strain the crystal. In practice this corresponds to applying stress to the crystal. The elastic energy can thus also be expressed in terms of stresses acting on the cell. The components of the stress tensor σ_i are given by the first derivative of E with respect to the strain

$$\sigma_i = \frac{1}{V_0} \frac{\partial E}{\partial \epsilon_i} = \sum_{j=1}^{6} C_{ij} \epsilon_j, \tag{4.2}$$

[2]Note that, in Voigt notation, $\epsilon_1 = \epsilon_{xx}$, $\epsilon_2 = \epsilon_{yy}$, $\epsilon_3 = \epsilon_{zz}$, $\epsilon_4 = 2\epsilon_{yz}$, $\epsilon_5 = 2\epsilon_{xz}$ and $\epsilon_6 = 2\epsilon_{xy}$.

whereas the elastic constants are given by the second derivatives:

$$C_{ij} = \frac{1}{V_0}\frac{\partial^2 E}{\partial\epsilon_i \partial\epsilon_j}. \tag{4.3}$$

Combining equations (4.1) and (4.2), we can also write

$$E = \frac{V_0}{2}\sum_i \sigma_i \epsilon_i. \tag{4.4}$$

Therefore, in the context of harmonic elasticity, the elastic constants C_{ij} can be calculated directly from the variation of the total energy E of the system with strain, Eq. (4.1), or directly from the variation of the stresses with strain, Eq. (4.2). The total energy E can in turn be expressed as an explicit function of the stresses σ_i, Eq. (4.4). We shall see next that, given the way in which the calculations of E and σ are implemented in DFT codes, the quantities given by Eq. (4.1) and Eq. (4.4) differ in general. Consequently, the elastic constants C_{ij} calculated from each of them differ too. When talking about the energy "calculated from the stresses", noted E_σ from now on, we will therefore be referring to Eq. (4.4). The "total energy", on the other hand, will be the total DFT free energy E_{tot}, given by our choice of DFT code, VASP, as explained in Section 4.2.3.

Equation (4.2) implies that the stresses vanish when the applied strain is zero. When performing an optimization of the lattice parameters of the cell there is no shear distortion, i.e. $\sigma_4 = \sigma_5 = \sigma_6 = 0$, and imposing the restrictions due to the crystal symmetry of the ZB and WZ lattices, Eq. (4.2) reduces (for small strains) to

$$\begin{aligned} \sigma_i^{\text{ZB}}(a) &= \beta_i^{\text{ZB}}(a - a_0), \\ \sigma_i^{\text{WZ}}(a, c) &= \beta_i^{\text{WZ}}(a - a_0) + \gamma_i^{\text{WZ}}(c - c_0), \end{aligned} \tag{4.5}$$

with $(i = 1, 2, 3)$, where β_i and γ_i are constants that must be fitted to the *ab initio* data. The lattice constants a_0 and c_0 for which all σ_i vanish are the equilibrium lattice constants. A similar procedure can be used to obtain the elastic constants by directly evaluating Eq. (4.2) once the equilibrium lattice parameters and internal strain parameters are known. In the case of a total energy calculation, the quadratic relation between energy and strain given by Eq. (4.1) means that a second order polynomial needs to be used for the fitting.

We note that although the total energy E_{tot} of a crystal system may have a non-zero minimum at the equilibrium lattice configuration, the *variation* of this total energy with strain must satisfy Eq. (4.1).

It is not explicit in any of the equations presented so far how the atomic degrees of freedom are updated as a function of strain. This *internal strain* is responsible, however, for an important contribution to the total elastic response of crystals, and will be dealt with in detail in the next section.

4.2.2 Internal strain in zinc-blende and wurtzite lattices

An optimization of the internal degrees of freedom in a periodic structure is necessary when the unit cell contains more than one atom. In some cases, the existence of internal strain can be ruled out based on symmetry arguments, as is the case for any combination of axial strains in a ZB lattice, in the usual [001] representation. For some crystals with low symmetry, as is the case for the WZ lattice, even hydrostatic strain induces an internal deformation of the unit cell. The internal strain parameters describe the redistribution of the coordinates of the atoms contained within a unit cell as a function of the macroscopic strain. Although the form of these internal strain parameters can be obtained from symmetry considerations, one has to pay attention to the specific energetics of the crystal in order to calculate their numerical values. The internal strain parameter ζ in a ZB lattice is called the Kleinman parameter, and was originally studied by its namesake [145], and by Keating [146]. More recently, the form of the internal strain parameters of a WZ lattice ζ_i have been studied by Camacho and Niquet [147], and by ourselves [3, 4]. In their work, Camacho and Niquet obtained only four internal strain parameters for the WZ unit cell via a valence force field (VFF) approach [147]. However, we have obtained the correct number, which is five, using a more general VFF model in which we allow the three-body interactions about anions and cations to have different force constants [3, 4]. Although Camacho and Niquet assumed that only the average of those constants appears in the expression for the energy, this is only true in the case of a ZB lattice. In ZB, because of the periodicity of the lattice, all the tetrahedra are equivalent and the bond angles about the cations are equal to those about the anions. In the WZ lattice, there are two sets of inequivalent tetrahedra and the different force constants enter the general expression for the energy independently. Although without explicitly referring to internal strain, Nusimovici

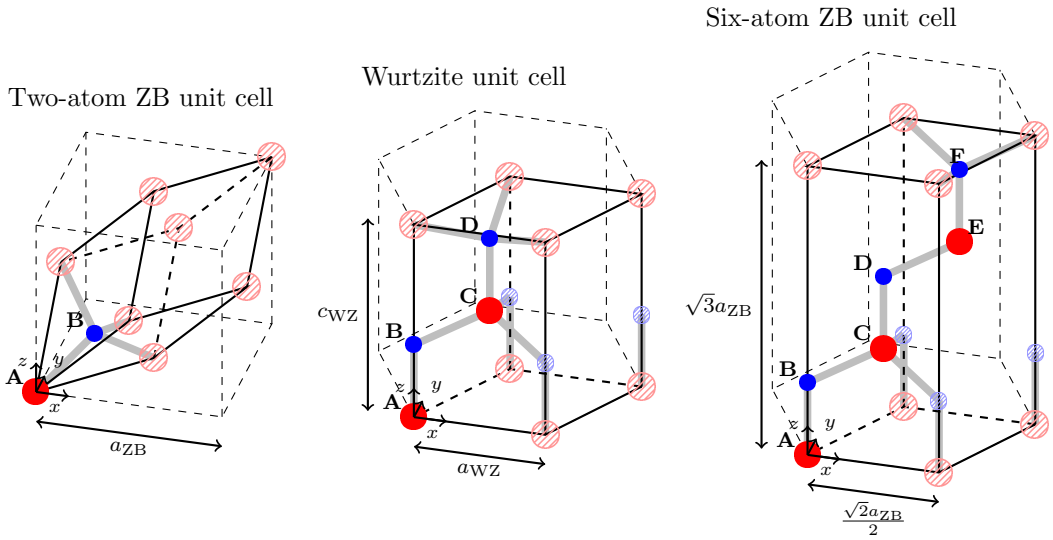


Figure 4.1: Different crystal systems used in this chapter. Big (red) circles indicate cations and small (blue) circles indicate anions. Solid circles are the atoms contained in the unit cell of the figure whereas shaded circles indicate atoms contained in neighbouring unit cells. Two bonds about atom D, and about atom E, that span across neighbouring unit cells, are omitted in the figure for the six-atom ZB cell.

and Birman [148] have previously reported different force constants for cation- and anion-centred three-body interactions for WZ CdS.

In Fig. 4.1, we show the spatial representation of the ZB (two- and six-atom unit cells) and WZ crystals that will be used throughout this chapter.[3] Before strain is applied, the positions of the atoms in the two-atom ZB unit cell are

$$\mathbf{r}^{\mathrm{ZB}}_{\mathrm{A},0} = [0,0,0]\,, \qquad \mathbf{r}^{\mathrm{ZB}}_{\mathrm{B},0} = \left[\frac{a_{\mathrm{ZB}}}{4}, \frac{a_{\mathrm{ZB}}}{4}, \frac{a_{\mathrm{ZB}}}{4}\right], \tag{4.6}$$

those of the WZ unit cell are

$$\begin{aligned} &\mathbf{r}^{\mathrm{WZ}}_{\mathrm{A},0} = [0,0,0]\,, \qquad \mathbf{r}^{\mathrm{WZ}}_{\mathrm{B},0} = [0,0,u_0 c_{\mathrm{WZ}}]\,, \\ &\mathbf{r}^{\mathrm{WZ}}_{\mathrm{C},0} = \left[\frac{a_{\mathrm{WZ}}}{2}, \frac{\sqrt{3}a_{\mathrm{WZ}}}{6}, \frac{c_{\mathrm{WZ}}}{2}\right], \qquad \mathbf{r}^{\mathrm{WZ}}_{\mathrm{D},0} = \left[\frac{a_{\mathrm{WZ}}}{2}, \frac{\sqrt{3}a_{\mathrm{WZ}}}{6}, \left(\frac{1}{2}+u_0\right)c_{\mathrm{WZ}}\right], \end{aligned} \tag{4.7}$$

[3]The [111]-oriented six-atom ZB unit cell will be employed in Section 4.3 to check for consistency of results when lowering the symmetry of the unit cell for the same material.

and the positions in the six-atom ZB (notated ZB_6) unit cell are

$$\begin{aligned}
\mathbf{r}_{\mathrm{A},0}^{\mathrm{ZB}_6} &= [0,0,0]\,, \qquad \mathbf{r}_{\mathrm{B},0}^{\mathrm{ZB}_6} = \left[0,0,\frac{\sqrt{3}a_{\mathrm{ZB}}}{4}\right], \\
\mathbf{r}_{\mathrm{C},0}^{\mathrm{ZB}_6} &= \left[\frac{\sqrt{2}a_{\mathrm{ZB}}}{4},\frac{\sqrt{6}a_{\mathrm{ZB}}}{12},\frac{\sqrt{3}a_{\mathrm{ZB}}}{3}\right], \qquad \mathbf{r}_{\mathrm{D},0}^{\mathrm{ZB}_6} = \left[\frac{\sqrt{2}a_{\mathrm{ZB}}}{4},\frac{\sqrt{6}a_{\mathrm{ZB}}}{12},\frac{7\sqrt{3}a_{\mathrm{ZB}}}{12}\right], \\
\mathbf{r}_{\mathrm{E},0}^{\mathrm{ZB}_6} &= \left[\frac{\sqrt{2}a_{\mathrm{ZB}}}{2},\frac{\sqrt{6}a_{\mathrm{ZB}}}{6},\frac{2\sqrt{3}a_{\mathrm{ZB}}}{3}\right], \qquad \mathbf{r}_{\mathrm{F},0}^{\mathrm{ZB}_6} = \left[\frac{\sqrt{2}a_{\mathrm{ZB}}}{2},\frac{\sqrt{6}a_{\mathrm{ZB}}}{6},\frac{11\sqrt{3}a_{\mathrm{ZB}}}{12}\right],
\end{aligned} \tag{4.8}$$

where a_{ZB} is the lattice constant of the ZB lattice, a_{WZ} and c_{WZ} are the lattice constants of the WZ lattice, and u_0 is the internal parameter of the latter ($u_0 = \frac{3}{8}$ for an ideal WZ lattice). After applying an arbitrary deformation characterized by the strain tensor ϵ:

$$\epsilon = \begin{pmatrix} \epsilon_{xx} & \epsilon_{xy} & \epsilon_{xz} \\ \epsilon_{xy} & \epsilon_{yy} & \epsilon_{yz} \\ \epsilon_{xz} & \epsilon_{yz} & \epsilon_{zz} \end{pmatrix}, \tag{4.9}$$

the positions for the two-atom ZB unit cell become

$$\mathbf{r}_{\mathrm{A}}^{\mathrm{ZB}} = [0,0,0]\,, \qquad \mathbf{r}_{\mathrm{B}}^{\mathrm{ZB}} = (\mathbb{1}+\epsilon)\,\mathbf{r}_{\mathrm{B},0}^{\mathrm{ZB}} + \mathbf{t}^{\mathrm{ZB}}, \tag{4.10}$$

for the WZ unit cell they become

$$\begin{aligned}
\mathbf{r}_{\mathrm{A}}^{\mathrm{WZ}} &= [0,0,0]\,, \qquad \mathbf{r}_{\mathrm{B}}^{\mathrm{WZ}} = (\mathbb{1}+\epsilon)\,\mathbf{r}_{\mathrm{B},0}^{\mathrm{WZ}} + \mathbf{t}_1^{\mathrm{WZ}}, \\
\mathbf{r}_{\mathrm{C}}^{\mathrm{WZ}} &= (\mathbb{1}+\epsilon)\,\mathbf{r}_{\mathrm{C},0}^{\mathrm{WZ}} + \mathbf{t}_2^{\mathrm{WZ}}, \qquad \mathbf{r}_{\mathrm{D}}^{\mathrm{WZ}} = (\mathbb{1}+\epsilon)\,\mathbf{r}_{\mathrm{D},0}^{\mathrm{WZ}} + \mathbf{t}_3^{\mathrm{WZ}},
\end{aligned} \tag{4.11}$$

and for the six-atom ZB unit cell the positions become

$$\begin{aligned}
\mathbf{r}_{\mathrm{A}}^{\mathrm{ZB}_6} &= [0,0,0]\,, \qquad \mathbf{r}_{\mathrm{B}}^{\mathrm{ZB}_6} = (\mathbb{1}+\epsilon)\,\mathbf{r}_{\mathrm{B},0}^{\mathrm{ZB}_6} + \hat{U}\mathbf{t}^{\mathrm{ZB}}, \\
\mathbf{r}_{\mathrm{C}}^{\mathrm{ZB}_6} &= (\mathbb{1}+\epsilon)\,\mathbf{r}_{\mathrm{C},0}^{\mathrm{ZB}_6}, \qquad \mathbf{r}_{\mathrm{D}}^{\mathrm{ZB}_6} = (\mathbb{1}+\epsilon)\,\mathbf{r}_{\mathrm{D},0}^{\mathrm{ZB}_6} + \hat{U}\mathbf{t}^{\mathrm{ZB}}, \\
\mathbf{r}_{\mathrm{E}}^{\mathrm{ZB}_6} &= (\mathbb{1}+\epsilon)\,\mathbf{r}_{\mathrm{E},0}^{\mathrm{ZB}_6}, \qquad \mathbf{r}_{\mathrm{F}}^{\mathrm{ZB}_6} = (\mathbb{1}+\epsilon)\,\mathbf{r}_{\mathrm{F},0}^{\mathrm{ZB}_6} + \hat{U}\mathbf{t}^{\mathrm{ZB}},
\end{aligned} \tag{4.12}$$

where $\mathbb{1}$ is the 3×3 identity matrix and $\hat{U}$ is the unitary transformation matrix

we have used to rotate the ZB system to a [111] orientation, given by

$$\hat{U} = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix}. \tag{4.13}$$

The different $\mathbf{t}$ are the internal strain vectors, given, for the ZB cell by [146]

$$\mathbf{t}^{\mathrm{ZB}} = \left[-\frac{a_{\mathrm{ZB}}}{2}\zeta\epsilon_{yz}, -\frac{a_{\mathrm{ZB}}}{2}\zeta\epsilon_{xz}, -\frac{a_{\mathrm{ZB}}}{2}\zeta\epsilon_{xy} \right], \tag{4.14}$$

and for the WZ cell by [3]

$$\begin{aligned} \mathbf{t}_1^{\mathrm{WZ}} &= c_{\mathrm{WZ}} \left[\zeta_1\epsilon_{xz} + \zeta_5\epsilon_{xy}, \zeta_1\epsilon_{yz} + \zeta_5\frac{(\epsilon_{xx} - \epsilon_{yy})}{2}, \zeta_2\left(\epsilon_{xx} + \epsilon_{yy}\right) - \zeta_3\epsilon_{zz} \right], \\ \mathbf{t}_2^{\mathrm{WZ}} &= a_{\mathrm{WZ}} \left[-\zeta_4\epsilon_{xy}, -\frac{(\epsilon_{xx} - \epsilon_{yy})}{2}\zeta_4, 0 \right], \\ \mathbf{t}_3^{\mathrm{WZ}} &= \mathbf{t}_1^{\mathrm{WZ}} + \mathbf{t}_2^{\mathrm{WZ}} - 2c_{\mathrm{WZ}} \left[\zeta_5\epsilon_{xy}, \zeta_5\frac{(\epsilon_{xx} - \epsilon_{yy})}{2}, 0 \right]. \end{aligned} \tag{4.15}$$

We have labelled the internal parameters of the WZ lattice to be consistent with Camacho and Niquet's original notation [147]: ζ_1, ζ_2, ζ_3 and ζ_4, including the missing parameter as ζ_5. About the equilibrium, the lattice energy exhibits a quadratic dependence on each of the internal strain parameters. The values of the strain parameters can therefore be obtained by applying a series of compatible macroscopic strains and then fitting the dependence of the total energy on the internal strain parameters to a parabola for each macroscopic strain used [3].

The calculations performed here, for which the internal degrees of freedom will be optimized at each strain configuration, are equivalent to the traditional *clamped-ion* method used to calculate elastic constants with an internal strain component [138]. In the clamped-ion scheme, there are two contributions to the elastic constants. The first contribution, the clamped-ion part, calculates the crystal energy when strain is applied to the lattice but the internal strain is fixed to zero; in other words, when all the $\mathbf{t}$ are set to zero in Equations (4.10), (4.11) and (4.12). The second then corrects for this over-estimation of the energy by subtracting from the total clamped-ion energy value the energy reduction due to the optimization of the internal coordinates. The energy density E/V_0 for an arbitrary lattice can be written, in the harmonic limit, as a quadratic explicit

function of macroscopic and internal strain:

$$\frac{E}{V_0} = \frac{1}{2}\sum_{i,j} C_{ij}\epsilon_i\epsilon_j + \frac{1}{2}\sum_{k,l} \alpha_{k,l}\left[t_k - t_{k,0}\left(\boldsymbol{\epsilon}\right)\right]\left[t_l - t_{l,0}\left(\boldsymbol{\epsilon}\right)\right]. \tag{4.16}$$

The first term in Eq. (4.16) is the same as Eq. (4.1), while the second one is a sum over all independent internal strains. t_k are the internal strains before optimization, whereas $t_{k,0}\left(\boldsymbol{\epsilon}\right)$ are the internal strains that minimize the total energy for some given strain configuration $\boldsymbol{\epsilon}$. The $\alpha_{k,l}$ can be expressed as $\frac{\partial^2(E/V_0)}{\partial t_k \partial t_l}$ and are constant in the harmonic approximation. Although, as previously seen, the number of independent internal strains depends on the particular symmetry of the lattice, Eq. (4.16) is general. In harmonic elasticity theory, we can assume the $t_{k,0}\left(\boldsymbol{\epsilon}\right)$ to be linear in ϵ_i. Therefore they can be expressed as

$$t_{k,0}\left(\boldsymbol{\epsilon}\right) = \sum_i \frac{\partial t_{k,0}}{\partial \epsilon_i}\epsilon_i, \tag{4.17}$$

where the partial derivatives are constant. Taking the second derivative of Eq. (4.16), and using Eq. (4.17) it is then possible to derive that

$$C_{ij} = \left.\frac{\partial^2(E/V_0)}{\partial\epsilon_i\partial\epsilon_j}\right|_{t_{k,l}=0} - \sum_{k,l}\frac{\partial^2(E/V_0)}{\partial t_k \partial t_l}\frac{\partial t_{k,0}}{\partial\epsilon_i}\frac{\partial t_{l,0}}{\partial\epsilon_j}, \tag{4.18}$$

where $C_{ij}^{(0)} = \left.\frac{\partial^2(E/V_0)}{\partial\epsilon_i\partial\epsilon_j}\right|_{t_{k,l}=0}$ gives the clamped-ion contribution to the elastic constant C_{ij}.

In contrast to Eq. (4.18), our calculations will use the derivative of the energy, either directly evaluated as a finite difference ["total energy method", Eq. (4.19)] or as given by the Hellmann-Feynman theorem ["stress method", Eq. (4.20)], where E is already the "true" total energy for a particular strain, in the sense that it has been minimized with respect to t_k and t_l, that is, $t_k = t_{k,0}$ and $t_l = t_{l,0}$ in Eq. (4.16). Performing different tests, we have found both methods to be in excellent agreement. This agreement is to be expected since the internal optimization is equivalent to the subtraction of the second term in Eq. (4.18). Although the two methods are equivalent, the approach which we will take seems to be more straightforward to implement, especially in the case of a large number of independent internal strains, such as for the WZ lattice, for which implementation of

Eq. (4.18) becomes tedious.

4.2.3 DFT implementation of total energy and stress tensor

VASP [72] is one of many available packages based on plane waves.[4] In these codes, the calculation of the total DFT energy E_{tot} is carried out in reciprocal space. The reciprocal space formulation of E_{tot} in the context of DFT was first introduced by Ihm, Zunger and Cohen [149] and later used by Nielsen and Martin [138, 150] to derive the stress tensor $\sigma_{\alpha\beta}$ using the Hellmann-Feynman theorem. Following Ref. [138], the energy density $\frac{E_{\text{tot}}}{V_0}$ is given by

$$\begin{aligned}\frac{E_{\text{tot}}}{V_0} = &\sum_{\mathbf{k},\mathbf{G},i} |\Psi_i(\mathbf{k}+\mathbf{G})|^2 \frac{\hbar^2}{2m}(\mathbf{k}+\mathbf{G})^2 + \frac{1}{2}4\pi e^2 {\sum_{\mathbf{G}}}' \frac{|\rho(\mathbf{G})|^2}{\mathbf{G}^2} + \sum_{\mathbf{G}} \epsilon_{\text{xc}}(\mathbf{G})\rho^*(\mathbf{G}) \\ &+ {\sum_{\mathbf{G},\tau}}' S_\tau(\mathbf{G}) V_\tau^{\text{L}}(G)\rho^*(\mathbf{G}) + \sum_{\mathbf{k},\mathbf{G},\mathbf{G}',i,l,\tau} S_\tau(\mathbf{G}-\mathbf{G}')\Delta V_{l,\tau}^{\text{NL}}(\mathbf{k}+\mathbf{G},\mathbf{k}+\mathbf{G}') \\ &\times \Psi_i(\mathbf{k}+\mathbf{G})\Psi_i^*(\mathbf{k}+\mathbf{G}') + \left(\sum_\tau \alpha_\tau\right)\left(\frac{1}{V_0}\sum_\tau Z_\tau\right) + \frac{\gamma_{\text{Ewald}}}{V_0}, \end{aligned} \tag{4.19}$$

and the stress tensor $\sigma_{\alpha\beta}$, expressed in Cartesian notation, by

$$\begin{aligned}\sigma_{\alpha\beta} = &\frac{1}{V_0}\frac{\partial E_{\text{tot}}}{\partial \epsilon_{\alpha\beta}} = \frac{\hbar^2}{m}\sum_{\mathbf{k},\mathbf{G},i} |\Psi_i(\mathbf{k}+\mathbf{G})|^2 (\mathbf{k}+\mathbf{G})_\alpha (\mathbf{k}+\mathbf{G})_\beta + \frac{1}{2}4\pi e^2 {\sum_{\mathbf{G}}}' \frac{|\rho(\mathbf{G})|^2}{\mathbf{G}^2} \\ &\times \left(\frac{2G_\alpha G_\beta}{\mathbf{G}^2} - \delta_{\alpha\beta}\right) + \delta_{\alpha\beta}\sum_{\mathbf{G}} \left\{\epsilon_{\text{xc}}(\mathbf{G}) - \mu_{\text{xc}}(\mathbf{G})\right\}\rho^*(\mathbf{G}) \\ &- {\sum_{\mathbf{G},\tau}}' S_\tau(\mathbf{G})\left(\frac{\partial V_\tau^{\text{L}}(G)}{\partial(G^2)} 2G_\alpha G_\beta + V_\tau^{\text{L}}(G)\delta_{\alpha\beta}\right)\rho^*(\mathbf{G}) \\ &+ \sum_{\mathbf{k},\mathbf{G},\mathbf{G}',i,l,\tau} S_\tau(\mathbf{G}-\mathbf{G}')\frac{\partial \Delta V_{l,\tau}^{\text{NL}}(\mathbf{k}+\mathbf{G},\mathbf{k}+\mathbf{G}')}{\partial \epsilon_{\alpha\beta}}\Psi_i(\mathbf{k}+\mathbf{G})\Psi_i^*(\mathbf{k}+\mathbf{G}') \\ &- \delta_{\alpha\beta}\left(\sum_\tau \alpha_\tau\right)\left(\frac{1}{V_0}\sum_\tau Z_\tau\right) + \frac{1}{V_0}\frac{\partial \gamma_{\text{Ewald}}}{\partial \epsilon_{\alpha\beta}}. \end{aligned} \tag{4.20}$$

In Eqs. (4.19) and (4.20), V_0 is the unit cell volume, Ψ_i is the wave function of each of the occupied states denoted by i, $\mathbf{k}$ represents all vectors in reciprocal space spanning the first Brillouin zone, $\mathbf{G}$ is a reciprocal lattice vector, ρ is the charge density, ϵ_{xc} is the exchange-correlation energy density, S_τ is the structure factor, where τ labels the atoms within the unit cell. V^{L} is a local (l-independent)

[4] Other popular choices are ABINIT, Quantum ESPRESSO, or CASTEP, to name a few.

potential, and ΔV^{NL} is a non-local (l-dependent) correction to the latter, where l is the quantum number for the angular momentum. α_τ is a constant term in V_τ^{L}, Z_τ is the number of valence electrons of atom τ, and γ_{Ewald} is the ion-ion electrostatic interaction compensated by a negative background that makes the average electrostatic potential zero. $\sum'$ indicates a sum excluding $\mathbf{G} = 0$ and μ_{xc} is the exchange-correlation potential. Further information can be found in Refs. [138] and [149]. The most relevant remarks for our discussion are, first, that for a particular set of lattice vectors, $\sigma_{\alpha\beta}$ can be obtained from E_{tot} *without changing the basis set*, that is, the summations run over the *same* $\mathbf{k}$ and $\mathbf{G}$ in both equations (4.19) and (4.20). Most available plane wave DFT codes, including VASP, implement the above equations, which means that the calculation of the stress tensor is done implicitly with a *fixed basis set*. Second, the contribution of the internal degrees of freedom to both energy and stress can be explicitly taken into account in both cases with the summation over τ. Third and last, applying strain or, equivalently, changing the lattice vectors, will have the effect of changing the basis set used in the calculation since the reciprocal lattice vectors $\mathbf{G}$ vary with strain.[5] It can be seen therefore that, using the finite difference method to calculate the stress σ_i from Eq. (4.2) as the derivative of the energy density given by Eq. (4.19), involves different basis sets for the different strain values used in the finite difference calculations. It can be expected that this may change the calculated stress value, and that it may differ from the direct output of Eq. (4.20). In the remainder of this chapter we show that this is indeed the case, unless convergence has been achieved with respect to the number of plane waves and $\mathbf{k}$ points.

4.3 Effects of basis set incompleteness and *k*-point sampling on results

As stated in Section 4.2.3, the stress tensor can in principle be numerically approximated as a finite difference for a series of calculations of E_{tot}/V_0 for different sets of lattice vectors, i.e. as a function of strain. If an infinite accuracy was available for such a calculation, it would yield results in perfect agreement with the stress tensor calculated directly from Eq. (4.20), for sufficiently small steps.

[5] Furthermore, an automatic k-point generation scheme will likely have the effect of also changing the $\mathbf{k}$ points used for the Brillouin zone sampling.

Unfortunately, the precision that can be obtained with VASP, or other plane wave codes, for both total energy and stress tensor depends (among other things) on the ability to achieve a particular degree of convergence with respect to the number of plane waves and the number of **k** points used in the calculation. In this section we present the results of our comparison between the energy-strain and stress-strain methods for the calculation of the elastic constants of our test case material AlN, and show how the convergence of the energy-strain affects the elastic constants values. The study is carried out for AlN with both zinc-blende (ZB) and wurtzite (WZ) crystal structures. In Section 4.3.1 we argue why AlN is a suitable material system to perform these tests. The impact of the plane wave basis set used in the calculations on the results is studied in Section 4.3.2. Section 4.3.3 deals with the behaviour of the results as the number of **k** points included in the calculation changes and as the cutoff energy is varied, where the cutoff energy determines the maximum **G** values used in the calculation. In Section 4.3.4 we analyse the effect that lowering the crystal symmetry has on the consistency of the elastic constants derived along different strain branches, i.e. for the different ways in which the crystal is deformed. Finally, for the sake of generality, we also show results for selected further materials, namely C, Si, GaAs and GaN, in Section 4.3.5.

To ensure that we are working within the harmonic limit, we verify that all the calculated energies (both free energies and stress energies) as a function of strain lie on a parabola (or parabolic surface, for several independent strains). Likewise, we also verify that all the calculated stresses as a function of strain lie on a straight line (or in a plane). This is required by Equations (4.1), (4.2) and (4.4), which are only valid in the context of harmonic elasticity, close enough to the equilibrium.

4.3.1 AlN as a test case material

The tests performed in this work include integration over very fine k meshes and the use of basis sets with very high cutoff energies. With the many different configurations studied, this translates into a very expensive computational cost. Popular semiconductors in the III-V category often include Ga- or In-containing compounds such as GaAs, InAs or GaN. Ga and In have shallow core ("semicore") d states that usually need to be explicitly included in accurate calculations, adding significantly to the overall computational cost. In contrast, the reduced number of shells present in the Al electronic configuration means that valence d orbitals can

be omitted from the calculation. Therefore, calculations for AlN run faster than for any other member of the III-V family, except perhaps for BN.

AlN is most stable in the WZ crystal structure, although attempts to grow it in the metastable ZB structure have also been successful [151]. Therefore, this material can be safely modelled for both structures without raising any issues related to phase instability. Given the crystal symmetries, Eq (4.1) can be simplified for ZB and WZ compounds as [144]

$$E_{\mathrm{ZB}} = V_0\Big\{\frac{1}{2}C_{11}\left({\epsilon_1}^2+{\epsilon_2}^2+{\epsilon_3}^2\right)+C_{12}\left(\epsilon_1\epsilon_2+\epsilon_1\epsilon_3+\epsilon_2\epsilon_3\right) + \frac{1}{2}C_{44}\left({\epsilon_4}^2+{\epsilon_5}^2+{\epsilon_6}^2\right)\Big\}, \tag{4.21}$$

and

$$E_{\mathrm{WZ}} = V_0\Big\{\frac{1}{2}C_{11}\left({\epsilon_1}^2+{\epsilon_2}^2\right)+\frac{1}{2}C_{33}{\epsilon_3}^2+C_{12}\epsilon_1\epsilon_2+C_{13}\left(\epsilon_1\epsilon_3+\epsilon_2\epsilon_3\right) + \frac{1}{2}C_{44}\left({\epsilon_4}^2+{\epsilon_5}^2\right)+\frac{1}{2}C_{66}{\epsilon_6}^2\Big\}, \tag{4.22}$$

respectively, where $C_{66}=\frac{C_{11}-C_{12}}{2}$.

We expect the results presented here for AlN to be also applicable, to different extents, to other semiconductor compounds, and possibly any crystalline material modelled via a plane wave formalism. Indeed, we have noted larger differences in the elastic constants calculated from total energy and stress for GaN and InN than for AlN. We provide selected calculations in Section 4.3.5 for C, Si, GaAs, and GaN, to further support that our conclusions should be generally valid.

4.3.2 Effect of the plane wave basis set on the calculation

For a certain cutoff energy E_{cut}, only the plane waves that obey the condition $|\mathbf{G}+\mathbf{k}|<G_{\mathrm{cut}}$ are included in the calculation, with

$$E_{\mathrm{cut}}=\frac{\hbar^2}{2m}{G_{\mathrm{cut}}}^2. \tag{4.23}$$

Thus, the number of plane waves depends on which $\mathbf{k}$ point is being considered and what the reciprocal lattice vectors look like. This means that in different calculations of E_{tot} with different sets of lattice vectors, the basis set changes *individually* for each $\mathbf{k}$ point used in the integration over the Brillouin zone. This would not

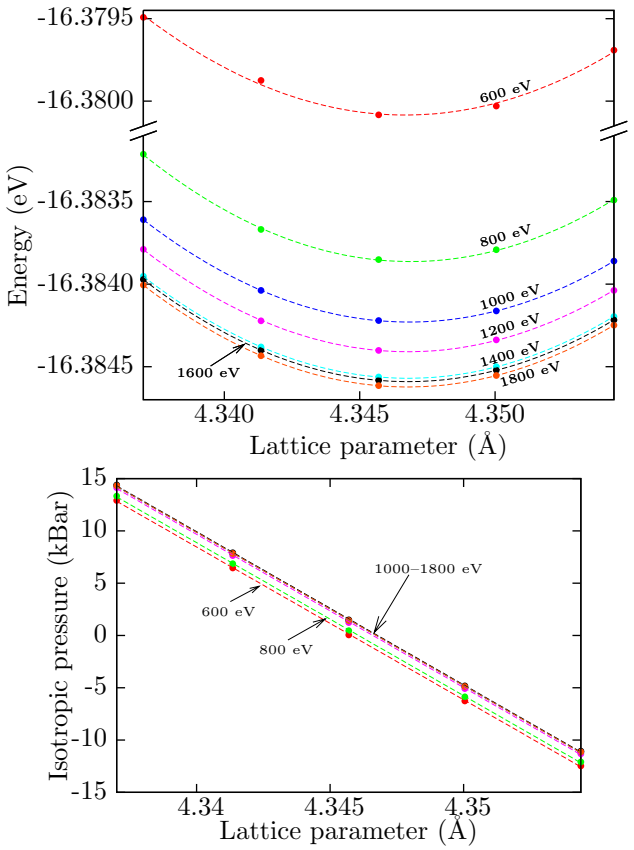


Figure 4.2: (top) Energy as a function of lattice parameter (or hydrostatic strain) for ZB AlN, with different cutoff energies for the plane wave basis set. k-mesh is $6\times6\times6$ Γ-centred. Results for PAW-LDA approximation; dashed lines are second order fit to data. (bottom) Isotropic stress as a function of lattice parameter for the same configuration as the total energy calculations above; dashed lines are linear fit to data.

be an issue if *absolute convergence* for the total energy could easily be achieved with respect to the cutoff, which is certainly not the case. Absolute convergence for E_{tot} is achieved *extremely* slowly with increasing cutoff, as can be observed in Fig. 4.2 (top) for ZB AlN. It should be noted that, within the PAW formalism, 600 eV is already considered to be a very high cutoff energy [72]. By contrast, convergence of the stress tensor is achieved much faster with increasing cutoff, as shown in Fig. 4.2 (bottom). In addition to this, the fact that each calculation is performed with a different basis set implies that there is an uncertainty regarding the value of the finite shift that occurs from one calculation to the next. This is true even for the same cutoff energy, that is, for the different data points along each of the curves in Fig. 4.2 (top). By contrast, there is no ambiguity regarding the minimum for the equilibrium configuration when calculating the stresses (that

Table 4.1: Elastic coefficient $C_{11} + 2C_{12}$, as a function of energy cutoff, obtained for ZB AlN from hydrostatic deformation using two different approaches: total energy and stress.

Cutoff (eV)	Total energy	Stress
	$C_{11} + 2C_{12}$ (GPa)	
600	633.8	634.8
800	653.2	635.7
1000	632.8	636.2
1200	635.2	636.2
1400	635.7	636.3
1600	635.7	636.3
1800	635.3	636.3

is, the minimum must always be at $\sigma = 0$ once convergence has been achieved with respect to the basis set). This ensures that calculations of the stress tensor done at different lattice vectors are consistent with each other, given that the basis set is sufficiently near completeness.

From Eq. (4.21), it can be seen that fitting to the data in Fig. 4.2 (hydrostatic strain, $\epsilon_1 = \epsilon_2 = \epsilon_3$) allows to determine $C_{11} + 2C_{12}$. The values calculated are shown in Table 4.1. The data in the table provides a first indication that stress calculations are cheaper (they converge faster) and more robust (they vary less for changing cutoff) than total energy calculations when deriving the elastic constants of crystalline solids.

It must be remarked that Fig. 4.2 shows results for a structure for which an optimization of the internal degrees of freedom is not necessary. Although for ZB AlN under *homogeneous* shear deformation the internal degrees of freedom need to be optimized, the crystal remains highly symmetric. Figure 4.3 shows the case of $\epsilon_4 = \epsilon_5 = \epsilon_6$, which gives C_{44}, and where the internal strain needs also to be determined (see Section 4.2.2). The off-diagonal components of the stress tensor converge even faster than the diagonal ones did in the hydrostatic strain case, because the small indeterminacy of the equilibrium lattice parameter has a direct impact on σ_1, σ_2 and σ_3 but not on σ_4, σ_5 and σ_6. The calculations shown in Fig. 4.3 yield $C_{44} = 188.7 \pm 0.3$ GPa for the total energy method and $C_{44} = 187.8 \pm 0.0$ GPa for the stress method.

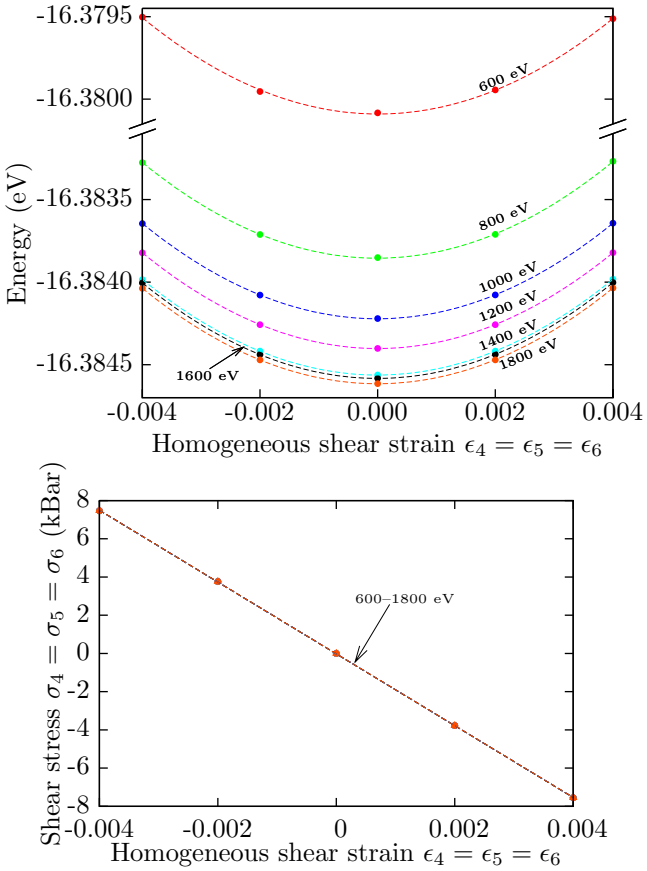


Figure 4.3: Same as Fig. 4.2 in the case of shear strain instead of hydrostatic strain.

4.3.3 Effect of the number of k points on the calculation

As we have seen in the previous sections, the integration in k space has to be done following a discretisation of the Brillouin zone. Therefore, as the number of **k** points included in the calculation increases, the quality of the numerical approximation to the integral will improve accordingly. Figure 4.4 (a) shows the convergence with the number of **k** points of the elastic properties of ZB AlN under hydrostatic strain. We have fixed the cutoff energy to 1000 eV (a1) and 800 eV (a2). It can be observed that, unlike in the previous section for the cutoff energy E_{cut}, both the total energy E_{tot} and stress energy E_σ methods do converge to the *same* value by increasing the number of **k** points. However, the cost of convergence in the case of total energy is much higher than the cost of a stress calculation, although increasing the cutoff energy enhances convergence in both cases. Indeed, for the particular case of 800 eV (which is already regarded as a high cutoff energy [72]), a calculation of $C_{11} + 2C_{12}$ with total energy with a $24 \times 24 \times 24$ k-mesh is of similar

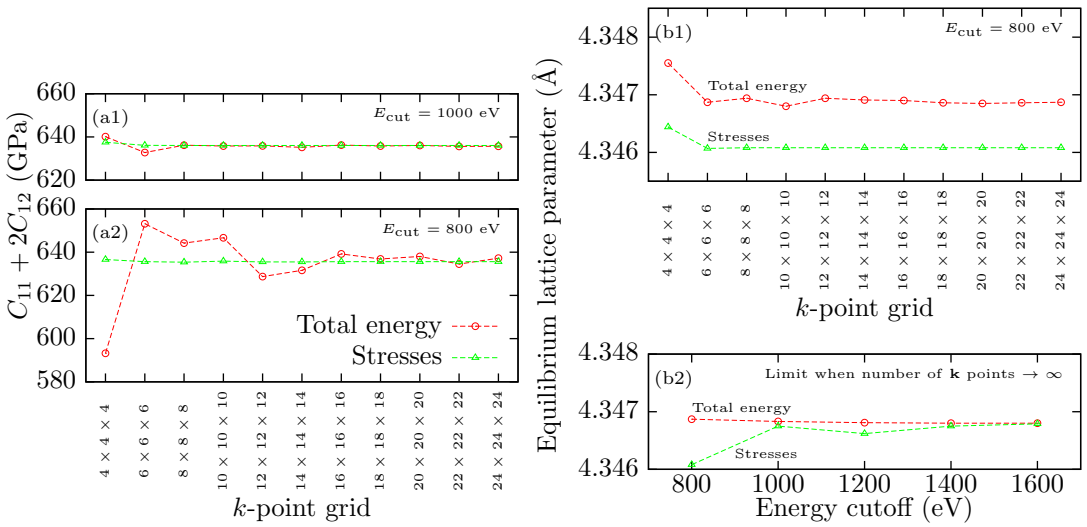


Figure 4.4: (left) Elastic coefficient $C_{11} + 2C_{12}$ calculated for ZB AlN as a function of the k-mesh used to sample the k space for total energy (open circles) and stress (open triangles) methods, for cutoff energies of 1000 (a1) and 800 eV (b2). (right) Equilibrium lattice parameter as obtained from total energy (open circles) and stress (open triangles) methods, (b1) for a fixed energy cutoff of 800 eV as the number of **k** points increases and (b2) asymptotic behaviour for very large number of **k** points and for varying cutoff energies. The k-grids are Γ-centred for all the calculations.

quality to a stress calculation of the same quantity with a $6 \times 6 \times 6$ mesh, which is almost converged to the correct value. For cubic semiconductors and insulators a $6 \times 6 \times 6$ k-mesh would be a typical choice for accurate calculations [72]. This is a clear indication that the stress method is, by far, the more appropriate approach to calculating elastic properties of semiconductors in the context of DFT, in terms of quality of the parameters and efficiency of the calculation. This conclusion is further supported by the data in Section 4.3.5, obtained for other materials. Hence, based on reliability and computational efficiency, we believe there is no objective justification to derive elastic constants using a total energy calculation. This is especially true in cases where the particularities of the problem limit the convergence parameters (i.e., number of plane waves and **k** points) which are affordable, for example when modelling large supercells or when employing hybrid functional schemes (see Section 4.4) [3].

In contrast to what happens to elastic constants, the equilibrium lattice parameter obtained from a total energy calculation, as shown in Fig. 4.4 (b), is closer to the value to which a series of calculations would converge as one increases *both* energy cutoff and number of **k** points [Fig. 4.4 (b2)]. The lattice parameter ob-

tained from a stress calculation converges faster with the number of **k** points than total energy [cf. Fig. 4.4 (b1)], but to a (slightly) wrong value, unless the cutoff energy is increased to very high values [cf. Fig. 4.4 (b2)].[6] The difference is quite small, and in any case always below 0.1% for a calculation of reasonable accuracy. Therefore we recommend using the lattice parameter obtained from the stresses when calculating elastic constants, for the sake of consistency.

There is, however, an important issue for the stress method related to the optimization of the internal degrees of freedom of the unit cell used to represent a crystal. For these, a minimization of the total energy still needs to be done in order to obtain the correct arrangement of the atoms in the asymmetric unit for fixed lattice vectors (cf. Section 4.2.2).

4.3.4 Effect of reduced crystal symmetry on the robustness of the total energy and stress methods

We have considered in Sections 4.3.2 and 4.3.3 high symmetry cases, corresponding to both hydrostatic ($\epsilon_1 = \epsilon_2 = \epsilon_3$) and homogeneous shear ($\epsilon_4 = \epsilon_5 = \epsilon_6$) deformations. We will now test the robustness of total energy and stress methods by reducing the symmetry of the crystal, for AlN in both ZB and WZ phases. Three different strain branches will be considered:

$$\begin{aligned} \boldsymbol{\epsilon}^{(1)} &= [\delta, \delta, 0, 0, 0, 0]\,, \\ \boldsymbol{\epsilon}^{(2)} &= [\delta, 0, 0, 0, 0, 0]\,, \\ \boldsymbol{\epsilon}^{(3)} &= [\delta, -\delta, 0, 0, 0, 0]\,. \end{aligned} \tag{4.24}$$

Note that we have chosen $\boldsymbol{\epsilon}^{(1)}$ so it does *not* lower the symmetry of WZ crystals, $\boldsymbol{\epsilon}^{(2)}$ reduces it, and $\boldsymbol{\epsilon}^{(3)}$ reduces it faster than $\boldsymbol{\epsilon}^{(2)}$. For ZB crystals all strain branches in Eq. (4.24) do reduce the symmetry of the lattice, with $\boldsymbol{\epsilon}^{(3)}$ causing the greatest reduction. First, a standard two-atom ZB unit cell will be considered, followed by a standard four-atom WZ unit cell. Finally, a special six-atom hexagonal ZB unit cell oriented along the [111] direction will be checked for consistency with the two-atom case. For a direct comparison, we plot throughout this section

[6] The absolute error in the diagonal components of the stress tensor is referred to as "Pulay stress", and is due to basis set incompleteness [72]. Pulay stress always leads to an underestimation of the lattice parameter using the stress method. This underestimation is small for the cutoff energies employed here.

the strain energy, E_{tot} and E_σ, calculated using the two methods. Here E_σ is shifted so that its minimum lies at the same value as the minimum of E_{tot}. The equilibrium configuration, that is, the equilibrium lattice parameter(s), will be taken as the ones for which $\sigma = 0$. This leads to the E_{tot} energy curve having its minimum slightly displaced from $\delta = 0$. The reason for this is that the total energy method predicts an equilibrium lattice parameter about 0.05% larger than the stress method, as seen in Fig. 4.4 (b). However, this will not affect the *curvature* of E_{tot} since our calculations are well within the linear regime, which has been specifically tested by ensuring that all the data points lie on a parabola.

A single elastic constant and some particular combinations of elastic constants can in principle be computed from stresses with only two calculations and from total energy with only three, which correspond to the number of points that can unambiguously define a straight line and a parabola, respectively. To check that calculations for different values of δ within the same branch are consistent with each other, one needs to ensure that all of them lie on the *same* straight line or parabola. Therefore, for each case, we use 11 different values of δ, centred about equilibrium, spread evenly within the range $\pm 0.5\%$. In general, it should be ensured that the calculations are within the harmonic limit by including during the fitting at least one more point than strictly necessary.

Schematics of the real-space representation of the different unit cells used in this section, and the form of the internal strain for each of them, were given in detail in Section 4.2.2.

Two-atom zinc-blende unit cell

The calculations for the standard two-atom ZB unit cell lead to the results shown in Fig. 4.5. The elastic coefficient extracted from the strain branch $\boldsymbol{\epsilon}^{(1)}$ is $C_{11}+C_{12}$ [cf. Eqs. (4.2) and (4.21)], giving values of 468.7 GPa and 471.4 GPa for E_σ and E_{tot}, respectively. From $\boldsymbol{\epsilon}^{(2)}$, C_{11} can be extracted, with values of 301.3 GPa and 302.7 GPa for the stress and total energy methods, respectively. Finally, from $\boldsymbol{\epsilon}^{(3)}$, $C_{11} - C_{12}$ can be extracted, with resulting values of 134.0 GPa and 135.5 GPa. The relative differences between the two methods for the three branches are 0.6%, 0.5% and 1.1%, respectively. These results are summarized in Table 4.2. Thus, for high symmetry crystals with small unit cells like ZB, both methods can give similar values for the elastic constants, although the stress method yields results

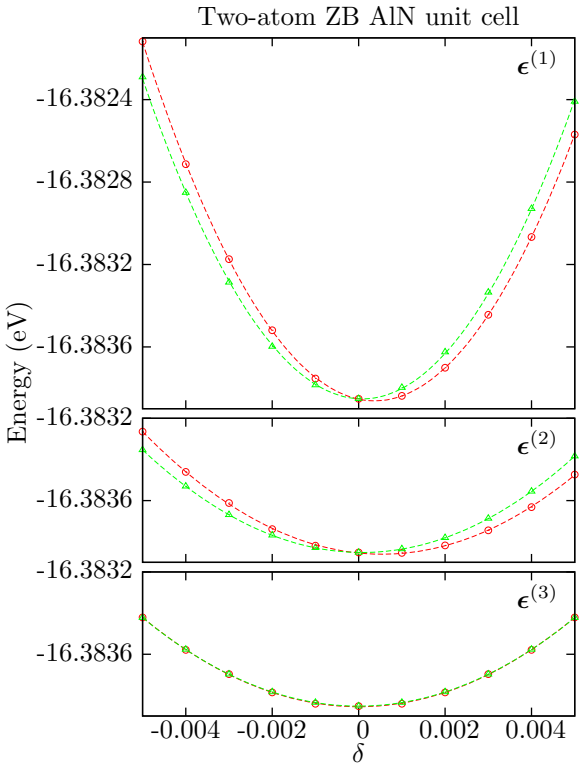


Figure 4.5: E_{tot} (circles) and E_σ (triangles) curves as a function of δ for all three strain branches $\boldsymbol{\epsilon}^{(1)}$, $\boldsymbol{\epsilon}^{(2)}$ and $\boldsymbol{\epsilon}^{(3)}$. Results for a two-atom ZB AlN unit cell, obtained with a $6 \times 6 \times 6$ Γ-centred k-mesh and cutoff energy of 800 eV. Curves are quadratic fit to the data. See text for further detail on the value of the coefficients.

that are more consistent for different strain branches. It must be noted that there were no internal parameters involved in these calculations, since internal strain in ZB is only present when (also) shear strain is applied [see Section 4.2.2].

Four-atom wurtzite unit cell

The case of WZ is extremely interesting given that the symmetry of the crystal is explicitly reduced with respect to the ZB case, including that there are two independent lattice parameters (a, c). The symmetry is further reduced implicitly due to the internal degrees of freedom. In contrast to the one independent internal strain parameter for ZB [146], there are five independent internal strain parameters in the WZ lattice (cf. Section 4.2.2) [3,4]. Constraining the internal degrees of freedom to what one would expect from the restricted crystal symmetry, improves the consistency of the optimized atomic positions for a strained WZ structure and the values of the internal strain parameters calculated for different strain branches.

Table 4.2: Elastic coefficients for ZB AlN extracted using two different approaches: total energy and stress from the three different strain branches considered in Fig. 4.5 and defined in Eq. (4.24).

Strain branch & coefficient (GPa)		Energy	Stress	Diff.
$\boldsymbol{\epsilon}^{(1)}$	$C_{11}+C_{12}$	471.4	468.7	0.6%
$\boldsymbol{\epsilon}^{(2)}$	C_{11}	302.7	301.3	0.5%
$\boldsymbol{\epsilon}^{(3)}$	$C_{11}-C_{12}$	135.5	134.0	1.1%
$\frac{1}{2}[\boldsymbol{\epsilon}^{(1)}+\boldsymbol{\epsilon}^{(3)}]$	C'_{11}	303.5	301.4	0.7%
$\frac{1}{2}[\boldsymbol{\epsilon}^{(1)}-\boldsymbol{\epsilon}^{(3)}]$	C'_{12}	168.0	167.4	0.4%
$\boldsymbol{\epsilon}^{(1)}-\boldsymbol{\epsilon}^{(2)}$	C''_{12}	168.7	167.4	0.8%
$\boldsymbol{\epsilon}^{(2)}-\boldsymbol{\epsilon}^{(3)}$	C'''_{12}	167.2	167.3	0.1%

Table 4.3: Elastic coefficients for the WZ AlN cell from the total energy and stress methods.

Strain branch & coefficient (GPa)		With internal constraints			No internal constraints		
		Energy	Stress	Diff.	Energy	Stress	Diff.
$\boldsymbol{\epsilon}^{(1)}$	$C_{11}+C_{12}$	541.4	543.7	0.4%	541.4	543.8	0.4%
$\boldsymbol{\epsilon}^{(2)}$	C_{11}	396.0	400.3	1.1%	394.0	398.4	1.1%
$\boldsymbol{\epsilon}^{(3)}$	$C_{11}-C_{12}$	254.1	256.9	1.1%	250.4	253.2	1.1%
$\frac{1}{2}[\boldsymbol{\epsilon}^{(1)}+\boldsymbol{\epsilon}^{(3)}]$	C'_{11}	397.8	400.3	0.6%	395.9	398.5	0.7%
$\frac{1}{2}[\boldsymbol{\epsilon}^{(1)}-\boldsymbol{\epsilon}^{(3)}]$	C'_{12}	143.7	143.4	0.2%	145.5	145.3	0.1%
$\boldsymbol{\epsilon}^{(1)}-\boldsymbol{\epsilon}^{(2)}$	C''_{12}	145.4	143.4	1.4%	147.4	145.4	1.4%
$\boldsymbol{\epsilon}^{(2)}-\boldsymbol{\epsilon}^{(3)}$	C'''_{12}	141.9	143.4	1.1%	143.6	145.2	1.1%

For comparison, we present in Fig. 4.6 results both constraining the internal degrees of freedom to what is allowed by the WZ crystal symmetry and allowing a free relaxation of all the atomic coordinates. These two sets of calculations should in principle give the same results. All the elastic coefficients that can be extracted from the calculations and how they compare to each other are summarized in Table 4.3. The numbers are similar although there are sizable discrepancies when constraining certain degrees of freedom, which gives an account of the impact of the internal strain on the macroscopic elastic constants. Also, it is important to note how the elastic constants derived with the stress method in different branches are much more consistent with each other than the ones obtained using the total energy method.

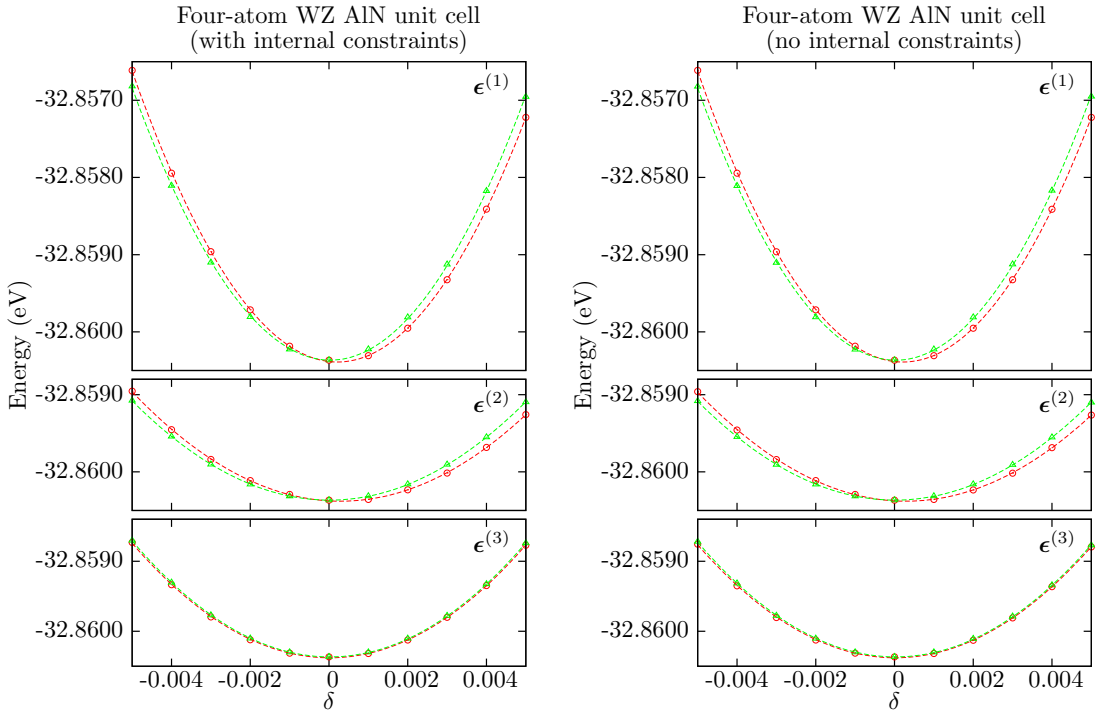


Figure 4.6: E_{tot} (circles) and E_σ (triangles) curves as a function of δ for all three strain branches $\boldsymbol{\epsilon}^{(1)}$, $\boldsymbol{\epsilon}^{(2)}$ and $\boldsymbol{\epsilon}^{(3)}$. Results for a four-atom WZ AlN unit cell, obtained with a $6 \times 6 \times 4$ Γ-centred k-mesh and cutoff energy of 800 eV. Curves are quadratic fit to the data. See text for further detail on the value of the coefficients.

Six-atom [111]-oriented zinc-blende unit cell

As seen in Section 4.2.2, it is possible to construct a "WZ-like" ZB unit cell consisting of a hexagonal cell with the c-axis, or [0001] direction of the lattice, oriented along the [111] direction of the original ZB primitive cell. The unit cell in this case contains six atoms. Using this unnecessarily large unit cell allows to check the behaviour of the elastic properties of a ZB material once the symmetry of the lattice is artificially lowered with respect to the original case. Also, the number of internal degrees of freedom is increased. In this way, any behaviour inconsistent with the two-atom primitive cell will be exposed. This will uncover potential drawbacks to the plane wave formalism related to symmetry and unit cell size, since the results for converged calculations should agree perfectly, regardless of the size of the unit cell. Therefore, any differences arising from calculations of similar accuracy would indicate that convergence is "unit cell dependent".

The transformation of the strain tensor is done via a unitary transformation with the appropriate rotation matrix. More information can be found in our

work on [111]-oriented ZB heterostructures [6]. The rotated strain transformation matrices for hydrostatic deformation and for homogeneous shear deformation along the [111] direction, $\hat{T}^{[111]}_{\text{hydrostatic}}$ and $\hat{T}^{[111]}_{\text{shear}}$, take a very simple form:

$$\hat{T}^{[001]}_{\text{hydrostatic}} = \begin{pmatrix} 1+\epsilon_h & 0 & 0 \\ 0 & 1+\epsilon_h & 0 \\ 0 & 0 & 1+\epsilon_h \end{pmatrix}$$

$$\rightarrow \qquad \hat{T}^{[111]}_{\text{hydrostatic}} = \begin{pmatrix} 1+\epsilon_h & 0 & 0 \\ 0 & 1+\epsilon_h & 0 \\ 0 & 0 & 1+\epsilon_h \end{pmatrix}$$

$$\hat{T}^{[001]}_{\text{shear}} = \begin{pmatrix} 1 & \frac{\epsilon_s}{2} & \frac{\epsilon_s}{2} \\ \frac{\epsilon_s}{2} & 1 & \frac{\epsilon_s}{2} \\ \frac{\epsilon_s}{2} & \frac{\epsilon_s}{2} & 1 \end{pmatrix} \qquad \rightarrow \qquad \hat{T}^{[111]}_{\text{shear}} = \begin{pmatrix} 1-\frac{\epsilon_s}{2} & 0 & 0 \\ 0 & 1-\frac{\epsilon_s}{2} & 0 \\ 0 & 0 & 1+\epsilon_s \end{pmatrix} \tag{4.25}$$

whereas for the strain branches $\boldsymbol{\epsilon}^{(1)}$, $\boldsymbol{\epsilon}^{(2)}$ and $\boldsymbol{\epsilon}^{(3)}$, defined in Eq. (4.24), a more complicated matrix is obtained:

$$\hat{T}^{[001]}_{12} = \begin{pmatrix} 1+\epsilon_1 & 0 & 0 \\ 0 & 1+\epsilon_2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\rightarrow \qquad \hat{T}^{[111]}_{12} = \begin{pmatrix} 1+\frac{\epsilon_1}{2} & \frac{-\epsilon_1}{2\sqrt{3}} & \frac{\epsilon_1}{\sqrt{6}} \\ \frac{-\epsilon_1}{2\sqrt{3}} & 1+\frac{\epsilon_1+4\epsilon_2}{6} & \frac{2\epsilon_2-\epsilon_1}{3\sqrt{2}} \\ \frac{\epsilon_1}{\sqrt{6}} & \frac{2\epsilon_2-\epsilon_1}{3\sqrt{2}} & 1+\frac{\epsilon_1+\epsilon_2}{3} \end{pmatrix} \tag{4.26}$$

which we have obtained using the unitary transformation $\hat{U}$ given in Eq. (4.13). The matrices in Eqs. (4.25) and (4.26) are given in terms of the original [001] strain components. For consistency, the rotation of the lattice vectors as well as the atomic positions from the [001] convention to the [111] also have to be made employing $\hat{U}$. For a fair comparison with the two-atom ZB unit cell, we have chosen a Γ-centred $6\times6\times2$ k-mesh, which maintains the same density of $\mathbf{k}$ points in the Brillouin zone as we used in the previous test. The results of the tests for hydrostatic and shear deformation calculations for this six-atom ZB cell are shown in Fig. 4.7. The pure shear and hydrostatic cases serve as a test for the validity of the comparison, since in these high symmetry cases one should obtain the same results as for the two atom case. In the case of shear deformation ($\epsilon_4 = \epsilon_5 = \epsilon_6$),

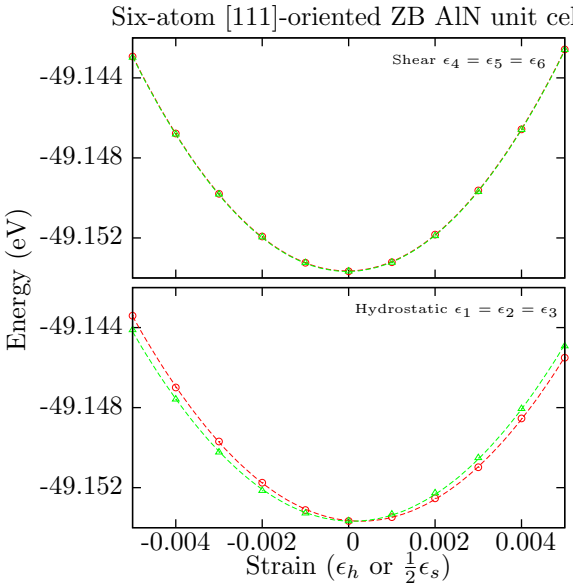


Figure 4.7: E_{tot} (circles) and E_σ (triangles) curves as a function of strain for the hydrostatic and shear cases. Results for a six-atom [111]-oriented ZB AlN unit cell, obtained with a $6 \times 6 \times 2$ Γ-centred k-mesh and cutoff energy of 800 eV. Curves are quadratic fits to the data. See text for further detail on the value of the coefficients.

the elastic constant C_{44} derived is 189.1 GPa from total energy and 188.8 GPa from stress. These values are in good agreement with the values obtained for the two-atom primitive cell, which were 188.7 GPa and 187.8 GPa, respectively. From the hydrostatic strain calculation ($\epsilon_1 = \epsilon_2 = \epsilon_3$), $C_{11} + 2C_{12}$ can be extracted. Values of 638.2 GPa and of 635.0 GPa are obtained respectively from the total energy and the stress calculations. These values are again in good agreement with the values obtained for the two-atom case (Table 4.2). Within the numerical accuracy of the calculations, whether a relaxation of the internal degrees of freedom is allowed or not, does not have any effect on the calculated value of $C_{11}+2C_{12}$. It has therefore been shown that both the primitive and enlarged ZB cells are well suited for the derivation of elastic properties, yielding similar results for these high symmetry cases, with less than a half percent difference in the worst case.

Figure 4.8 shows the results for the strain branches $\boldsymbol{\epsilon}^{(1)}$, $\boldsymbol{\epsilon}^{(2)}$ and $\boldsymbol{\epsilon}^{(3)}$, where the strain is defined in terms of the unrotated [001]-oriented ZB crystal. The derived values of the elastic coefficients are shown in Table 4.4, which are independent of whether a relaxation of the internal degrees of freedom was allowed or not. Three main observations can be made. First, the elastic constants derived for different strain branches are more consistent with each other when obtained using the stress

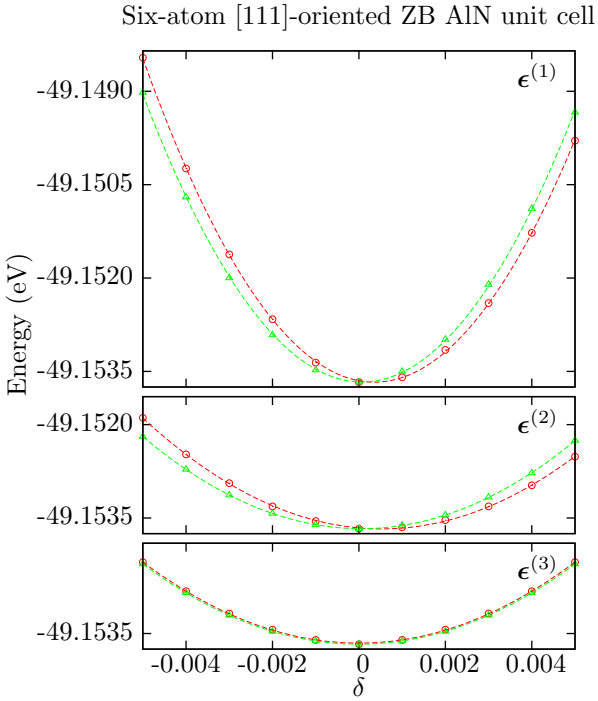


Figure 4.8: E_{tot} (circles) and E_σ (triangles) curves as a function of strain for the strain branches $\boldsymbol{\epsilon}^{(1)}$, $\boldsymbol{\epsilon}^{(2)}$ and $\boldsymbol{\epsilon}^{(3)}$. Results for a six-atom [111]-oriented ZB AlN unit cell, obtained with a $6 \times 6 \times 2$ Γ-centred k-mesh and cutoff energy of 800 eV. Curves are quadratic fit to the data. See text for further detail on the value of the coefficients.

method than with the total energy method. This does not come across directly from the data shown in Table 4.4, only because the choice of a $6 \times 6 \times 2$ k-point mesh happens to be a rather fortunate one, giving an uncertainty in the value of C_{12} of 0.2 GPa for total energy and 0.4 GPa from the stress method. Surprisingly, testing with denser meshes we obtained higher uncertainties for the total energy method of 1.4 GPa ($6 \times 6 \times 4$), 1.6 GPa ($8 \times 8 \times 4$) and 0.7 GPa ($8 \times 8 \times 6$), while the uncertainty of the stress method remained constant. This highlights again the dependence of the total energy results on the choice of k-mesh, as previously shown in Fig. 4.4. Second, whether the internal degrees of freedom are allowed to relax or not does not have any significant impact on the result of the calculations for this particular case. This reflects the symmetry of the forces in the ZB lattice under axial deformation (i.e. no internal strain), and cannot be generalized to other problems where internal strain is present (see Section 4.2.2). Third and last, changing the representation of the crystal does not seem to have a sizable impact on the elastic constants for calculations of similar accuracy, at least when

Table 4.4: Elastic coefficients that can be extracted from the three different strain branches, obtained for the six-atom ZB AlN cell using the two different approaches: total energy and stress.

Strain branch & coefficient (GPa)		Energy	Stress	Diff.
$\boldsymbol{\epsilon}^{(1)}$	$C_{11}+C_{12}$	471.2	467.7	0.7%
$\boldsymbol{\epsilon}^{(2)}$	C_{11}	303.5	301.0	0.8%
$\boldsymbol{\epsilon}^{(3)}$	$C_{11}-C_{12}$	135.6	133.9	1.3%
$\frac{1}{2}[\boldsymbol{\epsilon}^{(1)}+\boldsymbol{\epsilon}^{(3)}]$	C'_{11}	303.4	300.8	0.9%
$\frac{1}{2}[\boldsymbol{\epsilon}^{(1)}-\boldsymbol{\epsilon}^{(3)}]$	C'_{12}	167.8	166.9	0.5%
$\boldsymbol{\epsilon}^{(1)}-\boldsymbol{\epsilon}^{(2)}$	C''_{12}	167.7	166.7	0.6%
$\boldsymbol{\epsilon}^{(2)}-\boldsymbol{\epsilon}^{(3)}$	C'''_{12}	167.9	167.1	0.5%

there is no internal strain contribution to them, or when there is a preferential direction for the optimization of the internal degrees of freedom. This is the case for homogeneous shear deformation in ZB and strain branch $\boldsymbol{\epsilon}^{(1)}$ in WZ. It should be noted that the differences between methods are largest for the strain branch which causes the largest symmetry reduction, $\boldsymbol{\epsilon}^{(3)}$, in the two ZB and the WZ cells used.

Yu *et al.* [152] have recently presented a method, based on the calculation of stresses, to efficiently obtain all the independent elastic constants of a material, with as few calculations as possible. They have derived a series of "universal linear-independent coupling strains" (ULICS) for this purpose. Their approach involves the use of highly asymmetric strain tensors, corresponding to the different ULICS. They differ from our strain branches in that they involve an arbitrary reduction of the crystal symmetry, which we have shown to affect the results. Therefore, this should be taken into account when employing Yu's method for a very precise calculation of elastic constants. We also note that by reducing the symmetry of the Brillouin zone the number of equivalent $\mathbf{k}$ points is decreased and more of them need therefore to be explicitly included in the calculation, adding to the computational cost. We have carried out calculations for a two-atom ZB AlN unit cell for all the six different ULICS given in [152]. For a cutoff energy of 800 eV and a $6\times6\times6$ Γ-centred k-mesh, we obtain C_{11} in the range 295.5–306.0 GPa, C_{12} in the range 159.7–170.1 GPa, and C_{44} in the range 184.2–189.0 GPa. These uncertainty margins, in line with the 1–2% reported in [152], are relatively large compared to those for the results we present in this section for the two-atom ZB

unit cell. That data shows an almost negligible difference of about 0.1 GPa for C_{11} and C_{12} calculated from stresses in different branches. Increasing the number of **k** points and/or cutoff energy did not lead to any significant improvement in the agreement between the elastic constants obtained for different ULICS. The accuracy of Yu's approach can be improved by employing two calculations for each ULICS: a regular one, and an additional calculation where all the strain components change sign. Interpolating the stresses for these two calculations we managed to reduce the disagreement between elastic constants calculated with different ULICS to 0.2 GPa.

4.3.5 Additional results for selected materials

To test how the convergence properties of elastic constants apply to materials other than AlN, we have repeated the hydrostatic strain calculation ($\epsilon_1 = \epsilon_2 = \epsilon_3$) for several common materials, mostly semiconductors. These are C (insulator) and Si, with diamond structure, and GaAs and GaN, with ZB structure. The test calculation is the same as carried out for AlN, whose results were shown in Fig. 4.4 (a). This calculation allows to obtain $C_{11} + 2C_{12}$. It can be observed in Fig. 4.9 that the results are consistent with our findings for AlN: the stress-strain calculation converges much more rapidly than an energy-strain calculation with increasing number of **k** points. Again, increasing the cutoff energy allows to increase the convergence rate of the energy method. The optimum cutoff energy varies depending on the pseudopotentials used for the specific material. The default cutoff energies in VASP, for the PAW-LDA pseudopotential library, are 400 eV for C, 246 eV for Si, 283 eV for Ga, 209 eV for As and 400 eV for N [72]. The recommended cutoff for a calculation involving the computation of the stress tensor is 1.3 times the default cutoff, in order to avoid issues related to Pulay stress [72]. The cutoff energy used in the bottom panels of Fig. 4.9 for each material is roughly 1.5 times the highest cutoff of any of the elements present in the calculation, (except for Si, where it is about 1.2 times), which rules out Pulay stress as the origin of the difference between the two methods. These calculations confirm that the analysis carried out here for AlN can also be applied to other semiconductors and insulators, and that the stress method should be generally preferred for the calculation of the elastic constants of these materials.

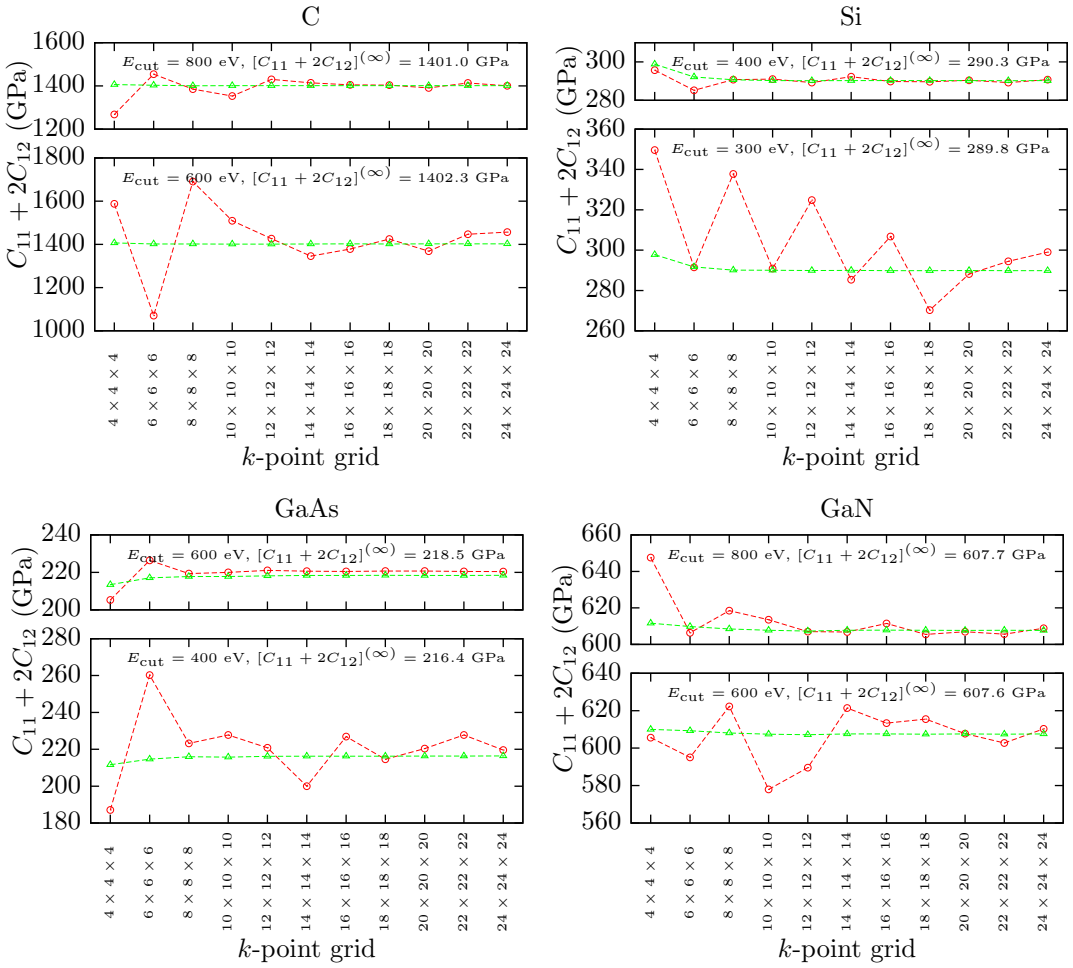


Figure 4.9: Same calculation as Fig. 4.4 (a), repeated for C and Si, both with the diamond structure, and GaAs and GaN, with the ZB structure. Triangles indicate results for the stress-strain method and circles indicate results for energy-strain method. Note that the cutoff energies needed to improve convergence of the total energy method vary for the different compounds and relate to the default cutoff for the specific PAW-LDA pseudopotentials used for each compound (see text). The given value $[C_{11}+2C_{12}]^{(\infty)}$ indicates the value to which the stress-strain calculation converges in each case as the number of **k** points tends to infinity.

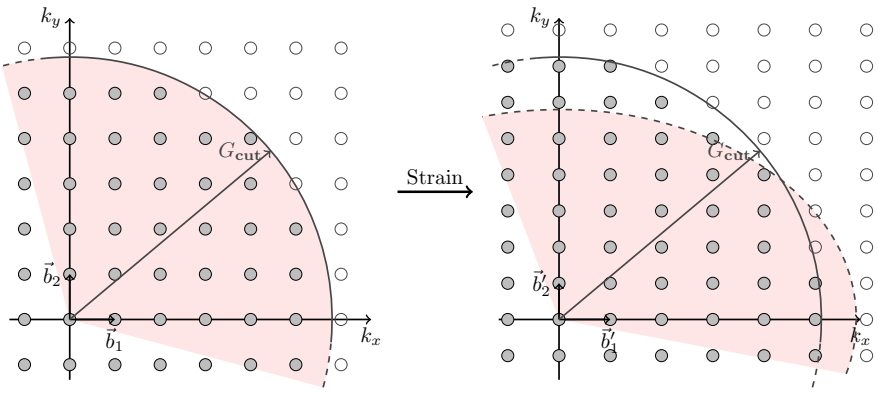


Figure 4.10: (left) Undistorted and (right) distorted reciprocal lattices. For a fixed cutoff energy, all the plane waves contained within the cutoff sphere are included in the calculation. The basis set originally used for the undistorted lattice calculation is now contained in an ellipsoid for the distorted lattice.

4.3.6 Correcting the total energy method: strain-adapted basis set

The effect of changing the basis set between two different calculations is a finite shift in the total energy, as previously shown in Figs 4.2 and 4.3. If this basis set change is carried out at the same time as a change of the lattice vectors, then that finite shift due to basis set difference is superimposed on the strain energy. This in turns leads to errors in the calculation of elastic constants with a total energy method.

What happens when the lattice vectors are changed and the same cutoff energy is kept is schematically shown with a two-dimensional analogy in Fig. 4.10. The shaded areas indicate the plane waves (or reciprocal lattice translations) that are used in the undistorted lattice calculation (inside the cutoff sphere) and their direct correspondence in the distorted lattice (ellipsoid). Dark dots indicate the plane waves that are included in the actual calculation in each case, when the cutoff energy is kept constant.

For shape-conserving calculations, the cutoff sphere and the basis set change isotropically. For distorted cells, the cutoff sphere becomes an ellipsoid and the basis set changes differently in different directions. As a consequence, stress and total energy methods show the biggest discrepancies for cell-distorting strains (e.g. biaxial) and the smallest discrepancies for hydrostatic deformation.

The total energy results can in principle be corrected by using a (in general anisotropic) strain-dependent cutoff energy. For small strains, $\pm 1\%$ change in lattice vectors corresponds to $\sim \mp 2\%$ change in cutoff energy, given the re-

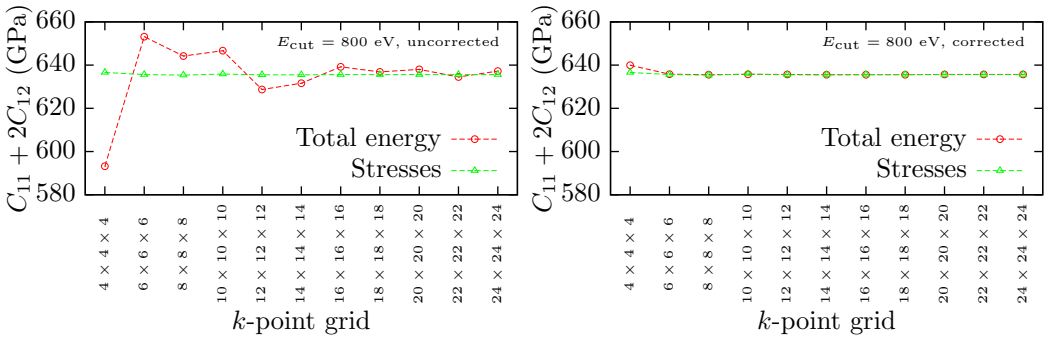


Figure 4.11: (left) Results for a fixed cutoff energy of 800 eV, previously given in Fig. 4.4 (a2), and (right) with a strain-corrected cutoff around 800 eV, for the same system (ZB AlN under hydrostatic deformation).

lation of Eq. (4.23). Implementing this correction leads to a remarkably improved agreement between total energy and stress methods, as shown in Fig. 4.11. Unfortunately, this correction can only be easily implemented for hydrostatic strain (i.e. only allows to calculate bulk modulus) because the basis set needs to change isotropically. To avoid these issues, the stress method should be used for reliable, consistent and computationally inexpensive calculation of elastic constants.

4.4 Elastic constants and structural parameters of group-III nitrides using the HSE hybrid functional scheme

For a reliable modelling of nitride-based devices, an accurate knowledge of material parameters such as lattice constants, elastic constants, piezoelectric coefficients, etc. is of crucial importance [1]. Ideally, these calculations should rely on experimental data. However, the access to some of the properties of crystalline solids is in many cases difficult. The experimental determination of elastic constants, for instance, depends on the ability to independently apply different stresses to a particular sample. Because of this, the obtained values of the elastic constants related to deformations that distort the unit cell are commonly accompanied by large uncertainties [153]. The measurements of internal parameters such as the Kleinman parameter [145] in a zinc-blende lattice are not easily available to experiment. In the case of materials where high quality samples are difficult to achieve, as is the case for zinc-blende nitrides [22], theoretical calculations are often a safer route towards the determination of material parameters. Therefore, accurate theoretical calculations are needed in order to bridge this gap and complement ex-

perimentally determined parameters with those that are not easily accessible in the laboratory. Density functional theory (DFT) within the Heyd-Scuseria-Ernzerhof (HSE) screened exchange hybrid functional scheme [62, 63] offers a high degree of accuracy. The HSE approach has been successful in solving the band gap problem for semiconductors, present when using standard Kohn-Sham approaches to calculating the exchange energy [64]. Paier *et al.* [65,66] also showed that HSE provides improved lattice parameters and bulk moduli for a wide range of semiconductors. Finally, HSE has been used by Yan *et al.* to predict other well known properties of group-III nitrides, such as the crystal-field splitting, in excellent agreement with experiment [113, 154]. This good agreement provides confidence in the quality of other parameters that have not been experimentally determined or whose experimental values are given with a large degree of uncertainty due to the quality of the samples or the complexity of the measurement procedure. We systematically apply the HSE approach here, together with the methodology presented in the previous sections, in order to calculate the structural and elastic properties of WZ as well as ZB GaN, InN and AlN.

The present calculations have been performed within the HSE scheme [62, 63], using the projector augmented-wave (PAW) method [68, 74], as implemented in VASP [72, 75]. For WZ materials we follow the recipe presented by Yan *et al.* [113]: the screening parameter μ was fixed to 0.2 and the mixing parameter α to 0.25 (which correspond to the HSE06 functional), the cutoff energy for plane waves was 600 eV, the semicore d-electrons of In and Ga were treated as valence electrons and a Γ-centred $6 \times 6 \times 4$ k-point grid was used. In the case of ZB nitrides, a similar scheme is used, with a Γ-centred $6 \times 6 \times 6$ k-point grid, accounting for the cubic symmetry of the crystal, but with the same settings otherwise. To check the accuracy of the calculations, finer k-point grids and higher cutoff energies were used in test calculations, for which we found no more than a 0.1% change in the equilibrium lattice parameter values.

To obtain the elastic constants and internal strain parameters, strains of ± 0.2 % are applied and the results of the calculation fitted to Eqs. (4.2), (4.14) and (4.15). The strain values have been tested and chosen to be small enough to be in the linear regime implied by those expressions, while still large enough to minimize the effects of any numerical error (noise) in the output of the calculation, which is higher for smaller strains. In order to minimize the impact of the numerical error and the finite k-sampling on the results, it is also crucial to constrain the

internal degrees of freedom of the cells to the displacements allowed by the crystal symmetry, as given by Eqs. (4.14) and (4.15). For example, a WZ material system that undergoes shear strain ϵ_{xy}, has fixed coordinates imposed for all the atoms contained in the unit cell, in accordance with the corresponding transformation $\mathbf{r} = (\mathbb{1} + \epsilon)\,\mathbf{r}_0$ [cf. Eq. (4.11)], except for the first coordinate of atoms B, C and D. Those particular components of the position vectors of B, C and D must be allowed to move according to the value of the internal strain parameters ζ_4 and ζ_5, as given by Eq. (4.15). An alternative way to calculate elastic constants is with a "clamped ion" model (see Nielsen and Martin [138] and Section 4.2.2), where the contribution of the internal strain to the macroscopic elastic constant is evaluated separately. Our tests indicate that the clamped-ion method yields the same results as the method presented here, in which an internal optimization is carried out for the strained structure.

The results of our calculations are listed in Table 4.5, together with the available experimental data (in brackets). The lattice parameters of the WZ materials are in line with the ones previously calculated by Yan *et al.* [113] except for InN, for which our values are in better agreement with experiment. Looking at the elastic constants, our results for GaN and AlN are in good agreement with the experimental data reported by Polian *et al.* [153] and McNeil *et al.* [155], respectively. Both authors used the same methodology (Brillouin scattering) to obtain their results. Our set of elastic constants for InN is also in good agreement with the recent experimental results of Serrano *et al.* [156], although as in other theoretical works, we obtain C_{13} about 20% lower than C_{12} [143, 156]. As already stated, the experimental determination of elastic constants related to strains that distort the unit cell, such as C_{12} or C_{13}, is usually accompanied by large uncertainties [153, 156]. In the case of the internal strain parameters, ζ_2 and ζ_3 are similar to the estimates that Camacho and Niquet [147] made from the *ab initio* calculations of Wagner and Bechstedt [105]. There is no previous published data for ζ_1, ζ_4 and ζ_5. The importance of an accurate theoretical determination of the internal strain parameters is two-fold. On the one hand, an experimental determination of the same is not possible with the technological capabilities currently available. On the other hand, it provides a benchmark to test the accuracy of empirical interatomic models, such as valence force field (VFF) models [146, 147], used in atomistic simulations for the calculation of atomic positions.

In the case of the ZB nitrides, the lattice parameters are in good agreement with

Table 4.5: Results of the HSE calculations for the WZ and ZB group-III nitrides. Experimental values (when available) are in parenthesis. Lattice parameters taken from Ref. [20] for WZ and ZB nitrides. Elastic constants of WZ GaN and AlN from Refs. [153, 155], and WZ InN from Ref. [156]. Elastic constant C_{11} of ZB GaN from Ref. [157].

	Wurtzite				Zinc-blende		
	GaN	InN	AlN		GaN	InN	AlN
a_0 (Å)	3.180	3.542	3.103	a_0 (Å)	4.489	4.988	4.363
	(3.189)	(3.545)	(3.112)		(4.50)	(4.98)	(4.38)
c_0 (Å)	5.172	5.711	4.970				
	(5.185)	(5.703)	(4.982)				
u_0	0.3772	0.3796	0.3818				
C_{11} (GPa)	368.6	233.8	410.2	C_{11} (GPa)	288.6	183.2	308.9
	(390)	(225)	(410.5)		(285)		
C_{12} (GPa)	131.6	110.0	142.4	C_{12} (GPa)	154.1	119.2	166.5
	(145)	(109)	(148.5)				
C_{13} (GPa)	95.7	91.6	110.1				
	(106)	(108)	(98.9)				
C_{33} (GPa)	406.2	238.3	385.0				
	(398)	(265)	(388.5)				
C_{44} (GPa)	101.7	55.4	122.9	C_{44} (GPa)	166.0	91.5	196.0
	(105)	(55)	(124.6)				
ζ_1	0.156	0.193	0.138	ζ	0.58	0.76	0.545
ζ_2	0.083	0.107	0.086				
ζ_3	0.159	0.218	0.191				
ζ_4	0.201	0.337	0.199				
ζ_5	0.141	0.107	0.143				

experiment and the elastic constants are in accordance with previous theoretical calculations, except for C_{44}, for which we obtain a value about 7% higher than Wright [143]. The values of the Kleinman parameters are also in good agreement with Wright's calculations. Little experimental data exists for ZB nitrides given the complexity of the growth process and the quality of the crystal samples. The only experimentally determined elastic constant, $C_{11} = 285$ GPa for GaN, reported by Moss *et al.* [157], is in good agreement with our value of 288.6 GPa.

4.5 Accuracy of theoretical calculations in the context of experimental results

We have so far extensively reviewed the methodology for the theoretical calculation of elastic properties of solids in the context of plane-wave based DFT. We have also presented strategies towards achieving converged results in an affordable way. However, we have not dealt with the issue of whether these converged results are more representative of the actual values compared with experiment, when taking the measurement uncertainty into account.

One of the obvious advantages of theoretical calculations is the ability to study ideal systems, whereas experiments are influenced by the quality of the material sample under study, given by the possible existence of dislocations, stacking faults, phase separation, built-in strain, etc. In the specific case of GaN, Polian *et al.*'s work [153] is the most cited source of elastic constants. However, they report errors as large as 19 % for C_{13} arising from the spread in experimental data used during the fitting – that is, assuming there is no intrinsic error in the experiment. Looking at coetaneous experimental results by Schwarz and collaborators [158] one observes significant disagreement between their results and those of Polian, with a difference of about 100 % between their values for C_{33}.

In this context, whether theoretical calculations can be trusted or not will mostly rely on whether the theoretical predictions match the experimental results when *accurate* experimental results are available. For the widely studied Si, the value of $C_{11} + 2C_{12} = 290.3$ GPa that we report in Fig. 4.9 agrees within approximately 1 % with the commonly accepted value [159]. Therefore, the fact that DFT results for the elastic constants can offer good agreement with the values for well established materials strengthens its validity to calculate these properties for less well studied compounds. This is also the case for quantities that are impractical to measure in the laboratory, such as internal strain parameters, for which relying on the theoretical results is a matter of necessity. Moreover, as we have discussed in the previous section and supported by Paier's results for a wide range of materials [65,66], the use of more accurate functionals, such as the HSE hybrid, leads to additional accuracy towards the goal of highly realistic theoretical elastic constants.

4.6 Summary

In this chapter, we have presented a complete methodology for the calculation of elastic and structural properties of tetrahedrally bonded semiconductors, including the effects of internal strain. We then have used ZB and WZ AlN to undertake a detailed comparison of the two methods commonly used to calculate elastic constants in the context of plane wave-based density functional theory. In particular, we have shown that calculating elastic constants using the stress tensor as given from the Hellmann-Feynman theorem is more robust than a total energy calculation in terms of the computational cost required to achieve convergence and in terms of the consistency of elastic constants obtained using different strain branches for the same material. The stress method converges considerably faster than the total energy method with respect to the cutoff energy and number of **k** points used, although both methods converge, as expected, to the same values. We have also shown how to correct the total energy results via a strain-dependent cutoff energy, allowing to achieve convergence similar to the stress method. However, the latter correction can only be easily carried out in the case of hydrostatic deformation, given the manner in which basis set selection is commonly implemented in modern DFT packages. We recommend therefore use of the stress method in any calculations for which cutoff energies and number of **k** points have to be kept down to modest values due to the computational costs involved. This is the case for large supercells (e.g. modelling of alloys) and hybrid functional schemes [3].

We have also found that in the three cases studied here, two- and six-atom ZB unit cells and the standard four-atom WZ unit cell, discrepancies between stress and total energy methods increase as the crystal symmetry is lowered: in all three cases differences are largest for the strain branch $\boldsymbol{\epsilon}^{(3)} \equiv [\delta, -\delta, 0, 0, 0, 0]$. It is generally found as well that elastic constants calculated for different strain branches are more consistent when obtained from stress than from total energy. For the systems studied, we found that whether a hexagonal or cubic representation of the ZB crystal is used does not seem to have a significant impact on the consistency and robustness of either method.

Finally, we have obtained a complete and consistent set of structural and elastic parameters for the group-III nitrides in both WZ and ZB phases. Our results were obtained within the highly accurate HSE hybrid functional scheme and are in good agreement with available experimental and theoretical data. Also, the ζ_1, ζ_4 and

ζ_5 internal strain parameters have been reported for the three WZ compounds studied, which were not previously available in the literature.

5 Theory of local electric polarization and its relation to internal strain in group-III nitrides

Abstract

We present a theory of local electric polarization in crystalline solids and apply it to study the case of wurtzite group-III nitrides. We show that a local value of the electric polarization, evaluated at the atomic sites, can be cast in terms of a summation over nearest-neighbour distances and Born effective charges. Within this model, the local polarization shows a direct relation to internal strain and can be expressed in terms of internal strain parameters. The predictions of the present theory show excellent agreement with a formal Berry phase calculation for random distortions of a test-case CuPt-like InGaN alloy and InGaN supercells with randomly placed cations. While the present level of theory is appropriate for highly ionic compounds, such as III-N materials, we show that a more complex model is needed for less ionic materials, such as GaAs, in which the strain dependence of Born effective charges has to be taken into account. We provide *ab initio* parameters for GaN, InN and AlN, including hybrid functional values for the piezoelectric coefficients and the spontaneous polarization. In order to calculate the local polarization potential, we also present a point dipole method. This method overcomes several limitations related to discretization and resolution which arise when obtaining the local potential by solving Poisson's equation on an atomic grid.

5.1 Introduction

The effect of polarization discontinuities in strained semiconductor nanostructures has been treated both theoretically and experimentally. Because of the difficulty to obtain direct access to piezoelectric fields in strained crystals, the theoretical approach has been preferred during recent times and an extensive collection of

studies on piezoelectric properties of bulk materials or nanostructures exists in the literature [5, 58, 89, 160]. Of particular interest are systems such as GaAs-based quantum dots (QDs), whose electronic properties are affected by the symmetry of strain and piezoelectric fields [6, 161]. This effect is even more important in GaN-based heterostructures, where the large piezoelectric response together with the intrinsic spontaneous polarization give rise to built-in electrostatic fields far exceeding those encountered for the other III-V [1,89,91,94,100,124,125]. Although these effects have been studied over the last two decades, only recently has come to attention the possible role that the *local* polarization potential might have on them [5].

Many theoretical studies that include a treatment of polarization fields operate at a continuum level (even if the strain itself is obtained from an atomistic calculation), and the polarization is assumed to have a smooth behaviour with local strain and composition, even in the case of alloys. We have previously shown for InGaN alloys that a local value of polarization can be obtained, observing large fluctuations in its value at a microscopic scale [5]. In this chapter we lay our theory of local polarization on more solid ground, giving general equations and providing a direct link with internal strain.

Furthermore, when trying to calculate the local polarization potential by solving Poisson's equation $\nabla \cdot (\varepsilon \nabla \phi) = \nabla \cdot \mathbf{P}$ in the same atomic grid where the polarization is given, one encounters two main difficulties. One is given by the discretization of the medium, which is irregular given the arrangement of the atoms in the strained crystal. The second, and most important, is a problem of resolution: because Poisson's equation needs to be solved in a finite difference or polynomial interpolation schemes, and its solution involves the calculation of several derivatives (see, for instance, Bester and Zunger [162]), approximate interpolations have to be made and the effects of abrupt local discontinuities are lost in the process. In order to compute the local polarization potential and overcome these limitations, we have previously used a point dipole model [5]. Here we give the details of our model and extend it, as well as assess its limitations and degree of validity for calculations involving a position-dependent value of the polarization.

5.2 Theory of local electric polarization

Piezoelectricity is the emergence of a net macroscopic polarization as a result of strain induced by stress applied to a material sample. Historically, this phenomenon has been defined and treated at the macroscopic level. However, the advent in recent times of an array of piezoelectric devices that operate at the nanoscale raises the question of how microscopic changes in strain affect piezoelectricity. Additionally, one needs to answer the question of what role the local lattice distortions present in an alloy play in the wider picture of piezoelectricity.

The piezoelectric response of a material to strain is modelled, in the linear regime,[1] via the piezoelectric tensor e_{ij}:

$$P_i^{\mathrm{pz}} = \sum_{j=1}^{6} e_{ij}\epsilon_j, \tag{5.1}$$

where P_i^{pz} are the components of the piezoelectric polarization vector and ϵ_j are the strains, given in Voigt notation.[2] The symmetry of the crystal determines the non-zero elements of e_{ij}. We shall see further on that, even for a bulk binary compound, one can define a *local* piezoelectric tensor e_{ij}^* whose *average* over the unit cell reduces to e_{ij}, but that has in general more non-zero elements than e_{ij}. The total polarization vector is given by

$$P_i = P_i^{\mathrm{pz}} + P_i^{\mathrm{sp}}, \tag{5.2}$$

where P_i^{sp} are the components of the spontaneous polarization vector, that will be present only if the crystal symmetry allows, as previously discussed.

As we have seen in Chapter 2, a rigorous frame for the computation of polarization in periodic solids was not available until as recently as the 1990s. The main developments were presented in the seminal papers by Vanderbilt and King-Smith [78,79], building up on an idea originally suggested by Resta [80], where the foundations of the *Berry-phase* theory of polarization, or *modern theory of polarization* [81], where laid. As we have discussed in Section 2.3, this theory allows a calculation of the dipole moment of the unit cell of a periodic insulating system,

[1] One can also define a second-order piezoelectric tensor to characterize piezoelectricity further away from equilibrium. See Grimmer [42].

[2] Note that in Voigt notation, $\epsilon_1 = \epsilon_{xx}$, $\epsilon_2 = \epsilon_{yy}$, $\epsilon_3 = \epsilon_{zz}$, $\epsilon_4 = 2\epsilon_{yz}$, $\epsilon_5 = 2\epsilon_{xz}$ and $\epsilon_6 = 2\epsilon_{xy}$.

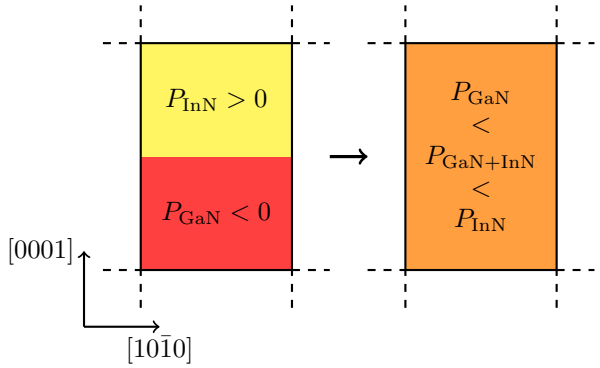


Figure 5.1: Periodic supercell consisting of alternating strained layers of GaN and InN. On the left, the polarization is given by the strain state of each layer. On the right, the polarization is given as the result of a Berry-phase calculation for the whole supercell, which gives a value between those of the individual layers. It is understood that P in each case is the component of the polarization vector along the [0001] axis.

which is well defined modulo $e\mathbf{R}$ (where e is the elementary charge and $\mathbf{R}$ is a lattice vector). The latter ambiguity can be removed in different ways, such that a meaningful value for the polarization can be obtained [78, 79, 82]. However, the obtainment of a position-dependent polarization vector, that varies *within* the unit cell in which the Berry phase is computed, is beyond the reach of this technique. Nevertheless, for systems where composition and/or strain change abruptly within the unit cell (e.g. random alloy InGaN QWs), the question of whether a local value of the polarization vector can be calculated becomes pertinent.

Consider, for instance, a superlattice made up of two different materials arranged in a periodic layered structure. For relevance to the materials under consideration in this chapter, in Fig. 5.1 we have constructed such a system with layers of strained GaN and InN, with the periodic unit cell being the supercell of the figure. If we choose the lattice parameter of the superlattice as the average between the values for GaN and InN, then the GaN layer would undergo tensile strain whereas the InN layer would be subjected to compressive strain. In such a case, it would seem reasonable to calculate the polarization in each of the layers separately as that due to the response of each of the corresponding materials, given their particular strain state. This is indicated by the sign of the polarization in each layer. However, in the context of the Berry-phase technique, only the average polarization of the periodic unit cell as a whole can be calculated *formally*. In such a calculation, the notion of a polarization discontinuity across the interface between the GaN and the InN layers is lost completely. Therefore any effect owing

to such discontinuity, like interfacial electric charge accumulation, of great importance in the case of nitride nanostructures [91], is lost. The problem in the case exemplified by Fig. 5.1 is easy to circumvent, in the sense that there is an obvious way to treat the GaN and InN layers separately and obtain the polarization for each layer. In a more general case, however, there must not necessarily be an obvious or straightforward way to partition the system into subsets in which the polarization can be easily calculated. Any knowledge of how the polarization varies within the supercell must therefore rely on a heuristic assumption. We are not troubled by depending on finding a phenomenological solution to the problem, insofar it gives access to physical information which would not be accessible otherwise. It will be shown that a position-dependent polarization, defined down to the unit volume of an ensemble of nearest-neighbours, yields results similar to a *formal* Berry-phase calculation, when extrapolated to calculate the average polarization of the supercell. We will take this as an indication that the local effects are indeed being computed correctly.

5.2.1 Formal definition of the local polarization

As already discussed, the total macroscopic polarization has two components: spontaneous and piezoelectric. Since the spontaneous polarization is a reference state, establishing a local value for it formally might prove rather non trivial: one would need to devise an adiabatic transformation which keeps the system insulating from the equivalent centrosymmetric to the polar crystal structure (Section 2.3) that allows to evaluate the difference in polarization locally (at each atomic site). Therefore, to avoid this complexity, in this first work we assume the spontaneous polarization for a given binary compound to be position-independent and direct our attention towards the piezoelectric polarization instead.

Our aim is a reformulation of Eq. (5.1) that allows an evaluation of the local and macroscopic contributions to the polarization separately. For the sake of clarity and conciseness, we constrain ourselves to changes in P_i^{pz} that are linear in the strains. Future work will extend our description to second-order piezoelectric polarization. As we will see later on, the linear approximation breaks down quickly for some III-Vs but is reasonably good up to moderate strain for the highly ionic III-nitrides. In analogy to elasticity [4], we can generalize Eq. (5.1) for arbitrary

internal strains as follows:

$$P_i^{\mathrm{pz}} = \sum_{j=1}^{6} e_{ij}\epsilon_j + \sum_{\alpha=1}^{N_{\mathrm{atoms}}} \sum_{k=1}^{3} \underbrace{\frac{\partial P_i^{\mathrm{pz}}}{\partial t_k^\alpha}}_{e\mathcal{Z}_{ik}^\alpha/V} \left[t_k^\alpha - t_{k,0}^\alpha(\boldsymbol{\epsilon})\right], \tag{5.3}$$

where N_{atoms} is the number of atoms in the sample, t_k^α is the k^{th} component of the internal strain vector for atom α, e is the elementary charge, V is the volume of the sample, and $\mathcal{Z}_{ik}^\alpha$ is the ik component of the Born effective charge tensor for atom α. $t_{k,0}^\alpha(\boldsymbol{\epsilon})$ are the internal strains that minimize the total energy of the crystal for any given strain state $\boldsymbol{\epsilon}$ [4]. Although Eq. (5.3) is general, because we are working in the linear approximation we will assume that the off-diagonal components of the Born effective charges are zero. Equation (5.3) therefore reduces to

$$P_i^{\mathrm{pz}} = \sum_{j=1}^{6} e_{ij}\epsilon_j + \sum_{\alpha=1}^{N_{\mathrm{atoms}}} \frac{e\mathcal{Z}_i^\alpha}{V} \left[t_i^\alpha - t_{i,0}^\alpha(\boldsymbol{\epsilon})\right], \tag{5.4}$$

where we have employed an implicit notation $\mathcal{Z}_i^\alpha \equiv \mathcal{Z}_{ii}^\alpha$. Again, in the linear limit, the $t_{i,0}^\alpha$ are linear in $\boldsymbol{\epsilon}$ and we can write

$$P_i^{\mathrm{pz}} = \sum_{j=1}^{6} \underbrace{\left(e_{ij} - \sum_{\alpha=1}^{N_{\mathrm{atoms}}} \frac{e\mathcal{Z}_i^\alpha}{V} \frac{\partial t_{i,0}^\alpha}{\partial \epsilon_j}\right)}_{e_{ij}^{(0)}} \epsilon_j + \sum_{\alpha=1}^{N_{\mathrm{atoms}}} \frac{e\mathcal{Z}_i^\alpha}{V} t_i^\alpha, \tag{5.5}$$

where $e_{ij}^{(0)}$ is the piezoelectric coefficient obtained from a "clamped-ion" calculation [82], in which the ionic coordinates are not allowed to relax. Note that in Eq. (5.5), the first term $e_{ij}^{(0)}$ is macroscopic, that is, defined for the unit cell as a whole, while the second one is evaluated locally.

Consider now that V_0 is the volume comprising an atomic site and all of its nearest neighbours (in the context of the four-fold coordinated ZB and WZ lattices this would correspond to each of the tetrahedra that make up the crystal). We label the central atomic site 0 and each of its nearest neighbours by $\alpha = 1, 2, 3, \ldots, N_{\mathrm{coor}}^0$. Then, the relevant quantity in Eq. (5.5) to be evaluated locally (at the atomic site

0) is

$$P^{\mathrm{pz}}_{i,\mathrm{local}}(0) \equiv \frac{e}{V_0}\left(\mathcal{Z}^0_i t^0_i + \sum_{\alpha=1}^{N^0_{\mathrm{coor}}} \frac{\mathcal{Z}^\alpha_i}{N^\alpha_{\mathrm{coor}}} t^\alpha_i\right), \tag{5.6}$$

where N^α_{coor} is the number of nearest neighbours of atom α. By dividing the contribution of each of the nearest neighbours $\mathcal{Z}^\alpha_i$ by their own number of nearest neighbours N^α_{coor} we ensure no double counting when extending the evaluation of Eq. (5.6) to the whole crystal.

The internal strains can be obtained relatively straightforwardly for binary compounds [3,4]. However, for an irregular material, such as an alloy, establishing a reference lattice structure with respect to which the internal strains could be calculated would carry a high degree of arbitrariness. Furthermore, an *exact* evaluation of Eq. (5.6) would rely on knowing the value of $\mathcal{Z}^\alpha_i$ for all the atoms present in the crystal. For an irregular material, $\mathcal{Z}^\alpha_i$ would differ, in general, for each atom, even (by a small amount) for atoms of the same species. Therefore, our choice is to deduce an approximation to Eq. (5.6) valid for a representative reference system (such as a binary), and use that approximation to estimate the local polarization in irregular systems. We propose the following spherical approximation for the local environment of the central atom (atomic site 0):

$$\sum_{\alpha=1}^{N^0_{\mathrm{coor}}} \frac{\mathcal{Z}^\alpha_i}{N^\alpha_{\mathrm{coor}}} t^\alpha_i \approx -\frac{\mathcal{Z}^0_i}{N^0_{\mathrm{coor}}} \sum_{\alpha=1}^{N^0_{\mathrm{coor}}} t^\alpha_i. \tag{5.7}$$

The approximation given by Eq. (5.7) would be exact if all the nearest neighbours ($\alpha = 1, 2, 3, \ldots, N^0_{\mathrm{coor}}$) of atom 0 were piezoelectrically equivalent, that is, if all of them have the same Born effective charges. This is the case for binary ZB and WZ compounds. Further on, we will deal with how different approximations work out for alloys.

We can characterize the bonds between atom 0 and atoms $\alpha = 1, 2, 3, \ldots, N^0_{\mathrm{coor}}$ by a vector $\boldsymbol{\ell}^\alpha$ as indicated in Fig. 5.2. If $\boldsymbol{\ell}^\alpha_0$ is the bond vector of the unstrained case, we can write $\boldsymbol{\ell}^\alpha$ in terms of the macroscopic and internal strains:

$$\ell^\alpha_i = \sum_{j=1}^{3} \left(\delta_{ij} + \epsilon_{ij}\right) \ell^\alpha_{j,0} + t^\alpha_i - t^0_i, \tag{5.8}$$

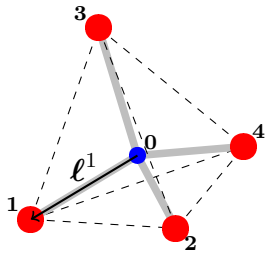


Figure 5.2: First nearest-neighbour environment in a tetrahedrally bonded crystal. The vector pointing from atom 0 (the central atom) towards atom α is denoted $\boldsymbol{\ell}^\alpha$.

where ϵ_{ij} are the components of the strain tensor in Cartesian notation and δ_{ij} is the Kronecker delta function. With the approximation of Eq. (5.7) and the definition given by Eq. (5.8) we rewrite Eq. (5.5) as

$$P_i^{\mathrm{pz}} = \sum_{j=1}^{6} e_{ij}^{(0)} \epsilon_j - \frac{e}{V_0} \frac{\mathcal{Z}_i^0}{N_{\mathrm{coor}}^0} \left(\underbrace{\sum_{\alpha=1}^{N_{\mathrm{coor}}^0} \ell_i^\alpha}_{\mu_i} - \sum_{j=1}^{3} (\delta_{ij} + \epsilon_{ij}) \underbrace{\sum_{\alpha=1}^{N_{\mathrm{coor}}^0} \ell_{j,0}^\alpha}_{\mu_{j,0}} \right), \tag{5.9}$$

where $\boldsymbol{\mu}$, defined as a summation over nearest-neighbour distances, is the *bond asymmetry parameter* [5]. $\boldsymbol{\mu}_0$ is the bond asymmetry parameter of the unstrained system, that would be zero for binary ZB materials and would have a non-zero component along the polar axis $\mu_{3,0}$ for WZ materials [5].

Finally, we write for the *total* polarization at atomic site 0:

$$P_i = \underbrace{\sum_{j=1}^{6} e_{ij}^{(0)} \epsilon_j}_{\text{macroscopic}} + \underbrace{P_i^{\mathrm{sp}} - \frac{e}{V_0} \frac{\mathcal{Z}_i^0}{N_{\mathrm{coor}}^0} \left(\mu_i - \sum_{j=1}^{3} (\delta_{ij} + \epsilon_{ij}) \mu_{j,0} \right)}_{\text{local}}. \tag{5.10}$$

Equation (5.10) is a central result of this chapter, which separates the contributions to the polarization arising from macroscopic effects, given by the clamped-ion piezoelectric coefficient $e_{ij}^{(0)}$, and local effects, dominated by internal strain.

5.2.2 Validity of the model

We have made a number of approximations in the previous section. Depending on the nature of the compound at hand, each of them will have a different impact on the results, and will limit the accuracy that can be achieved. These approximations

are:

1. We have assumed that $\mathbf{P}_{\mathrm{sp}}$ is constant throughout the crystal for binaries. However, we have defined it as a *local* quantity (this will prove helpful when dealing with alloys).

2. For the piezoelectric part, we have truncated our description to first order in both macroscopic and internal strain.

3. We have assumed that the off-diagonal terms of the Born effective charge tensor are zero.

4. We have performed a spherical approximation for the Born effective charge of the nearest neighbours of the atom where the local polarization is evaluated.

As discussed in Section 5.2.1, it is not trivial to establish wether approximation 1 is good or not. It is possible to separate the contributions to $\mathbf{P}_{\mathrm{sp}}$ into that arising from the *initial* bond asymmetry parameter $\boldsymbol{\mu}_0$ that we have defined previously (which in WZ is related to the internal parameter u), and the purely electronic contribution [5, 109, 163]. In this context, it is at least possible to assign a local value for the initial bond asymmetry contribution, which in the case of WZ would be equal in both cation and anion sites. It seems therefore that assuming the electronic part to be also constant between different atomic sites for the binaries might be reasonable.

Approximation 2 is indeed the main limitation to this model, but possibly the most straightforward one to overcome. This theory can be extended to include second-order piezoelectric effects at the expense of complicating the formulas. We opt here to limit ourselves to a first-order description to emphasize the conceptual implications of the theory. The linear limit should be valid for highly ionic compounds, such as group-III nitrides, as will be shown in the next section. For the nitrides, although the second-order effects are large, the first-order terms dominate up to strain values that are typically found in realistic alloys and heterostructures (up to 5%) [1, 89, 112]. For the other III-Vs, however, second-order piezoelectric coefficients are relatively much larger if compared with the linear ones. For instance, for the Al compounds AlP, AlAs and AlSb, Beya-Wakata *et al.* [58] found that the first-order piezoelectricity can practically be neglected and second-order effects dominate even for small strains. For the technologically important GaAs

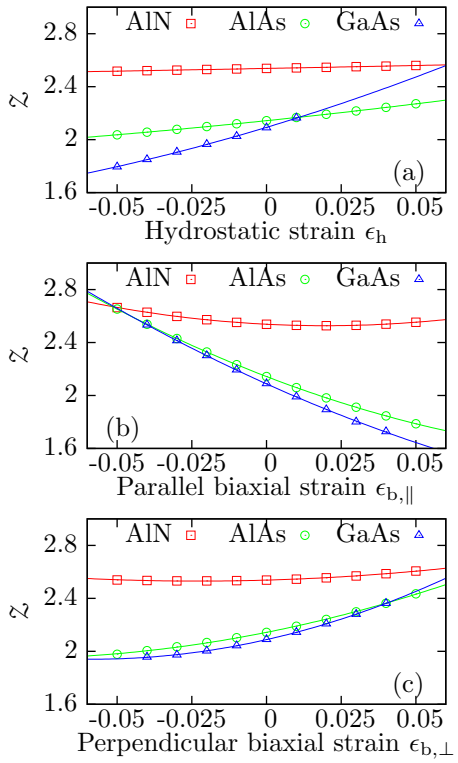


Figure 5.3: Born effective charges of the corresponding cation for ZB AlN, AlAs and GaAs, as a function of (a) hydrostatic and (b–c) biaxial strain. "Parallel biaxial strain" means that the inequivalent strain axis coincides with the axis along which the Born effective charge is calculated, that is $\epsilon_1 = \epsilon_2 = \epsilon_{b,\parallel}$, $\epsilon_3 = -2\epsilon_{b,\parallel}$ and $\mathcal{Z} \equiv \mathcal{Z}_3$. "Perpendicular biaxial strain" refers to the opposite situation: $\epsilon_2 = \epsilon_3 = \epsilon_{b,\perp}$, $\epsilon_1 = -2\epsilon_{b,\perp}$ and $\mathcal{Z} \equiv \mathcal{Z}_3$. Open symbols are the results of LDA-DFT calculations (see Section 5.2.3 for details) while solid lines are quadratic fits to the data. The missing points for GaAs within this strain range cannot be calculated because the LDA predicts a conducting state which is not compatible with the Berry phase formalism (see discussion for InN in Section 5.2.3) [78, 79].

the situation is intermediate and the present level of theory should be accurate for small strains below 1 or 2 %. This complication is also present when computing the Born effective charges. As we show in Fig. 5.3 for the hydrostatic and biaxial strain dependence of $\mathcal{Z}$ (see figure caption and next section for details of the calculation), the linear approximation for the Born effective charge gets worse as one moves from the highly ionic AlN to the less ionic GaAs and AlAs. Note that strain-dependent Born effective charges also have an impact on the clamped-ion piezoelectric coefficient, as given by Eq. (5.5). Therefore, a more complete and accurate treatment for general materials should eventually include the dependence of the $\mathcal{Z}_i$ with strain.

Approximation 3 is generally good, since for binary compounds the off-diagonal components of the Born effective charge are typically zero, and in any case the ratio $\mathcal{Z}_{ij}/\mathcal{Z}_{ii}(i \neq j)$ is usually small.

The validity of approximation 4 relies greatly on the specific crystalline structure and whether the nearest neighbours of the central atom where the polarization is being calculated are equivalent (that is, have the same Born effective charge) or not. For this reason, in the case of binary tetrahedrally bonded compounds, where all the nearest neighbours for one given site are of the same atomic species, this approximation should be good for small strains. As observed in Fig. 5.3 for biaxial strain, lattice distortions that change the symmetry of the bonds have a large impact on the Born effective charge for some compounds. Therefore, the validity of Eq. (5.10) would be limited for low ionicity and the more general form, Eq. (5.5), should be used. On the other hand, for ionic compounds such as nitrides, Eq. (5.10) retains its validity and offers an accurate description of the local effects, as will be shown in Section 5.2.3. In both cases (low and high ionicity in tetrahedrally bonded binaries) the approximation is exact for the linear piezoelectric limit (see Section 5.2.3).

5.2.3 Testing the theory for group-III nitrides

As a first validation test and application of our theory, we have chosen group-III nitrides. The large difference in bond lengths between the nitride binaries leads to considerable local strains in these alloys, with measurable effects such as large band gap bowings [21,136]. We have previously shown how these local strain fields affect the electric polarization for InGaN alloys, retrieving the macroscopic limit with the advantage of giving a description of the local effects at the same time [5]. We have now presented in Section 5.2 a refined and more general form of that model. In the following, we will thoroughly apply this theory to test its validity for the III-N.

Parameters involved in the calculation of the local polarization

The first step is to derive the necessary parameters for the WZ III-N binaries GaN, AlN and InN: piezoelectric tensor e_{ij}, spontaneous polarization P_i^{sp}, Born effective charges $\mathcal{Z}_i$, lattice parameters a_0 and c_0, internal parameter u_0, and internal strain parameters ζ_i. For our calculations we have used the plane wave implementation

Table 5.1: Parameters involved in the calculation of polarization-related quantities for WZ group-III nitrides, obtained from DFT calculations as explained throughout the text. The HSE lattice parameters a_0, c_0, internal parameter u_0, and internal strain parameters ζ_i are taken from Ref. [3] (Chapter 4). The **k** grids are $6\times6\times4$ Γ-centred for a four-atom hexagonal cell in all cases except for the calculation of e_{ij}, $e_{ij}^{(0)}$, P^{sp} and $\mathcal{Z}_i$ for InN. For those quantities we use an orthorhombic-equivalent 16-atom supercell and the sampling in k space is $4\times4\times4$, following the standard Monkhorst scheme implemented in VASP, which does not include Γ in the integration (see text for details) [72]. Note that in all cases, the positive sign for $\mathcal{Z}_i$ implies a displacement of the cation sublattice: the corresponding Born effective charge of the anions is $-|\mathcal{Z}_i|$. $P^{\mathrm{sp}}_{\mathrm{idWZ}}$ is the spontaneous polarization of the ideal WZ lattice (lattice parameters and internal parameter extrapolated from the ZB phase).

	GaN		InN		AlN	
	HSE	LDA	HSE	LDA	HSE	LDA
a_0 (Å)	3.180	3.154	3.542	3.507	3.103	3.092
c_0 (Å)	5.172	5.141	5.711	5.668	4.970	4.947
u_0	0.3772	0.3765	0.3796	0.3787	0.3818	0.3820
ζ_1	0.156	0.168	0.193	0.204	0.138	0.145
ζ_2	0.083	0.089	0.107	0.112	0.086	0.091
ζ_3	0.159	0.168	0.218	0.226	0.191	0.200
ζ_4	0.201	0.210	0.337	0.339	0.199	0.224
ζ_5	0.141	0.148	0.107	0.118	0.143	0.140
e_{15} (C/m^2)	-0.32	-0.36	-0.42	-0.47	-0.39	-0.43
e_{31} (C/m^2)	-0.44	-0.49	-0.58	-0.63	-0.63	-0.69
e_{33} (C/m^2)	0.74	0.83	1.07	1.09	1.46	1.59
P_3^{sp} (C/m^2)	-0.040	-0.029	-0.049	-0.041	-0.091	-0.096
$P^{\mathrm{sp}}_{3,\mathrm{idWZ}}$ (C/m^2)	-0.019	-0.016	-0.019	-0.016	-0.031	-0.033
$e_{15}^{(0)}$ (C/m^2)	0.43	0.45	0.39	0.35	0.28	0.28
$e_{31}^{(0)}$ (C/m^2)	0.40	0.41	0.37	0.38	0.26	0.25
$e_{33}^{(0)}$ (C/m^2)	-0.87	-0.87	-0.87	-0.95	-0.51	-0.47
$\mathcal{Z}_1(=\mathcal{Z}_2)$	2.64	2.58	2.85	2.83	2.53	2.52
$\mathcal{Z}_3$	2.77	2.72	3.02	3.00	2.68	2.67

of density functional theory (DFT) available from the VASP package [72,75], within the projector augmented-wave (PAW) method [68, 74]. We perform calculations using both local density approximation (LDA) and the Heyd-Scuseria-Ernzerhof (HSE) screened-exchange hybrid functional [3, 62, 63]. For the LDA we use VASP's implementation of the Perdew-Zunger parametrization [53], while the settings for the HSE functional correspond to HSE06, with mixing parameter $\alpha = 0.25$ and screening parameter $\mu = 0.2$. In all calculations the cutoff energy for plane waves is 600 eV. All the quantities involving a calculation of the polarization have been obtained using Martijn Marsman's implementation of the Berry phase technique [79] available in VASP. We use HSE to obtain high quality parameters for the binaries and LDA to perform test calculations for larger supercells and for statistical evaluation of the accuracy of the theory. In our experience, LDA-DFT gives a relatively good description of elastic properties and internal strain, while at the same time being computationally affordable. Also, LDA-DFT seems to give results in better agreement with experiments than generalized-gradient approximations (GGAs) for the calculated electric polarization, at least for the III-Vs [58]. The more computationally demanding HSE functional, on the other hand, reduces the band gap problem existent in standard Kohn-Sham DFT [64], that potentially leads to a conducting phase being incorrectly predicted for narrow gap semiconductors, such as InN. HSE also provides lattice parameters and elastic properties in better agreement with experiment [3].

We have calculated all the relevant structural and polarization-related parameters, which are summarized in Table 5.1. In the context of the Berry phase approach, a meaningful value for the polarization can only be calculated if the system remains insulating [78, 79, 81]. As already discussed, in the case of the III-N this is not a problem for the HSE functional, which predicts a positive gap [113]. In the LDA, AlN and GaN are predicted to have (underestimated) positive gaps. However, our settings lead to the prediction of a band crossing at the Γ point for InN, and therefore an incorrect metallic phase that renders the calculation of a meaningful value of the polarization uncertain. Previous data have been given for InN by Fiorentini and collaborators in a series of papers on the piezoelectric properties and spontaneous polarization of group-III nitrides [57, 82, 109]. While their LDA calculations obtain the correct insulating phase of InN,[3] ours must rely

[3]Private communication with V. Fiorentini and D. Vanderbilt.

on a different approach: since the band crossing occurs only at the Γ point and immediate surroundings, we skip this area in the k-point integration by shifting the $\mathbf{k}$ mesh away from Γ. The resulting values of the polarization-related quantities in Table 5.1 for the LDA show almost perfect agreement with Fiorentini *et al.*'s LDA data [57, 109], although InN remains technically a metal in our case. The good agreement with the HSE calculation further supports that our LDA values should be correct.

Local piezoelectric tensor

To obtain the relation between piezoelectric coefficients e_{ij} and internal strain parameters ζ_i, one can apply Eq. (5.5) to the internal strain vectors for the WZ geometry. We have previously obtained the relation between macroscopic and internal strain for the WZ lattice and provided the definition of the WZ internal strain parameters ζ_i in Ref. [3] (Chapter 4). The results can conveniently be expressed in the following compact form:

$$\begin{aligned} e_{15} &= e_{15}^{(0)} - \frac{2e\mathcal{Z}_1}{\sqrt{3}{a_0}^2}\zeta_1, \\ e_{31} &= e_{31}^{(0)} - \frac{4e\mathcal{Z}_3}{\sqrt{3}{a_0}^2}\zeta_2, \\ e_{33} &= e_{33}^{(0)} + \frac{4e\mathcal{Z}_3}{\sqrt{3}{a_0}^2}\zeta_3. \end{aligned} \tag{5.11}$$

We have incorporated in Eq. (5.11) none of the assumptions leading to Eq. (5.10). Therefore, Eq. (5.11) is an exact result for WZ in linear piezoelectricity. It is thus surprising that ζ_4 and ζ_5, although breaking the cell symmetries, do not appear in the expressions for the e_{ij}. The reason for this will become clear when obtaining the e_{ij} as $\frac{\partial P_i}{\partial \epsilon_j}$, calculated from Eq. (5.10). Therefore, following the convention of Fig. 5.4, the *local* piezoelectric tensor, notated e_{ij}^*, can be calculated at the atomic sites A and C, corresponding to the two cations present in the unit cell, as the derivative of Eq. (5.10) with respect to the strains:

$$e_{ij}^{*,X} = e_{ij}^{(0)} - \frac{e\mathcal{Z}_i^X}{\sqrt{3}{a_0}^2 c_0}\left(\frac{\partial \mu_i^X}{\partial \epsilon_j} - \sum_{k=1}^{3}\frac{\partial \epsilon_{ik}}{\partial \epsilon_j}\mu_{k,0}^X\right), \tag{5.12}$$

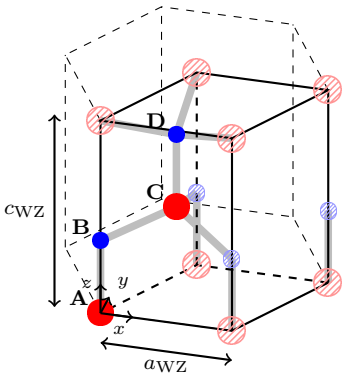


Figure 5.4: Standard four-atom WZ unit cell. A and C are cations, B and D are anions.

where X indicates A or C. For WZ, the only non-zero $\mu_{k,0}$ is $\mu_{3,0} = 4(u_0 - 3/8)c_0$ [5]. Expressing $\boldsymbol{\mu}$ in terms of macroscopic strains, lattice parameters and internal strain parameters, each of the non-zero components of e^*_{ij} can be obtained:[4]

$$
\begin{aligned}
e^{*,\mathrm{A}}_{15} &= e^{*,\mathrm{C}}_{15} = e^{(0)}_{15} - \frac{2e\mathcal{Z}_1}{\sqrt{3}{a_0}^2}\zeta_1, \\
e^{*,\mathrm{A}}_{16} &= -e^{*,\mathrm{C}}_{16} = \frac{\sqrt{3}e\mathcal{Z}_1}{2a_0c_0}\zeta_4 + \frac{e\mathcal{Z}_1}{\sqrt{3}{a_0}^2}\zeta_5, \\
e^{*,\mathrm{A}}_{21} &= -e^{*,\mathrm{C}}_{21} = e^{*,\mathrm{A}}_{16}, \\
e^{*,\mathrm{A}}_{22} &= -e^{*,\mathrm{C}}_{22} = -e^{*,\mathrm{A}}_{16}, \\
e^{*,\mathrm{A}}_{31} &= e^{*,\mathrm{C}}_{31} = e^{(0)}_{31} - \frac{4e\mathcal{Z}_3}{\sqrt{3}{a_0}^2}\zeta_2, \\
e^{*,\mathrm{A}}_{33} &= e^{*,\mathrm{C}}_{33} = e^{(0)}_{33} + \frac{4e\mathcal{Z}_3}{\sqrt{3}{a_0}^2}\zeta_3.
\end{aligned}
\tag{5.13}
$$

That is, the expressions for e_{15}, e_{31} and e_{33} are retrieved exactly, but additional piezoelectric components appear, that change sign going from A to C. To elucidate the effect of this on the symmetry of the piezoelectric tensor, we write e^*_{ij} in matrix form:

$$
e^{*,\mathrm{A/C}}_{ij} \equiv \begin{pmatrix}
0 & 0 & 0 & 0 & e_{15} & \pm e^*_{16} \\
\pm e^*_{16} & \mp e^*_{16} & 0 & e_{15} & 0 & 0 \\
e_{31} & e_{31} & e_{33} & 0 & 0 & 0
\end{pmatrix}. \tag{5.14}
$$

[4] An example calculation for $e^{*,\mathrm{A}}_{15}$ is given in Appendix A.1.

When averaging $e_{ij}^{*,\mathrm{A}}$ and $e_{ij}^{*,\mathrm{C}}$ one retrieves the WZ macroscopic limit:

$$\frac{1}{2}\left(e_{ij}^{*,\mathrm{A}}+e_{ij}^{*,\mathrm{C}}\right)\equiv\begin{pmatrix}0&0&0&0&e_{15}&0\\0&0&0&e_{15}&0&0\\e_{31}&e_{31}&e_{33}&0&0&0\end{pmatrix}. \tag{5.15}$$

The anion sites B and D have the same expressions for e_{15}, e_{31} and e_{33} and slightly different expressions for e_{16}^{*}:

$$e_{16}^{*,\mathrm{B}}=-e_{16}^{*,\mathrm{D}}=-\frac{\sqrt{3}e\mathcal{Z}_1}{2a_0c_0}\zeta_4-\frac{2e\mathcal{Z}_1}{\sqrt{3}{a_0}^2}\zeta_5. \tag{5.16}$$

The macroscopic limit is of course also retrieved when averaging for the anion sites. Note that the values of e_{16}^{*} are comparable to those of the macroscopic piezoelectric tensor. For instance, for GaN, $|e_{16}^{*}|$ amounts to 0.79 C/m^2 and 1.13 C/m^2 for cation and anion sites, respectively.

Equation (5.14) is the (site-dependent) local piezoelectric tensor of the WZ lattice. It reflects the fact that there exist two sets of inequivalent tetrahedra in a WZ lattice, and that the macroscopic strain affects the nearest-neighbour environment of each of them differently [3, 4]. This is *a priori* an unexpected result, that would have the implication that crystals that are non-polar and non-piezoelectric *on average* could present a local, perhaps measurable, piezoelectric-like polarization.

Finally, note the very interesting similarity between the local piezoelectric tensor of WZ and that of ZB in a (111)-oriented description [Eq. (27) of Ref. [6]]. The (111)-oriented ZB systems present three-fold symmetry which is not present in WZ [6].

Local polarization in InGaN alloys: strategies and testing

We have seen so far that for wurtzite nitride binaries there is an exact correspondence between local and macroscopic polarization that is retrieved when averaging the local part over the unit cell. Although some solid-state devices might operate employing binary compounds, the most interesting applications of the nitrides arise through the use of their alloys and the corresponding property tunability (e.g. of band gap).

The main problem associated to performing a local polarization calculation for

an alloy is the increased complexity of the atomic environment of each of the sites where the local polarization is to be evaluated. This is due to the fact that the Born effective charges of all the atoms involved in the calculation are affected by the interaction with all the other atoms present in the crystal. In a periodic cell calculation this number would be reduced to the number of atoms in the supercell. Since there is an arbitrarily large number of possible configurations depending on alloy composition and supercell size, establishing an exact correspondence between local and macroscopic polarization in the fashion of Section 5.2.1 becomes virtually impossible. To overcome this limitation, we will assume for the nitrides, and InGaN in particular, that the Born effective charge of the cations in the alloy remains the same as for the binary, and that the spherical approximation still holds [5]. We have devised two tests in order to establish how good this approximation is. First, we will use the smallest alloy cell, which is a CuPt-like (CP-like) InGaN unit cell consisting only of 4 atoms [5, 109], and will perform random distortions of the atomic positions within the unit cell. The result of the averaged local polarization, calculated using Eq. (5.10), will be compared to the formal Berry-phase result. Second, 32- and 128-atom $In_{0.5}Ga_{0.5}N$ supercells will be considered and the cation sites occupied randomly with either a Ga or an In atom, with the only requirement that the stoichiometric ratio of 1/1 be preserved (i.e. the nominal composition of all cells is the same). The internal atomic positions will then be allowed to relax by minimizing the supercell LDA-DFT total energy, and the result of the averaged local polarization will again be compared to that of a Berry-phase calculation. The statistical treatment of both tests will reveal the validity of the approximation for InGaN alloys.

The results of the first test are depicted in Fig. 5.5. The figure shows a comparison of the average polarization of the CP-like InGaN cell calculated both within the present local polarization model and with the Berry-phase technique. We have performed random displacements of up to ± 0.2 Å (which is equivalent to approximately a 10% of the equilibrium bond lengths) to each of the Cartesian coordinates of each of the 4 atoms in the unit cell. For the local polarization model, we have computed the local polarization contributions at the Ga and In sites using Eq. (5.10) and then obtained its average for the whole cell. Since only differences in polarization are meaningful within the Berry-phase formalism [78, 79], we compare in Fig. 5.5 the difference ΔP between the polarization of the equilibrium CP-like InGaN structure and the distorted one. As can be seen, the agreement

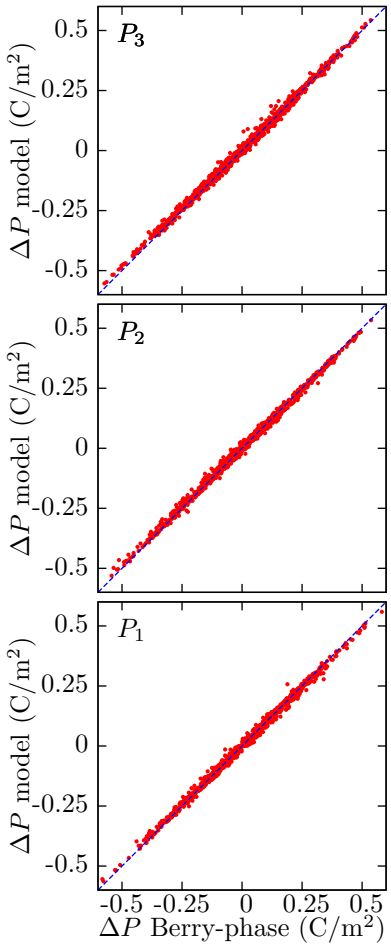


Figure 5.5: Comparison between the polarization predicted by the present model and a Berry-phase calculation for a large number (1000) of randomly distorted CP-like InGaN cells (four-atom unit cell). ΔP is the difference in polarization between the equilibrium and distorted structures, where the lattice vectors are fixed but the coordinates of each atom in the unit cell are varied randomly up to ± 0.2 Å in each Cartesian direction. The Berry-phase values are LDA-DFT results. The dashed line indicates perfect agreement between the two methods, that is $\Delta P_{\text{model}} = \Delta P_{\text{Berry-phase}}$. A few random distortions within the range lead to a metallic phase being predicted by LDA, and were left out of the comparison.

between the two methods is remarkable, with all the data points lining up against the dashed line that corresponds to perfect agreement $\Delta P_{\text{model}} = \Delta P_{\text{Berry-phase}}$.

Even more enlightening is the comparison between the present model and the Berry-phase results depicted in Fig. 5.6 for random $In_{0.5}Ga_{0.5}N$ orthorhombic supercells. In that figure, ΔP is the difference between the polarization of the supercell before and after internal strain relaxation. The supercells are constructed with either 32 or 128 atoms and the In and Ga atoms are placed randomly at the cation sites. The lattice vectors of the supercells are kept fixed and chosen as the average between the LDA values for the binaries. The "site count" panels show the number of cation sites that present a particular *local* polarization value within different ranges, for the combined supercells. We note two main features. The first observation is that the local polarization model succeeds at very accurately predicting the average supercell polarization even though the latter is calculated from a sum over many local contributions whose values vary within limits one order of magnitude higher. Second, our results show that the average polarization is highly dependent on the specific atomic arrangement, even for a large number of atoms. Bernardini and Fiorentini [109] have previously calculated the spontaneous polarization for the same material using a 32-atom special quasirandom structure (SQS) [164], and have proposed that disorder plays only a secondary role in the calculation of the polarization, both spontaneous and piezoelectric [89, 107, 109, 112]. We have found that this is indeed the case for the spontaneous polarization of the supercells studied *before* the optimization of the atomic degrees of freedom: all the 128-atom configurations studied yielded the same value of ~ -0.009 C/m^2 within less than 0.001 C/m^2 of each other. However, our results suggest i) that a 32-atom supercell might not be large enough to study the effect of disorder (see e.g. clustering of calculated values for P_2 in Fig. 5.6) and ii) that internal strain relaxation introduces large corrections to the polarization value, even for supercells containing as many as 128 atoms. Note, for instance, that the average in-plane components of the polarization P_1 and P_2, which are not symmetry-allowed for the binaries, do not vanish for the alloys in the case of finite-size supercells.

All of these considerations do not only support the validity of the local model discussed here, but also highlight the need for one, in order to be able to treat the effects of disorder and associated internal strain accurately.

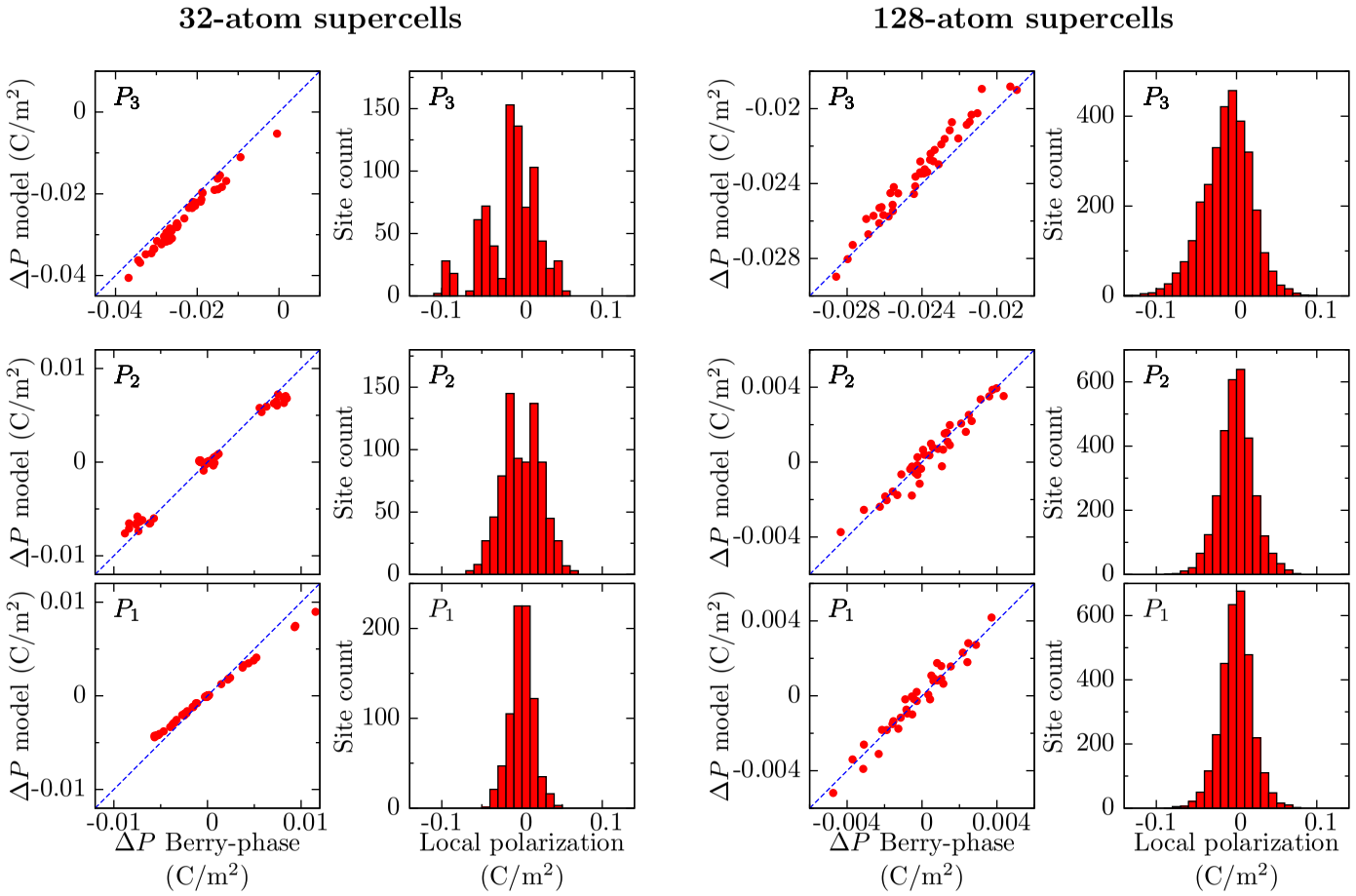


Figure 5.6: Comparison between the spontaneous polarization calculated using the present local model and the Berry-phase technique for a series of $In_{0.5}Ga_{0.5}N$ random supercells with 32 and 128 atoms. ΔP is the difference between the polarization of the supercells before and after internal strain relaxation. The Berry-phase values and the relaxed atomic positions are LDA-DFT results. The dashed line indicates perfect agreement between the two methods, that is $\Delta P_{\text{model}} = \Delta P_{\text{Berry-phase}}$. The supercells are orthorhombic, with the lattice vectors given by the average of the InN and GaN lattice parameters. In terms of cation layers, the 32-atom supercells have a $2 \times 4 \times 2$ arrangement while the 128-atom supercells are $4 \times 4 \times 4$. The "site count" panels for each series refers to the number of cation sites that registered a local polarization value within the ranges shown (of width 0.01 C/m^2), for the combined supercells.

5.3 Point dipole method for the calculation of the polarization potential

The point dipole model is a solution to the challenge of solving Poisson's equation on an atomic grid where abrupt changes in the polarization vector occur [5]. However, before the polarization potential can be obtained from the point dipoles, a remapping of polarization density into dipole moment on the system's grid has to be performed. The latter is dealt with in Section 5.3.1. The general solution for the polarization potential arising from the ensemble of point dipoles is obtained in Section 5.3.2 in an image dipole scheme, for a QW system (or layered structure, in general) where a different arbitrary dielectric constant is allowed for all three neighbouring layers of material. The effect of different levels of approximation for this general solution is also treated in Appendix A. In Section 5.3.3 we present a comparison between the solution of Poisson's equation for a problem with an available analytical solution and different levels of implementation of our method. Finally, in Section 5.3.4 we discuss computational aspects. Further material complementary to each of the sections is given in Appendix A.

5.3.1 From polarization to dipole moment

Before establishing the form of the potential due to a point dipole ensemble, we focus our attention on the transformation between polarization density $\mathbf{P}$, which is the quantity usually calculated in strained crystals, and dipole moment $\mathbf{p}$, which is the quantity involved in the equations that will be presented in the next section.

The polarization $\mathbf{P}$ can be understood as a "density of dipole moment". Indeed, the total dipole moment of a finite size sample in which the polarization density is constant is simply the product of $\mathbf{P}$ and the volume of the sample. Therefore, when dealing with constant polarization in a continuum-based description, a standard cubic discretization of the material, with step size Δ, is well suited to the representation of $\mathbf{P}$ as an ensemble of dipoles of magnitude $\mathbf{p} = \mathbf{P}\Delta^3$ located at each of the mesh points. However, our main interest is the representation of the material as an ensemble of point dipoles in an atomistic scheme. For tetrahedrally bonded compounds this involves the discretization in a mesh with either cubic (zinc-blende) or hexagonal (wurtzite) coordination, in the ideally undistorted lattice. After strain is applied, the former grids will suffer a deviation from cubic and

hexagonal symmetries and the assignment of a finite volume to each mesh point becomes cumbersome.

In the description of local polarization that we have previously employed, the values of $\mathbf{P}$ were given at the sites of each of the cations present in the crystal [5]. The latter is a useful description, in the sense that the representation of the whole crystal as a collection of deformed tetrahedra can be done via the relative positioning of the nearest neighbours: each cation and its four neighbouring anions unambiguously define each tetrahedron. Labelling the anions immediately surrounding a cation as 1, 2, 3 and 4 (Fig. 5.2), we refer to the volume of the corresponding tetrahedron as V_{1234}. If the positions of the anions are $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{r}_3$ and $\mathbf{r}_4$, then V_{1234} is given by

$$V_{1234} = \frac{1}{6}|\,(\mathbf{r}_1 - \mathbf{r}_4)\cdot[(\mathbf{r}_2 - \mathbf{r}_4)\times(\mathbf{r}_3 - \mathbf{r}_4)]\,|. \tag{5.17}$$

However, it can be easily shown that V_{1234} only accounts for the volume of the tetrahedron itself and that a summation of the volumes of all the tetrahedra contained within a material sample would underestimate the volume of the sample by exactly a factor of 6. Therefore, we define the volume *corresponding* to a tetrahedron as

$$\tilde{V}_{1234} = 6V_{1234}. \tag{5.18}$$

Now, the value of the dipoles can be easily obtained once a map of the polarization is available. For simplicity, we denote each grid point by i and the volume of the corresponding tetrahedron, as given by Eq. (5.18), as $\tilde{V}_i$:

$$\mathbf{p}_i = \mathbf{P}_i\tilde{V}_i, \tag{5.19}$$

being $\mathbf{p}_i$ located at the position $\mathbf{r}_i$ of cation i.

Our choice for a cation-based description stems from convenience. In a nitride alloy all the anions are nitrogen atoms and therefore applying the spherical approximation of Eq. (5.7) (which is based on nearest neighbours only) leads to one Born effective charge definition per cation atomic species: Ga, In and Al for conventional III-N. Using an anion-based description would lead, in the case of nitrides, to defining 15 different Born effective charges for N, which correspond to the 15 possible combinations of Ga/In/Al atoms that can be nearest neighbours

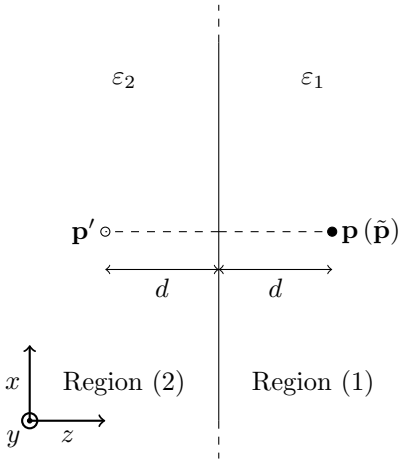


Figure 5.7: Schematic representation of the two media with different dielectric constant and point dipole $\mathbf{p}$ problem. The image dipoles $\mathbf{p}'$ and $\tilde{\mathbf{p}}$ are needed in order to solve it.

to N (e.g, 4 Ga, 3 Ga and 1 In, 2 Al and 2 In, etc.).

5.3.2 Solution for materials with different dielectric constant

Given the multipole expansion of a distribution of electric charge (see, for example, Ref. [45]), the contribution to the electrostatic potential $\phi_{\mathbf{p}}$ calculated at $\mathbf{r}$ due to a point dipole $\mathbf{p}$ is given by

$$\phi_{\mathbf{p}}(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0\varepsilon_r}\frac{\mathbf{p}\cdot(\mathbf{r}-\mathbf{r}_{\mathbf{p}})}{|\mathbf{r}-\mathbf{r}_{\mathbf{p}}|^3}, \tag{5.20}$$

where $\mathbf{r}_{\mathbf{p}}$ is the position of the dipole $\mathbf{p}$, ε_0 is the permittivity of the vaccum, and ε_r is the dielectric constant of the material. Eq. (5.20) is only valid when both the dipole $\mathbf{p}$ at $\mathbf{r}_{\mathbf{p}}$ and the point $\mathbf{r}$ where the potential is calculated are contained within an infinite (or big enough to neglect surface effects) sample of a dielectric material with dielectric constant ε_r. For the more general case in which there are boundaries between materials with different dielectric constants, e.g. a quantum well, it is appropriate to use the method of images to obtain a form of Eq. (5.20) that accounts for the discontinuity of ε_r across the different interfaces. Proceeding in a similar manner to the one employed by Jackson for a point charge [45], we can obtain the exact analytic solution for the potential due to a point dipole when only one interface is present, as schematically shown in Fig. 5.7:

$$\phi_{\mathbf{p}}^{(1)}(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0\varepsilon_1}\frac{\mathbf{p}\cdot(\mathbf{r}-\mathbf{r}_{\mathbf{p}})}{|\mathbf{r}-\mathbf{r}_{\mathbf{p}}|^3} + \frac{1}{4\pi\varepsilon_0\varepsilon_1}\frac{\mathbf{p}'\cdot(\mathbf{r}-\mathbf{r}_{\mathbf{p}'})}{|\mathbf{r}-\mathbf{r}_{\mathbf{p}'}|^3},$$
$$\phi_{\mathbf{p}}^{(2)}(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0\varepsilon_2}\frac{\tilde{\mathbf{p}}\cdot(\mathbf{r}-\mathbf{r}_{\tilde{\mathbf{p}}})}{|\mathbf{r}-\mathbf{r}_{\tilde{\mathbf{p}}}|^3}, \tag{5.21}$$

with

$$\mathbf{p}' = \frac{\varepsilon_1-\varepsilon_2}{\varepsilon_1+\varepsilon_2}[p_x, p_y, -p_z], \qquad \mathbf{r}_{\mathbf{p}'} = [x_{\mathbf{p}}, y_{\mathbf{p}}, z_{\mathbf{p}}-2d],$$
$$\tilde{\mathbf{p}} = \frac{2\varepsilon_2}{\varepsilon_1+\varepsilon_2}[p_x, p_y, p_z], \qquad \mathbf{r}_{\tilde{\mathbf{p}}} = \mathbf{r}_{\mathbf{p}}, \tag{5.22}$$

where $\mathbf{p}'$ is the image dipole used, together with the original dipole $\mathbf{p}$, for the calculation of the potential $\phi_{\mathbf{p}}^{(1)}(\mathbf{r})$ in region (1) and $\tilde{\mathbf{p}}$ is the image dipole used for the calculation of the potential $\phi_{\mathbf{p}}^{(2)}(\mathbf{r})$ in region (2). Their positions are given by $\mathbf{r}_{\mathbf{p}'}$ and $\mathbf{r}_{\tilde{\mathbf{p}}}$, respectively. The results for a test dipole of arbitrary magnitude when one of the materials has a dielectric constant twice as big as that of the material in which the dipole is contained are shown in Fig. 5.8(a–c) for three different orientations of the dipole.

The calculation of the potential when a second interface is included is far more complicated, as additional mirror images have to be added to balance the two initial image dipoles about each interface. As a result, an infinite number of reflections (and hence, image dipoles) have to be considered in order to obtain the exact form of the potential. These reflections up to third order are shown in Fig. 1 of Appendix A. The treatment for a point charge in such a situation has been already done by Barrera [165]. For the case of a point dipole, we find the expressions to be similar although the transformation of the point dipole is somehow different compared to the point charge due to the vector nature of the former. Details of our treatment and expressions for the three-media case are given in Section A.2.

5.3.3 Comparison to the solution of Poisson's equation for simple structures

Before applying the model to calculate the local polarization potential in realistic structures, it is necessary to test its accuracy against well established methods. An excellent test is the calculation of the polarization potential in a capacitor-like structure. In such an example, a layer of dielectric material (1) of thickness h, in which the polarization $\mathbf{P} = P_0\,\hat{\mathbf{z}}$ (where $\hat{\mathbf{z}}$ is a unit vector along the z axis) is

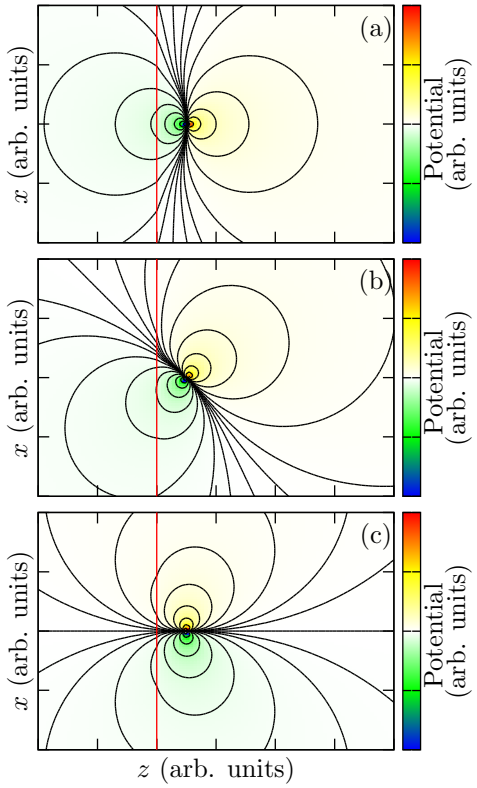


Figure 5.8: Potential profiles for three dipole orientations in the case of only one planar interface (indicated by the red line) and two different dielectric constants. The potential isolines are chosen so they decay exponentially.

constant and perpendicular to the neighbouring interfaces, is surrounded by two infinite layers of a dielectric material (2), with a different dielectric constant, in which the polarization is zero. An exact analytical solution to Poisson's equation can be obtained for the latter case. If we assume the first interface is located at $z = 0$, the potential is given by

$$\phi(\mathbf{r}) = \frac{P_0}{2\varepsilon_0 \varepsilon_r^{(1)}} \left(|z| - |z - h| \right), \tag{5.23}$$

where $\varepsilon_r^{(1)}$ is the dielectric constant of material (1). Figure 5.9(a) shows the potential profile as calculated exactly and analytically using Eq. (5.23) for the special case in which $P_0 = 0.1$ C/m^2, $\varepsilon_r^{(1)} = 8.4$ and $h = 30$ Å, which would be typically the situation in an InGaN QW surrounded by GaN barriers in which, for simplicity, the polarization has been switched off in the barriers. Within this simplified continuum picture, a spatial discretization of the current problem in a cubic grid

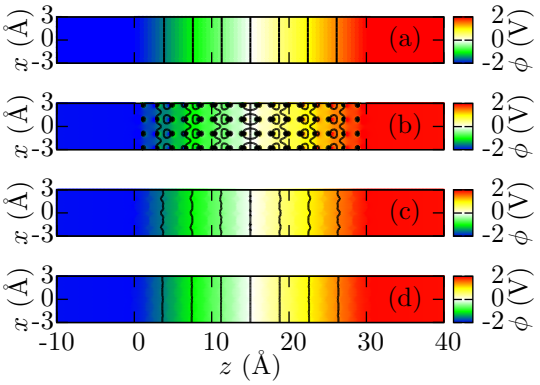


Figure 5.9: Potential obtained at different levels of approximation for a QW of width $h = 30$ Å for which $P_0 = 0.1$ C/m^2 and $\varepsilon_r = 8.4$. In the barrier $P = 0$ and $\varepsilon_r = 9.6$. (a) Analytic solution [Eq. (5.23)], (b) direct application of the present dipole method, (c) dipole method with cutoff radius $r_{\text{cutoff}} = 1$ Å and (d) dipole method with Gaussian smearing implementation, $r_{\text{smear}} = 1.5$ Å and $\sigma = 1$ Å (see Appendix A).

of steps $\Delta \approx 2$ Å, as discussed in the previous section, creates an ensemble of point dipoles which are of similar size to the ones encountered in typical InGaN QW situations. The application of our dipole method to first order reflections (see Appendix A) leads to a potential profile as in Fig. 5.9(b). In that figure, it can be observed how the potential changes brusquely in the surroundings of the dipoles (the plane of the figure has been deliberately chosen to be one that contains dipoles in it to dramatize this effect). This is due to the fact that Eq. (5.20) is a valid solution for a distribution of charge only if the position where the potential is calculated is sufficiently far away from the location of the point dipole that represents that distribution. We acknowledged this limitation in our previous work and proposed a cutoff radius around $\mathbf{r}$ for which only the dipoles that obey the condition $|\mathbf{r} - \mathbf{r}'| > r_{\text{cutoff}}$ are taken into account [5]. The potential profile for the present example and $r_{\text{cutoff}} = 1$ Å is shown in Fig. 5.9(c). Although this solution certainly improves the results and leads to a much better agreement with the analytical solution, it has the inconvenience of creating sharp transitions at the cutoff distances around the dipoles. To complement this treatment, we have now substituted the elimination of dipoles below the cutoff radius by a Gaussian smearing of dipoles that obey the condition $|\mathbf{r} - \mathbf{r}'| < r_{\text{smear}}$, as detailed in Section A.3 of Appendix A. This solution leads to smoother potentials and a much better agreement with the analytic solution for this test case, as observed in Fig. 5.9(d).

5.3.4 Computational aspects: method of layers and application to quantum wells

It is clear that when dealing with real size structures, for which the polarization is sampled at a very elevated number of sites, the calculation of the potential $\phi(\mathbf{r})$ becomes very expensive. In particular, for each $\mathbf{r}$, a summation over *all* the dipoles present in the system has to be done:

$$\phi(\mathbf{r}) = \sum_{\mathbf{p}} \phi_{\mathbf{p}}(\mathbf{r}). \tag{5.24}$$

In a system where the density of dipoles $n_{\mathbf{p}}$ is approximately constant, for instance one dipole located at each cation site in Ref. [5], the number of dipoles $\delta N_{\mathbf{p}}$ contributing to Eq. (5.24) located at distances between R and $R + \delta R$ from $\mathbf{r}$ is proportional to the surface area of a sphere of radius R:

$$\delta N_{\mathbf{p}} \propto 4\pi R^2 n_{\mathbf{p}} \delta R, \tag{5.25}$$

where δR is an infinitesimal increment in R. Because the contribution to $\phi(\mathbf{r})$ from each dipole decreases like $1/R^2$, as given by Eq. (5.20), Eq. (5.25) implies that the contribution to $\phi(\mathbf{r})$ due to the dipoles located at $\mathbf{r_p}$ for which $R < |\mathbf{r}-\mathbf{r_p}| < R+\delta R$ is of the same order of magnitude as the contribution due to dipoles for which $R' < |\mathbf{r} - \mathbf{r_p}| < R' + \delta R$, for any arbitrary $R' > R$. In other words, in principle, the sum in Eq. (5.24) does not converge. In practice, for real structures such as InGaN/GaN QWs, the facts that there is a dot product involved in the calculation of the potential due to each dipole, and that the dipoles in the barrier typically point in the same direction, give rise to opposite contributions that tend to cancel each other as R increases, as schematically shown in Fig. 5.10. In that case, the sum does converge although rather slowly. We propose two different methods to speed up the convergence of the sum in Eq. (5.24), one of which can be applied to any system, the "method of layers". The other method can be applied to systems where some assumption can be made about the value of the polarization being constant in the greatest part of the system, as is the case in QWs. An outline of these methods is given in Section A.4 of Appendix A.

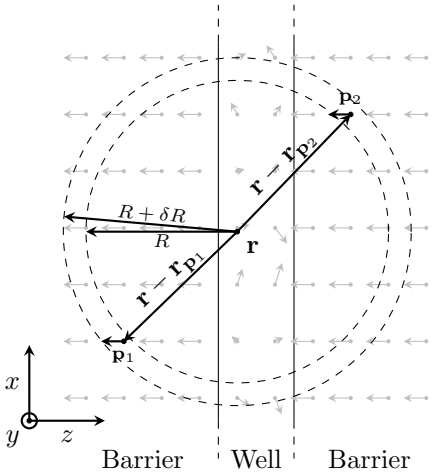


Figure 5.10: Schematic representation of the dipoles present in a typical nitride QW structure. In a sphere of radius R from $\mathbf{r}$ there exist a certain number of dipole pairs for which $\mathbf{p}_1 \cdot \left(\mathbf{r} - \mathbf{r}_{\mathbf{p}_1}\right) = -\mathbf{p}_2 \cdot \left(\mathbf{r} - \mathbf{r}_{\mathbf{p}_2}\right)$ and therefore tend to neutralize each other (they do not exactly cancel each other due to the image dipole effect that depends on how far $\mathbf{r}$ and $\mathbf{r}_\mathbf{p}$ are from *each* interface). For large R this effect is bigger as the polarization is usually constant in the barrier.

5.4 Selected results for InGaN quantum wells

Once the method for calculating the local polarization potential has been established, we can turn our attention towards achieving a local description of that quantity in relevant nanostructures. In the present example, we look at InGaN/GaN QWs grown along polar and non-polar[5] directions. Polar structures are grown along the c-axis, whereas in the case of non-polar structures the c-axis lies within the growth plane. In a macroscopic picture of the polarization, this ensures that there are no discontinuities in $\mathbf{P}$ between the well and barriers in the non-polar case. However, as we shall see, in a microscopic description discontinuities occur locally depending on local strain and composition.

While in Section 5.2.3 we have used DFT to optimize the atomic positions of the supercells studied, in the context of atomistic calculations involving large supercells highly accurate first-principles calculations are unaffordable in terms of computer time and memory usage. The usual approach to relax the atomic degrees of freedom in such cases is to use a classical interatomic force method. For tetrahedrally bonded compounds, Keating's valence force field (VFF) model [146]

[5]Non-polar QWs structures have precisely been proposed as a possible solution to the built-in field issue in nitride heterostructures [90].

is by far the most popular [147, 166]. Camacho and Niquet have previously used a modified version of Keating's model adapted to the WZ crystal structure to account for the deviation of the c/a ratio of lattice parameters with respect to its ideal value [147]. We have instead chosen an approach based on Martin's VFF [167] that includes the electrostatic interaction explicitly [5]. At a higher computational cost, this model succeeds at predicting the deviation of the c/a ratio while maintaining the correct symmetry of the interatomic interactions. For instance, the two-body interactions directed along the WZ c-axis have the same functional form, including the equilibrium bond length, as the other ones. This allows to obtain a much more flexible set of potentials that are transferable between similar polymorphs of the same compound, i.e. WZ and ZB in this case. With our model we are able to predict elastic and structural properties of binary and ternary nitrides in excellent agreement with first-principles DFT calculations, therefore providing solid grounds for using the supercells relaxed using this method as high-quality input for the subsequent local polarization calculation. An extensive article with the details and validity of our method is currently in preparation and will be published elsewhere (see also Chapter 7).

Making use of the expressions derived throughout this chapter, and the VFF just outlined, we have calculated the local polarization for InGaN/GaN QWs with 30% In content in both polar and non-polar orientations, as shown in Figs. 5.11(a) and (b), respectively. Note that the component shown in the colour code is the component of the polarization along the c-axis. The correspondent polarization potential is shown in Figs. 5.11(c), for the polar case, and (d), for the non-polar situation. It can be observed that the polar structure shows a potential profile with the main features of a capacitor-like structure, although fluctuations occur. For a constant value of the polarization, i.e. no local effects taken into account, the isolines in Fig. 5.11(c) would be perfectly parallel to each other, as seen already in Fig. 5.9. In the non-polar case [Fig. 5.11(d)] there are no main features in the potential but only local effects. Ongoing studies we are carrying out show how these local fluctuations have a sizeable impact on the electronic properties of these structures. Specifically, we have found that the local polarization potential affects the valence band edge bowing in InGaN alloys. An outline of this work will be presented in Chapter 7.

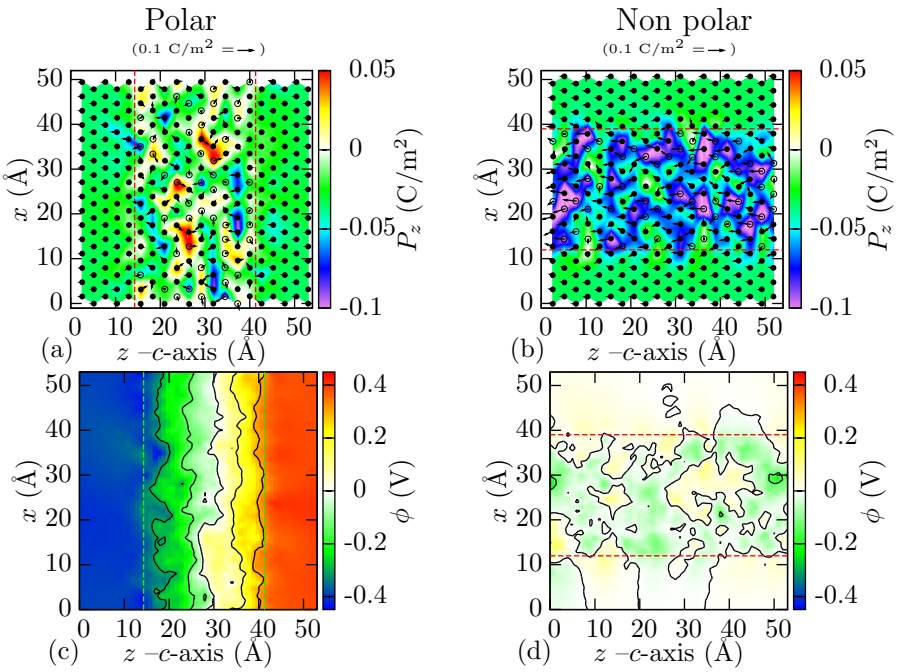


Figure 5.11: Sections in a plane parallel to the c-axis of $In_{0.3}Ga_{0.7}N$/GaN QWs in polar and non-polar orientations. The component of the polarization along the c-axis, P_z, for the polar and non-polar structures is shown in (a) and (b), respectively. The corresponding polarization potential is shown in (c) and (d). The dashed lines denote the approximate location of the interfaces between well and barriers (because of the atomistic description of the system, the concept of interface is ill-defined and an "exact" location cannot be provided). The arrows in (a) and (b) give direction of the polarization in the xz plane, as well as its magnitude in the same units as the arrow in the legend, which indicates 0.1 C/m^2. Solid circles are Ga atoms and open circles are In atoms.

5.5 Summary

We have presented a complete theory of local electric polarization in the linear piezoelectric limit. The connection between the local polarization and local internal strain is obtained in an elegant manner through the use of Born effective charges and internal strain parameters. We have validated the theory against the highly ionic III-N wurtzite compounds, obtaining a high degree of agreement between our model and Berry-phase calculations. We have cast these local effects in the form of a local piezoelectric tensor, which elucidates the implications of local strain and tetrahedron orientation on the polarization. In addition to this, we have obtained a consistent series of polarization-related *ab initio* parameters for the III-N.

We have also presented a point dipole method for the calculation of the local polarization potential that overcomes resolution problems encountered when solving directly Poisson's equation. The method involves the discretization of the polarization field as a series of point dipoles. The accuracy of the method has been tested against a well known problem with analytical solution.

Finally, as an example, we have applied our theory and methodology to study the local polarization and local polarization potential in polar and non-polar InGaN/GaN QW structures, where we have observed large local fluctuations in both quantities.

6 Staggered-grid formulation of a continuum elasticity method for efficient structural relaxation of zinc-blende and wurtzite crystal lattices

Abstract

In this chapter we present a continuum elasticity method for very efficient strain energy minimization of conventional zinc-blende and wurtzite crystal lattices. We also present results for symmetry-adapted (111)-oriented zinc-blende systems. The formulation of the strain energy, the displacement components and their derivatives is done via finite differences on a staggered grid, leading to a more accurate description of those quantities and avoiding common problems related to averaging and ill-defined boundaries between different materials. The implementation of structures of arbitrary shape is straightforward as the staggered grid formulation ensures that the specification of the material properties at the grid sites is sufficient to determine the geometry of the system. We implement a hybrid evaluation of the condition of minimum energy, that allows an extremely efficient analytical derivation of the linear system, that can be solved numerically using standard sparse-matrix methods. The results obtained with our method are in good agreement with previously reported data with the additional advantage of being obtained significantly faster.

6.1 Introduction

Great improvement has been achieved in recent years in the realm of atomistic calculations of the electronic properties of nanoscale systems. As a proof of this, and except for Chapter 3, all of the present work has aimed at improving the knowledge of semiconductor materials at a level of accuracy that can only be achieved when taking the discrete nature of matter into account. However, the

large number of atoms present in structures such as quantum logical devices [168], makes such methods impractical beyond certain sizes. For the study of those (relatively) large systems, continuum methods, such as the $\mathbf{k} \cdot \mathbf{p}$ method, need to be employed.

The use of highly lattice-mismatched semiconductor materials, such as III-N heterostructures can lead to large built-in strain and piezoelectric fields in devices based on such materials, as we have previously seen. The determination of strain in nanostructures is computationally a heavy task, and a considerable part of the time for any simulation that takes strain into account is usually spent on its calculation. Methods that admit analytical solutions in some cases, such as the surface integral method [101, 115], can provide approximate solutions. These methods are helpful for identifying trends or carrying out cheaper calculations when a large number of systems is involved (e.g. optimization procedure of Chapter 3). However, they rely on assumptions regarding isotropy of the system properties, or apply restrictions to the complexity of the system's geometry [101, 115]. Therefore, for a more accurate description of strain for systems with an arbitrary geometry whose size lies beyond the reach of atomistic calculations, and to get insights into the impact of strain on their electronic and optical properties, one must rely on numerical implementations of continuum elasticity. Here we present a hybrid analytical-numerical method that helps substantially alleviate the computational requirements of such calculations while at the same time improving their accuracy. This is achieved through use of a staggered-grid formulation of the elastic energy problem.

In Section 6.2.1 we give an introduction to strain theory in discrete space, while Section 6.2.2 treats the formulation for a staggered grid. In Section 6.3, an analytical evaluation of the elastic energy minimization is presented. Some results of the application of the present method, as well as a study of its performance for large systems, are presented in Section 6.4, and finally our conclusions are summarized in Section 6.5.

6.2 General form of the strain on a discrete mesh

6.2.1 Forward and backward finite differences

In continuum elasticity theory, the strain at each point in a crystal is given as a function of the derivatives of the displacement field $\mathbf{u}$ with respect to the Cartesian

coordinates at that point [39]:

$$\epsilon_{ij}^{\mathbf{u}}(\mathbf{r}) = \frac{1}{2}\left(\frac{\partial u_i(\mathbf{r})}{\partial j} + \frac{\partial u_j(\mathbf{r})}{\partial i}\right), \tag{6.1}$$

where $i, j \equiv x, y, z$. It is important to note that, although the strain $\epsilon_{ij}^{\mathbf{u}}$ is defined as a symmetric quantity ($\epsilon_{ij}^{\mathbf{u}} = \epsilon_{ji}^{\mathbf{u}}$), the derivatives in Eq. (6.1) are not necessarily symmetric (i.e., in general $\frac{\partial u_i}{\partial j} \neq \frac{\partial u_j}{\partial i}$). In discrete space, either forward or backward finite differences are the common approach found in the literature when approximating the derivatives, because symmetric finite differences lead to oscillatory solutions, and should therefore be avoided [161,166]. In this scheme, in discrete space Eq. (6.1) takes the form

$$\epsilon_{ij}^{\mathbf{u},\alpha\beta}(\mathbf{r}) = \frac{1}{2}\left(\Delta_j^{\alpha} u_i(\mathbf{r}) + \Delta_i^{\beta} u_j(\mathbf{r})\right), \tag{6.2}$$

where Δ_i^{α} is the finite difference operator along the x_i axis, forward for $\alpha \equiv (+)$ and backward for $\alpha \equiv (-)$:

$$\begin{aligned}
\Delta_i^{+} u_j(\mathbf{r}) &= \frac{u_j\left(\mathbf{r} + h_i\hat{\boldsymbol{i}}\right) - u_j(\mathbf{r})}{h_i}, \\
\Delta_i^{-} u_j(\mathbf{r}) &= \frac{u_j(\mathbf{r}) - u_j\left(\mathbf{r} - h_i\hat{\boldsymbol{i}}\right)}{h_i},
\end{aligned} \tag{6.3}$$

where h_i is the step size along the i coordinate axis (that need not be necessarily equal to the step sizes along the other axes) and $\hat{\boldsymbol{i}}$ is a unit vector along the same axis.

We will refer to Eqs. (6.1) and (6.2), which are the classical definition of strain in continuum elasticity theory, as the *displacement field strain* (in contrast to the *physical strain*), because they do not give the physical strain when two materials with a different lattice constant are present in the crystal, as is the case in quantum dot (QD) systems, for instance. To prove the latter statement, let us consider an extremely simple one-dimensional system consisting of two different materials with lattice constants a_1 and a_2, with $a_2 > a_1$ (see Fig. 6.1). It is natural and simplest to work in a coordinate system in which at least one of the materials is relaxed (i.e. there is no strain) if the displacement field is zero. Let this be the case for material (1) in the present example. The spatial axis is discretized

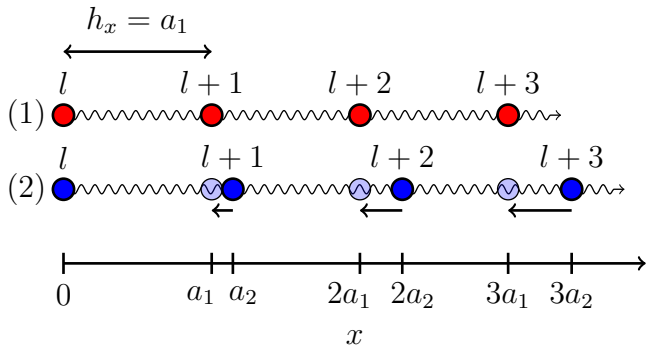


Figure 6.1: Representation of materials (1) and (2), with different lattice constants, on a one-dimensional grid. The solid circles give the particular discretization for which each material would be in their preferential coordinate system (in which they are relaxed if $\mathbf{u} = 0$), and the blue shaded circles represent the case in which material (2) is discretized in the preferential coordinate system of material (1).

for simplicity in steps $h_x = a_1$. To be represented in this coordinate system with this particular discretization, material (2) has to suffer an initial compression accounting for $\epsilon^0_{xx} = (a_1 - a_2)/a_2$, as shown in Fig. 6.1. It is now clear why in continuum elasticity theory the displacement field does not account for the physical strain when several materials are involved: the offset due to transforming one (or more) materials from their preferential coordinate system (in which they would be relaxed if $\mathbf{u} = 0$) to an arbitrary one has to be taken into account. However, the relation $\epsilon_{xx} = \epsilon^{\mathbf{u}}_{xx} + \epsilon^0_{xx}$ alone is not sufficient. Additionally, for material (2) to be relaxed, i.e. $\epsilon_{xx} = 0$, it must have $\epsilon^{\mathbf{u}}_{xx} = \frac{a_2 - a_1}{a_1}$, which would correspond to the light circles in Fig. 6.1 going back to their original positions. Assuming that ϵ_{xx} is linear in $\epsilon^{\mathbf{u}}_{xx}$ and making use of the two former requirements, the physical strain can be finally defined, for this particular setup, as

$$\epsilon_{xx} = \begin{cases} \epsilon^{\mathbf{u}}_{xx} & \text{for material (1).} \\ \frac{a_1}{a_2}\epsilon^{\mathbf{u}}_{xx} + \frac{a_1 - a_2}{a_2} & \text{for material (2).} \end{cases} \tag{6.4}$$

In this one-dimensional example, and with a forward difference approximation for the derivative, the displacement field strain $\epsilon^{\mathbf{u},+}_{xx}$ at point l would be given by

$$\epsilon^{\mathbf{u},+}_{xx} = \Delta^+_x u_x = \frac{u_{x,l+1} - u_{x,l}}{a_1}. \tag{6.5}$$

Equations (6.4) and (6.5) can be readily generalized to three-dimensional systems

and position-dependent lattice constants. The diagonal strain components are given by

$$\epsilon_{ii}^{\alpha}(\mathbf{r}) = \frac{a_i^{\mathrm{ref}}}{a_i(\mathbf{r})}\Delta_i^{\alpha} u_i(\mathbf{r}) + \frac{a_i^{\mathrm{ref}} - a_i(\mathbf{r})}{a_i(\mathbf{r})}, \tag{6.6}$$

and the shear strain components by

$$\epsilon_{ij}^{\alpha\beta}(\mathbf{r}) = \frac{1}{2}\left(\frac{a_j^{\mathrm{ref}}}{a_j(\mathbf{r})}\Delta_j^{\alpha} u_i(\mathbf{r}) + \frac{a_i^{\mathrm{ref}}}{a_i(\mathbf{r})}\Delta_i^{\beta} u_j(\mathbf{r})\right) \qquad \text{with} \qquad i \neq j, \tag{6.7}$$

where a_i^{ref} is the lattice constant of the reference material along the i axis. This means that the physical strain for the reference material is given directly by the displacement field strain. The most intuitive choice of a reference, in terms of interpretation of the results, is perhaps the barrier material in the case of a QD, or the most abundant material in the cell used for the calculation otherwise. However, any particular choice, even of a material that is not contained in the cell, does not affect the value of the physical strain, calculated as in Eqs. (6.6) and (6.7). Equations (6.6) and (6.7) are consistent with the fact that in wurtzite lattices, for instance, the lattice constant is anisotropic and, therefore, a different prefactor a^{ref}/a needs to be considered for derivatives performed in different directions. Equations (6.6) and (6.7) are the general form of the strain in continuum elasticity theory for a discrete mesh, in a forward/backward finite difference scheme. However, in the next section, we shall see why forward and/or backward finite differences should be avoided and how the use of staggered grids overcomes several issues related to them.

6.2.2 Formulation of the strain in a staggered grid

The reason why forward/backward finite differences should be avoided in the definition of the strain is related to the fact that the resultant derivatives of the displacement field are not defined in the natural grid of the problem, that is, the grid where the material parameters (lattice and elastic constants) are given. Let us consider a standard rectangular grid for which the position of each grid element is given by $\mathbf{r} \equiv (lh_x, mh_y, nh_z)$, where $l, m, n \in \mathbb{Z}$ and $h_x, h_y, h_z \in \mathbb{R}^+$ are the step sizes in each direction. Once the step sizes are known, the grid elements can be unambiguously identified by the indices (l, m, n) alone. When forward and back-

ward finite differences are used, both material properties and displacement field components are given at (l, m, n). However, the derivatives of the displacement field, which are the key quantities involved in the calculation of the strain, are *assigned* to the grid points (l, m, n) by convention. For instance, the quantity $[u_x(l+1, m, n) - u_x(l, m, n)]/h_x$ is *chosen* to be the derivative $\partial u_x/\partial x$ at (l, m, n) when forward differences are used *and* the same derivative at $(l+1, m, n)$ when backward differences are the choice. The problems associated with this approach are two-fold. First, the value of the derivative is arbitrarily defined *by convention* and second, in a system where (l, m, n) corresponds to one particular material and $(l+1, m, n)$ to another (that is, at an interface) the interface itself is ill-defined and its position shifted between one convention and the other. Also, any particular choice of sign for α and β in the derivatives involved in Eqs. (6.6) and (6.7) leads to a preferential direction being singled out during an elastic energy minimization procedure. To prevent this, all eight configurations corresponding to all possible permutations of the three spatial derivatives must be considered and the corresponding energies averaged [161, 166]. Averaging leads to additional computational effort and loss of resolution.

One way to overcome these limitations is by using a staggered grid. Following the previous example, it would be natural to identify the quantity $[u_x(l+1, m, n) - u_x(l, m, n)]/h_x$ as the derivative $\partial u_x/\partial x$ at $(l+\frac{1}{2}, m, n)$. However, the latter is not a natural element of the grid and the material properties are not clearly defined at that point. Alternatively, one could define the displacement u_x at $(l+\frac{1}{2}, m, n)$ and $(l-\frac{1}{2}, m, n)$, yielding

$$\frac{\partial u_x(l, m, n)}{\partial x} = \frac{u_x(l+\frac{1}{2}, m, n) - u_x(l-\frac{1}{2}, m, n)}{h_x}, \tag{6.8}$$

which is defined at the natural grid elements (l, m, n). The displacement field components are not a physical quantity strictly speaking, but rather a mathematical entity. The *derivatives* of the displacement field, however, are involved in the calculation of the strain and are physically meaningful. Thus, Eq. (6.8) should be preferred when calculating $\partial u_x/\partial x$, in the sense that it ensures that the derivative is correctly defined in the natural grid of the problem. Using the same approach as in Eq. (6.8), we can systematically set up a series of staggered grids where all

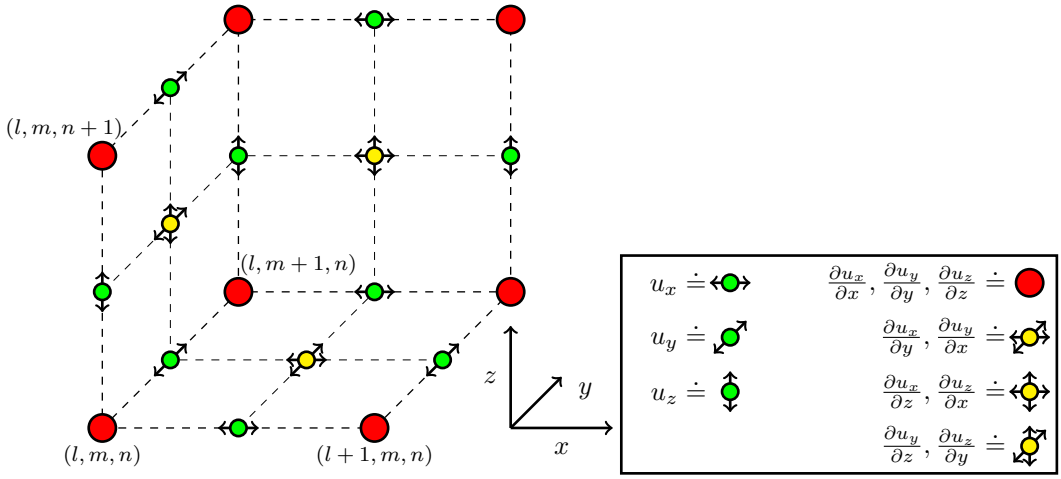


Figure 6.2: Staggered grid proposed in the present work. The big red circles give the natural grid, where the properties of the materials (lattice and elastic constants) are well defined; the axial derivatives are also defined at those points. The small circles give the positions at which the displacement field (green) and crossed derivatives (yellow) are defined.

the relevant quantities are defined. The axial displacement field is defined at

$$u_x \doteq \left(l + \frac{1}{2}, m, n\right), \qquad u_y \doteq \left(l, m + \frac{1}{2}, n\right), \qquad u_z \doteq \left(l, m, n + \frac{1}{2}\right),$$

the axial derivatives at

$$\frac{\partial u_x}{\partial x} \doteq \frac{\partial u_y}{\partial y} \doteq \frac{\partial u_z}{\partial z} \doteq \epsilon_{xx} \doteq \epsilon_{yy} \doteq \epsilon_{zz} \doteq (l, m, n),$$

and the crossed derivatives at

$$\begin{aligned}
\frac{\partial u_x}{\partial y} &\doteq \frac{\partial u_y}{\partial x} \doteq \epsilon_{xy} \doteq \left(l + \frac{1}{2}, m + \frac{1}{2}, n\right), \\
\frac{\partial u_x}{\partial z} &\doteq \frac{\partial u_z}{\partial x} \doteq \epsilon_{xz} \doteq \left(l + \frac{1}{2}, m, n + \frac{1}{2}\right), \\
\frac{\partial u_y}{\partial z} &\doteq \frac{\partial u_z}{\partial y} \doteq \epsilon_{yz} \doteq \left(l, m + \frac{1}{2}, n + \frac{1}{2}\right),
\end{aligned}$$

where $\doteq$ means "defined on the grid element". This set of grids is better pictured in Fig. 6.2. As can be observed, the derivatives corresponding to the axial strains $(\partial u_i/\partial i)$ are naturally defined in the same grid that gives the material geometry. At those points the material parameters (lattice and elastic constants) are well

defined. As a consequence, the crossed derivatives involved in the calculation of the shear strains ($\partial u_i/\partial j$ with $i \neq j$) are defined at the centre of the faces of the box with corners (l, m, n), $(l+1, m, n)$, etc. Therefore, as will be seen in the next section, some averaging needs to be done for the material properties at those points. Although an inconvenience, this is still preferred to averaging over the eight possible different combinations of forward and backward differences. This is because with our staggered grid the axial strains ϵ_{ii}, which are discontinuous at the interfaces [there is an initial misfit strain as given by Eq. (6.6)], are calculated exactly. By contrast, shear strains are usually much smoother at the interfaces between different materials [Eq. (6.7)], and their averaging has a correspondingly smaller impact on the total elastic energy.

Equations (6.6) and (6.7) are rewritten on the staggered grid as

$$\epsilon_{ii}(\mathbf{r}) = \frac{a_i^{\mathrm{ref}}}{a_i(\mathbf{r})}\Delta_i^{\dagger} u_i(\mathbf{r}) + \frac{a_i^{\mathrm{ref}} - a_i(\mathbf{r})}{a_i(\mathbf{r})}, \tag{6.9}$$

and

$$\epsilon_{ij}(\mathbf{r}) = \frac{1}{2}\left(\frac{a_j^{\mathrm{ref}}}{a_j(\mathbf{r})}\Delta_j^{\dagger} u_i(\mathbf{r}) + \frac{a_i^{\mathrm{ref}}}{a_i(\mathbf{r})}\Delta_i^{\dagger} u_j(\mathbf{r})\right) \qquad \text{with} \quad i \neq j, \tag{6.10}$$

where the "staggered" operators $\Delta_i^{\dagger}$ are given by

$$\Delta_i^{\dagger} u_i(l,m,n) = \frac{u_i\left(l + \frac{\delta_{xi}}{2}, m + \frac{\delta_{yi}}{2}, n + \frac{\delta_{zi}}{2}\right) - u_i\left(l - \frac{\delta_{xi}}{2}, m - \frac{\delta_{yi}}{2}, n - \frac{\delta_{zi}}{2}\right)}{h_i}, \tag{6.11}$$

and

$$\Delta_j^{\dagger} u_i\left(l + \frac{\theta_{xij}}{2}, m + \frac{\theta_{yij}}{2}, n + \frac{\theta_{zij}}{2}\right) = \frac{u_i\left(l + \frac{\delta_{xi}}{2} + \delta_{xj}, m + \frac{\delta_{yi}}{2} + \delta_{yj}, n + \frac{\delta_{zi}}{2} + \delta_{zj}\right) - u_i\left(l + \frac{\delta_{xi}}{2}, m + \frac{\delta_{yi}}{2}, n + \frac{\delta_{zi}}{2}\right)}{h_j} \qquad \text{with} \quad i \neq j, \tag{6.12}$$

where δ_{ij} is the Kronecker delta function and $\theta_{\nu ij}$ is defined as

$$\theta_{\nu ij} = \delta_{\nu i} + \delta_{\nu j}.$$

Equations (6.9) and (6.10) are the general form of the strain on the staggered grid presented here. The next section will deal with the implementation of these equations in an efficient algorithm for elastic energy minimization.

6.3 Elastic energy minimization

6.3.1 Form of the elastic energy in discrete continuum elasticity theory

In continuum elasticity theory, the elastic energy of a volume element V_0 of a crystal is related to the strain via the elastic (or *stiffness*) tensor:

$$E = \frac{V_0}{2} \sum_{i,j,k,l} C_{ijkl} \epsilon_{ij} \epsilon_{kl}, \tag{6.13}$$

where the i, j, k and l indices refer to the Cartesian coordinates ($i, j, k, l \equiv x, y, z$), C_{ijkl} are the elastic constants, ϵ_{ij} are the strain components and V_0 is the equilibrium volume. Given the particular symmetry of the crystal, Eq. (6.13) can usually be greatly simplified. As we saw in Chapter 4, for zinc-blende lattices, E becomes

$$\begin{aligned} E^{\mathrm{ZB}} =& \frac{V_0}{2} C_{11} \left({\epsilon_1}^2 + {\epsilon_2}^2 + {\epsilon_3}^2 \right) + V_0 C_{12} \left(\epsilon_1 \epsilon_2 + \epsilon_1 \epsilon_3 + \epsilon_2 \epsilon_3 \right) \\ &+ \frac{V_0}{2} C_{44} \left({\epsilon_4}^2 + {\epsilon_5}^2 + {\epsilon_6}^2 \right), \end{aligned} \tag{6.14}$$

whereas the expression is slightly more complicated for wurtzite lattices:

$$\begin{aligned} E^{\mathrm{WZ}} =& \frac{V_0}{2} C_{11} \left({\epsilon_1}^2 + {\epsilon_2}^2 \right) + \frac{V_0}{2} C_{33} {\epsilon_3}^2 + V_0 C_{12} \epsilon_1 \epsilon_2 + V_0 C_{13} \left(\epsilon_1 \epsilon_3 + \epsilon_2 \epsilon_3 \right) \\ &+ \frac{V_0}{2} C_{44} \left({\epsilon_4}^2 + {\epsilon_5}^2 \right) + \frac{V_0}{2} C_{66} {\epsilon_6}^2, \end{aligned} \tag{6.15}$$

with $C_{66} = \frac{C_{11} - C_{12}}{2}$. Voigt notation has been employed in Eqs. (6.14) and (6.15) for the sake of simplicity.[1] Equation (6.14) is a simplification of Eq. (6.15) when $C_{33} = C_{11}$, $C_{13} = C_{12}$ and $C_{66} = C_{44}$. Therefore, for simplicity, Eq. (6.15) will

[1] Note that in Voigt notation, $\epsilon_1 = \epsilon_{xx}$, $\epsilon_2 = \epsilon_{yy}$, $\epsilon_3 = \epsilon_{zz}$, $\epsilon_4 = 2\epsilon_{yz}$, $\epsilon_5 = 2\epsilon_{xz}$ and $\epsilon_6 = 2\epsilon_{xy}$.

be used throughout the following discussion for both ZB and WZ lattices without loss of generality. C_{66} will be kept explicitly since the $C_{66} = \frac{C_{11}-C_{12}}{2}$ relation is not valid for zinc-blende.

As mentioned in the previous section, in our staggered grid the axial strains and the shear strains are not defined on the same grid. Therefore the elastic energy must be split into several contributions evaluated at different sites. In three-dimensional discrete space, for axial strains ϵ_{ii} the elastic energy is evaluated at the mesh elements (l, m, n) with equilibrium volume $V_0^{l,m,n}$ and takes the form

$$\begin{aligned} E_{l,m,n} =& \frac{V_0^{l,m,n}}{2} C_{11}^{l,m,n} \left([\epsilon_1(l,m,n)]^2 + [\epsilon_2(l,m,n)]^2 \right) + \frac{V_0^{l,m,n}}{2} C_{33}^{l,m,n} [\epsilon_3(l,m,n)]^2 \\ &+ V_0^{l,m,n} C_{12}^{l,m,n} \epsilon_1(l,m,n)\epsilon_2(l,m,n) \\ &+ V_0^{l,m,n} C_{13}^{l,m,n} \left[\epsilon_1(l,m,n)\epsilon_3(l,m,n) + \epsilon_2(l,m,n)\epsilon_3(l,m,n) \right], \end{aligned} \tag{6.16}$$

whereas for the shear strain contributions we have

$$\begin{aligned} E_{l,m+\frac{1}{2},n+\frac{1}{2}} &= \frac{V_0^{l,m+\frac{1}{2},n+\frac{1}{2}}}{2} C_{44}^{l,m+\frac{1}{2},n+\frac{1}{2}} \left[\epsilon_4 \left(l, m+\frac{1}{2}, n+\frac{1}{2} \right) \right]^2, \\ E_{l+\frac{1}{2},m,n+\frac{1}{2}} &= \frac{V_0^{l+\frac{1}{2},m,n+\frac{1}{2}}}{2} C_{44}^{l+\frac{1}{2},m,n+\frac{1}{2}} \left[\epsilon_5 \left(l+\frac{1}{2}, m, n+\frac{1}{2} \right) \right]^2, \\ E_{l+\frac{1}{2},m+\frac{1}{2},n} &= \frac{V_0^{l+\frac{1}{2},m+\frac{1}{2},n}}{2} C_{66}^{l+\frac{1}{2},m+\frac{1}{2},n} \left[\epsilon_6 \left(l+\frac{1}{2}, m+\frac{1}{2}, n \right) \right]^2. \end{aligned} \tag{6.17}$$

The equilibrium volume $V_0^{l,m,n}$ is given by the step sizes h_i and the lattice constants of the material at the $(l, m, n)^{\text{th}}$ site:

$$V_0^{l,m,n} = \frac{a_x(l,m,n) a_y(l,m,n) a_z(l,m,n)}{a_x^{\text{ref}} a_y^{\text{ref}} a_z^{\text{ref}}} h_x h_y h_z. \tag{6.18}$$

As previously mentioned, the lattice and elastic constants in Eq. (6.17) need to be evaluated in grid elements other than their natural ones. For those, we propose a linear interpolation between the values at the four neighboring points of the

natural grid (l, m, n) where they are well defined:

$$\begin{aligned}
&\xi(l+\frac{1}{2}, m+\frac{1}{2}, n) = \\
&\quad \frac{1}{4}\big[\xi(l,m,n) + \xi(l+1,m,n) + \xi(l,m+1,n) + \xi(l+1,m+1,n)\big], \\
&\xi(l+\frac{1}{2}, m, n+\frac{1}{2}) = \\
&\quad \frac{1}{4}\big[\xi(l,m,n) + \xi(l+1,m,n) + \xi(l,m,n+1) + \xi(l+1,m,n+1)\big], \\
&\xi(l, m+\frac{1}{2}, n+\frac{1}{2}) = \\
&\quad \frac{1}{4}\big[\xi(l,m,n) + \xi(l,m+1,n) + \xi(l,m,n+1) + \xi(l,m+1,n+1)\big]. \qquad (6.19)
\end{aligned}$$

The total energy of the cell used in the calculation is given by the sum over the $N_x \times N_y \times N_z$ points contained in the mesh:

$$E_{\text{total}} = \sum_{l,m,n} E_{l,m,n} + {\sum_{l,m,n}}' \big(E_{l,m+\frac{1}{2},n+\frac{1}{2}} + E_{l+\frac{1}{2},m,n+\frac{1}{2}} + E_{l+\frac{1}{2},m+\frac{1}{2},n}\big). \qquad (6.20)$$

The prime symbol in the second summation in Eq. (6.20) indicates that the summation rule is different depending on the boundary conditions: for periodic boundary conditions l, m and n run over all N_x, N_y and N_z, respectively. For fixed boundary conditions, the terms for which $l + \frac{1}{2} > N_x$, $m + \frac{1}{2} > N_y$ and $n + \frac{1}{2} > N_z$ have to be left out.

E_{total} is a function of the displacement field through the relation between ϵ and $\mathbf{u}$ [Eqs. (6.9) and (6.10)]. Therefore, in order to find the equilibrium configuration for any given structure, E_{total} must be minimized with respect to the different components of $\mathbf{u}$. The condition of minimum is

$$\frac{\partial E_{\text{total}}(\mathbf{u})}{\partial u_i\left(l + \frac{\delta_{xi}}{2}, m + \frac{\delta_{yi}}{2}, n + \frac{\delta_{zi}}{2}\right)} = 0. \qquad (6.21)$$

Again, the number of variables (and hence equations) depends on whether periodic or fixed boundary conditions are being used. The number of components of the displacement field $\mathbf{u}$ is $N = 3N_xN_yN_z$ for periodic boundary conditions, and $N = 3N_xN_yN_z - N_xN_y - N_xN_z - N_yN_z$ for fixed boundary conditions. Because E_{total} is quadratic in $\mathbf{u}$, Eq. (6.21) leads to a system of N equations, which are linear in the

N variables u_i. Whereas the usual approach is to directly evaluate the derivative in Eq. (6.21) numerically [166], it will be shown in the next section how an analytical evaluation of the same simplifies the problem substantially and allows for a very efficient calculation of the displacement field.

6.3.2 Analytical reformulation of the condition of minimum

We have shown already that the total elastic energy of the cell used in a calculation can be separated into smaller contributions given at different grid elements, each of which is given by Eqs. (6.16) or (6.17). For the elementary energy $E_{l,m,n}$, only the strains evaluated at the (l,m,n) site are involved. The same is true for the shear strain part of the energy $E_{l+\frac{1}{2},m+\frac{1}{2},n}$, $E_{l+\frac{1}{2},m,n+\frac{1}{2}}$ and $E_{l,m+\frac{1}{2},n+\frac{1}{2}}$, and their respective grid elements. This means that a particular component of the displacement field is involved in the calculation of a very small part of the total energy. It can be seen from Fig. 6.3 how $u_x\left(l+\frac{1}{2},m,n\right)$ is involved in the calculation of the energy elements $E_{l,m,n}$, $E_{l+1,m,n}$, $E_{l+\frac{1}{2},m+\frac{1}{2},n}$, $E_{l+\frac{1}{2},m-\frac{1}{2},n}$, $E_{l+\frac{1}{2},m,n+\frac{1}{2}}$ and $E_{l+\frac{1}{2},m,n-\frac{1}{2}}$ alone. Consequently, Eq. (6.21) can be expressed in a simplified way as

$$\begin{aligned}
&\frac{\partial E_{\text{total}}\left(\mathbf{u}\right)}{\partial u_x(l+\frac{1}{2},m,n)} = \\
&\quad\frac{\partial\left(E_{l,m,n}+E_{l+1,m,n}+E_{l+\frac{1}{2},m+\frac{1}{2},n}+E_{l+\frac{1}{2},m-\frac{1}{2},n}+E_{l+\frac{1}{2},m,n+\frac{1}{2}}+E_{l+\frac{1}{2},m,n-\frac{1}{2}}\right)}{\partial u_x(l+\frac{1}{2},m,n)} = 0, \\
&\frac{\partial E_{\text{total}}\left(\mathbf{u}\right)}{\partial u_y(l,m+\frac{1}{2},n)} = \\
&\quad\frac{\partial\left(E_{l,m,n}+E_{l,m+1,n}+E_{l+\frac{1}{2},m+\frac{1}{2},n}+E_{l-\frac{1}{2},m+\frac{1}{2},n}+E_{l,m+\frac{1}{2},n+\frac{1}{2}}+E_{l,m+\frac{1}{2},n-\frac{1}{2}}\right)}{\partial u_y(l,m+\frac{1}{2},n)} = 0, \\
&\frac{\partial E_{\text{total}}\left(\mathbf{u}\right)}{\partial u_z(l,m,n+\frac{1}{2})} = \\
&\quad\frac{\partial\left(E_{l,m,n}+E_{l,m,n+1}+E_{l+\frac{1}{2},m,n+\frac{1}{2}}+E_{l-\frac{1}{2},m,n+\frac{1}{2}}+E_{l,m+\frac{1}{2},n+\frac{1}{2}}+E_{l,m-\frac{1}{2},n+\frac{1}{2}}\right)}{\partial u_z(l,m,n+\frac{1}{2})} = 0.
\end{aligned} \tag{6.22}$$

The expression for the derivative in Eq. (6.22), although rather tedious, can be obtained analytically. The result is a system of N equations in which each equation depends on only 15 independent variables of the total N displacement field components u_i. This number is of course independent of the total number of vari-

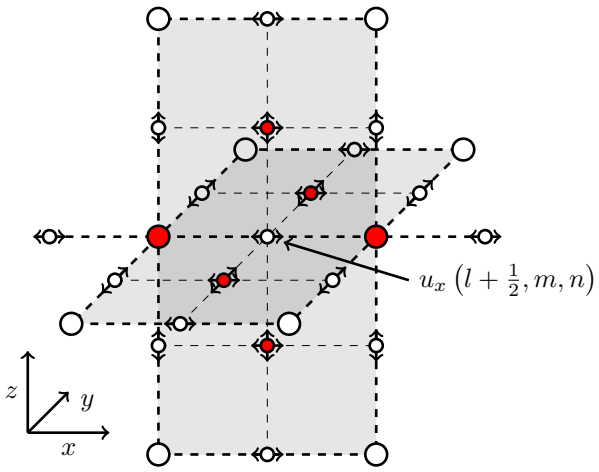


Figure 6.3: All the grid sites immediately surrounding $(l+\frac{1}{2},m,n)$. The dark red circles indicate the grid sites at which $u_x(l+\frac{1}{2},m,n)$ is involved in the calculation of energy elements: $E_{l,m,n}$, $E_{l+1,m,n}$, $E_{l+\frac{1}{2},m+\frac{1}{2},n}$, $E_{l+\frac{1}{2},m-\frac{1}{2},n}$, $E_{l+\frac{1}{2},m,n+\frac{1}{2}}$ and $E_{l+\frac{1}{2},m,n-\frac{1}{2}}$, in this case. Similar constructions can be made for the displacements $u_y(i,j+\frac{1}{2},k)$ and $u_z(i,j,k+\frac{1}{2})$.

ables N, which means that the size of the problem scales approximately linearly with the number of grid points.[2] In this manner, Eq. (6.22) can be given as the sum of the product of the variables $u_{q,i}^{l\,m\,n}$ and constant coefficients $C_{q,i}^{l\,m\,n}$:[3]

$$\sum_{q=1}^{15} C_{q,i}^{l\,m\,n} u_{q,i}^{l\,m\,n} + C_{0,i}^{l\,m\,n} = 0, \tag{6.23}$$

where $C_{0,i}^{l\,m\,n}$ is the constant term. The coefficient $C_{q,i}^{l\,m\,n}$ is the number multiplying variable $u_{q,i}^{l\,m\,n}$ in the expression for the derivative $\partial E_{\text{total}}/\partial u_i^{l\,m\,n}$ of Eq. (6.22). We have derived the explicit form of Eq. (6.22) for crystals with wurtzite symmetry, for which the lattice constants are given by $a_x, a_y = a$ and $a_z = c$, and implemented them in our code HUSH.[4] To obtain the zinc-blende coefficients from the wurtzite ones, c, C_{13}, C_{33} and C_{66} must be replaced by a, C_{12}, C_{11} and C_{44}, respectively. As an example, the explicit form of Eq. (6.22) for the derivative of E_{total} with respect

[2]The scaling is linear for periodic boundary conditions with an additional sublinear correction for fixed boundary conditions: $N_x \times N_y \times N_z$ is the number of points in the mesh and $N = 3N_xN_yN_z$ for periodic boundary conditions and $N = 3N_xN_yN_z - N_xN_y - N_xN_z - N_yN_z$ for fixed boundary conditions.

[3]For simplicity, we use an abbreviated notation for the components of the displacement field: $u_{q,i}^{l\,m\,n}$ (with $q = 1,\ldots,15$) are each of the displacement field components that appear in the derivative of E_{total} with respect to $u_i(l+\frac{\delta_{xi}}{2}, m+\frac{\delta_{yi}}{2}, n+\frac{\delta_{zi}}{2}) \equiv u_i^{l\,m\,n}$.

[4]HUSH stands for *Hybrid Utility for Strained Heterostructures*.

to $u_x(l+\frac{1}{2},m,n)$ is given in Appendix B.

Equation (6.23) can be rewritten in matrix form as

$$A\mathbf{X}+\mathbf{B}=\mathbf{0}, \tag{6.24}$$

where A is the matrix of coefficients $C_{q,i}^{l\,m\,n}$,[5] $\mathbf{X}$ is a column vector with all the N variables $u_i^{l\,m\,n}$, $\mathbf{B}$ is a column vector with the constant terms $C_{0,i}^{l\,m\,n}$ and $\mathbf{0}$ is a column vector of zeros. The dimension of A is $N\times N$, and the dimension of the different vectors is $N\times 1$. As already mentioned, all the elements of A are zero except for 15 elements per row, which means that whereas the total number of elements in A increases with the number of grid points as N^2, the number of non-zero elements increases linearly with N. A is therefore a sparse matrix and becomes more sparse as the size of the problem increases. There are many very efficient numerical algorithms available for solving sparse matrix systems. Although it is not the scope of this work to review all of them or point out the most efficient one for a particular problem, one of these algorithms will be used in Section 6.4 as an example to what kind of performance can be achieved by the implementation of Eq. (6.24).

6.3.3 Retrieving the strain from the displacement field

Once the displacement field $\mathbf{u}$ has been obtained by solving Eq. (6.24), the strains must be calculated from $\mathbf{u}$. For the axial strain components, the calculation is straightforward and is obtained directly from Eqs. (6.9) and (6.11) at the natural grid sites (l,m,n). However, the shear strains are defined at sites other than the ones given by the natural grid. The value of the shear strains at those sites might be a valuable piece of information if the calculated strain components are going to be used as input for, say, an electronic structure calculation, or for the calculation of piezoelectric fields. Equations (6.10) and (6.12) do not give ϵ_4, ϵ_5 and ϵ_6 at (l,m,n) and an interpolation is needed to estimate those quantities. The simplest way to do this is by averaging over the four neighboring sites where the shear strains are defined. The "staggered" finite difference operator for shear strains at

[5] Obviously, when implementing this equation in practice, the indices need to be remapped from $(i;l,m,n)$ into one single index.

(l, m, n) can thus be expressed as

$$\begin{aligned}\Delta_j^{\dagger} u_i\,(l,m,n) \equiv \frac{1}{4h_j}\Bigg[& u_i\left(l+\tfrac{\delta_{xi}}{2}+\delta_{xj}, m+\tfrac{\delta_{yi}}{2}+\delta_{yj}, n+\tfrac{\delta_{zi}}{2}+\delta_{zj}\right) \\ & + u_i\left(l-\tfrac{\delta_{xi}}{2}+\delta_{xj}, m-\tfrac{\delta_{yi}}{2}+\delta_{yj}, n-\tfrac{\delta_{zi}}{2}+\delta_{zj}\right) \\ & - u_i\left(l+\tfrac{\delta_{xi}}{2}-\delta_{xj}, m+\tfrac{\delta_{yi}}{2}-\delta_{yj}, n+\tfrac{\delta_{zi}}{2}-\delta_{zj}\right) \\ & - u_i\left(l-\tfrac{\delta_{xi}}{2}-\delta_{xj}, m-\tfrac{\delta_{yi}}{2}-\delta_{yj}, n-\tfrac{\delta_{zi}}{2}-\delta_{zj}\right)\Bigg] \\ \text{with} \quad & i \neq j. \end{aligned} \tag{6.25}$$

Equation (6.25) can now be employed in Eq. (6.10) to estimate ϵ_4, ϵ_5 and ϵ_6 at (l, m, n). Note that even though there is an averaging involved in the present method for shear strains, the axial strains are calculated directly at the natural grid. Also, this average is calculated over only four neighbouring sites instead of the eight site method commonly employed in the literature [161, 166].

6.3.4 Implementation in rotated (111)-oriented zinc-blende systems

We have previously shown how the implementation of a rotation formalism for zinc-blende QDs oriented along the [111] direction of the lattice helps overcome a series of issues related to the size of the cell used in the calculation of strain and polarization fields [6]. In particular, the explicit form of the elastic energy obtained under the rotation convention used throughout Ref. [6] includes the dependence on an extra off-diagonal component of the elastic tensor $\tilde{C}_{15}$:

$$\begin{aligned} E_{(111)}^{\mathrm{ZB}} = & \frac{V_0}{2}\tilde{C}_{11}\left({\epsilon_1}^2+{\epsilon_2}^2\right) + \frac{V_0}{2}\tilde{C}_{33}{\epsilon_3}^2 + V_0\tilde{C}_{12}\epsilon_1\epsilon_2 + V_0\tilde{C}_{13}\left(\epsilon_1\epsilon_3+\epsilon_2\epsilon_3\right) \\ & + \frac{V_0}{2}\tilde{C}_{44}\left({\epsilon_4}^2+{\epsilon_5}^2\right) + \frac{V_0}{2}\tilde{C}_{66}{\epsilon_6}^2 + V_0\tilde{C}_{15}\epsilon_5\left(\epsilon_1-\epsilon_2\right) - V_0\tilde{C}_{15}\epsilon_4\epsilon_6, \end{aligned} \tag{6.26}$$

where the tilde symbols in the elastic constants indicate that they are the components of the transformed (rotated) elastic tensor. Except for the last two terms in Eq. (6.26), the elastic energy of the rotated (111)-oriented systems has the same form as the WZ one [Eq. (6.15)]. The similarities between WZ and (111)-oriented ZB systems have already been pointed out in Ref. [6] and Chapters 4 and 5. The method presently described can be applied to these systems provided that the extra terms in Eq. (6.26) are added to the terms already present in Eq. (6.20). In

the discrete description, they can be written as

$$\begin{aligned}\Delta E^{(111)}_{l,m,n} =& V_0^{l,m,n}\tilde{C}_{15}^{l,m,n}\epsilon_5\left(l,m,n\right)\left[\epsilon_1\left(l,m,n\right)-\epsilon_2\left(l,m,n\right)\right],\\ \Delta E^{(111)}_{l+\frac{1}{2},m+\frac{1}{2},n+\frac{1}{2}} =& -V_0^{l+\frac{1}{2},m+\frac{1}{2},n+\frac{1}{2}}\tilde{C}_{15}^{l+\frac{1}{2},m+\frac{1}{2},n+\frac{1}{2}}\epsilon_4\left(l+\frac{1}{2},m+\frac{1}{2},n+\frac{1}{2}\right)\\ &\times\epsilon_6\left(l+\frac{1}{2},m+\frac{1}{2},n+\frac{1}{2}\right).\end{aligned} \tag{6.27}$$

Equation (6.27) requires that ϵ_4, ϵ_5 and ϵ_6 be defined in grid sites other than the ones we have used in Section 6.2.2. Thus, they must be interpolated from the strains defined in our staggered grid:

$$\begin{aligned}\epsilon_5\left(l,m,n\right) =& \frac{1}{4}\Bigg[\epsilon_5\left(l+\frac{1}{2},m,n+\frac{1}{2}\right)+\epsilon_5\left(l+\frac{1}{2},m,n-\frac{1}{2}\right)\\ &+\epsilon_5\left(l-\frac{1}{2},m,n+\frac{1}{2}\right)+\epsilon_5\left(l-\frac{1}{2},m,n-\frac{1}{2}\right)\Bigg],\\ \epsilon_4\left(l+\frac{1}{2},m+\frac{1}{2},n+\frac{1}{2}\right) =& \frac{1}{2}\left[\epsilon_4\left(l,m+\frac{1}{2},n+\frac{1}{2}\right)+\epsilon_4\left(l+1,m+\frac{1}{2},n+\frac{1}{2}\right)\right],\\ \epsilon_6\left(l+\frac{1}{2},m+\frac{1}{2},n+\frac{1}{2}\right) =& \frac{1}{2}\left[\epsilon_6\left(l+\frac{1}{2},m+\frac{1}{2},n\right)+\epsilon_4\left(l+\frac{1}{2},m+\frac{1}{2},n+1\right)\right].\end{aligned} \tag{6.28}$$

For $V_0^{l+\frac{1}{2},m+\frac{1}{2},n+\frac{1}{2}}$ and $\tilde{C}_{15}^{l+\frac{1}{2},m+\frac{1}{2},n+\frac{1}{2}}$ the average over the values of V_0 and $\tilde{C}_{15}$ at the eight nearest sites of the natural grid must be taken. Equation (6.28) implies a dependence of the derivatives in Eq. (6.22) on a number of additional energy elements. The resultant extra coefficients must be included into the corresponding linear system [Eq. (6.24)].[6]

6.4 Testing the method

The procedure described in Section 6.3 has been implemented in our code HUSH,[7] together with the generalized minimum residual (GMRES) iteration method, with incomplete *LU* factorization preconditioning, as available from the DLAP linear algebra software library [169]. GMRES is an efficient method for solving linear

[6] For the rotated system, there is a total of 39 non-zero elements corresponding to the derivative of E_{total} with respect to u_x and 35 non-zero elements corresponding to its derivatives with respect to both u_y and u_z.

[7] The code is not currently publicly released. A precompiled version can be obtained from the author upon request.

systems, as the one given in Eq. (6.24), for sparse and asymmetric matrices. Figure 6.4 shows the results of a series of strain calculations for QD test systems. The first two rows [Figs. 6.4 (a) to (d)] correspond to strain profiles for the archetypical case of a pyramidal InAs QD grown in a GaAs matrix with a wetting layer, with periodic boundary conditions for the cell used in the calculation. Our results compare well with Grundmann *et al.* [161] and Pryor *et al.* [166]. The last two rows [Figs. 6.4 (e) to (h)] correspond to a (111)-oriented lens-shaped InAs/GaAs dot, similar to those studied in Ref. [6], with fixed boundary conditions [the interaction between the strain field and the edge of the box can be observed in Fig. 6.4 (h)]. The three-fold symmetry expected from this kind of system and discussed in detail by Schulz *et al.* [6] can be readily identified in the hydrostatic strain plot of Fig. 6.4(f). The results for the pyramidal QD were obtained through the implementation of the method for conventional zinc-blende systems described in Section 6.3, whereas the lens-shaped QD was treated under the "rotated" formalism described in Ref. [6] and the extra off-diagonal terms of the elastic energy were added as described in Section 6.3.4. Our code has been applied to accurately describe strain fields in a number of studies on QD systems. For instance, we have used it to survey the piezoelectric properties of ZB QDs [8], and to study non-polar wurtzite GaN/AlN QDs [7].

6.4.1 Performance

Finally we present a performance test to address the computational advantage of the method presented in this work. For comparison, we use a numerical simulation tool. The space discretization is done in a finite difference scheme and a strain energy minimization routine is included with the package. In both the purely numerical code and our hybrid code HUSH, the minimization procedure occurs in two steps. First, the system of linear equations in Eq. (6.24) needs to be set up. Computationally, this corresponds mostly to the calculation of the matrix A. Second, an iterative method for the solution of the system is used. Figure 6.5 shows a comparison between the time that it took each code to perform a test case calculation for the first part, that is, obtaining the system given in Eq. (6.24). The test case consisted of an InAs cubic QD of side 9 nm embedded in a cubic GaAs matrix of side 40 nm, and was repeated for different step sizes and, therefore, different numbers of grid points. For consistency, all the calculations were done on

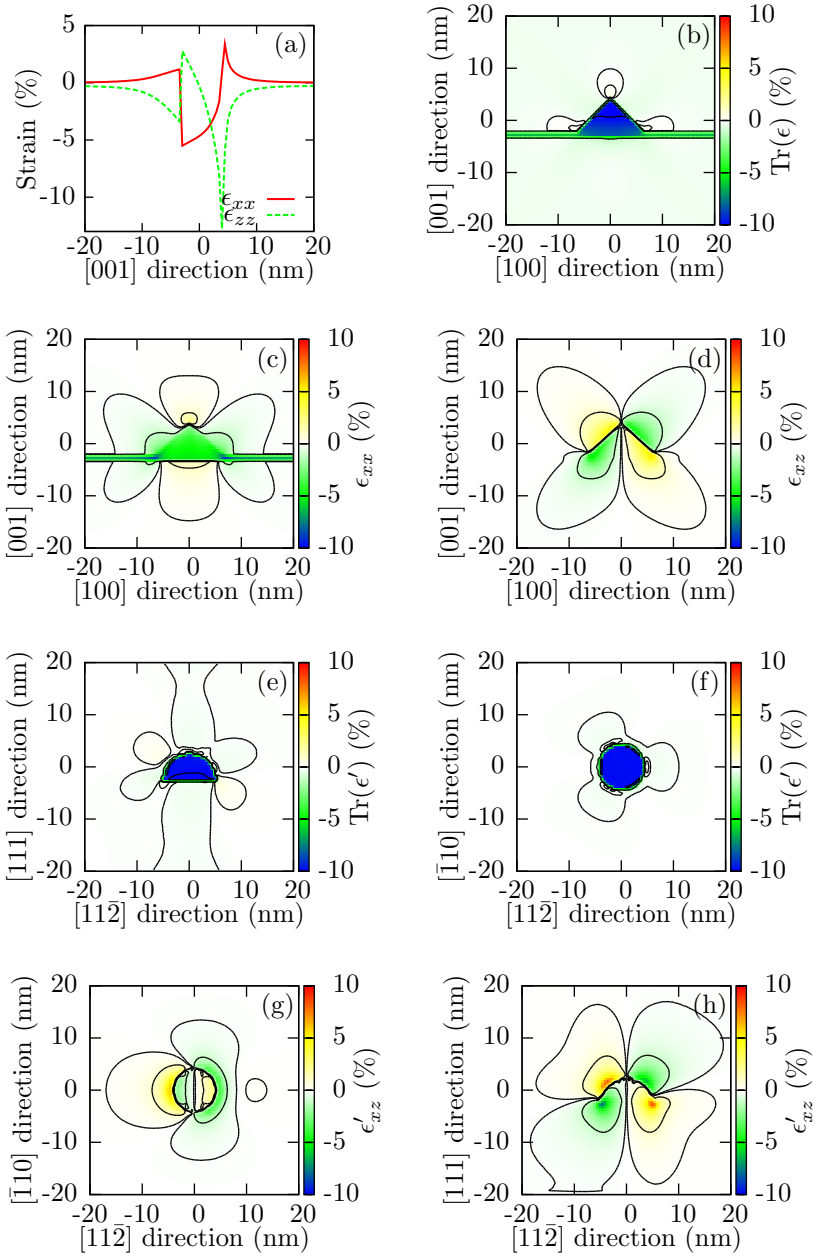


Figure 6.4: Test-case QD structures. (a–d) Pyramidal InAs/GaAs QD with wetting layer (WL), grown along the [001] direction; the sides of the pyramid base are 12 nm long and placed along [100] and [010]; the height of the pyramid is 6 nm and the WL is 1 nm thick. (e–h) Lens-shaped (half sphere) InAs/GaAs QD grown along the [111] direction; the base radius is 4.5 nm; the hydrostatic components Tr(ϵ') in (e) and (f) are independent of the coordinate system, (g) and (h) are given in the rotated coordinate convention from Ref. [6].

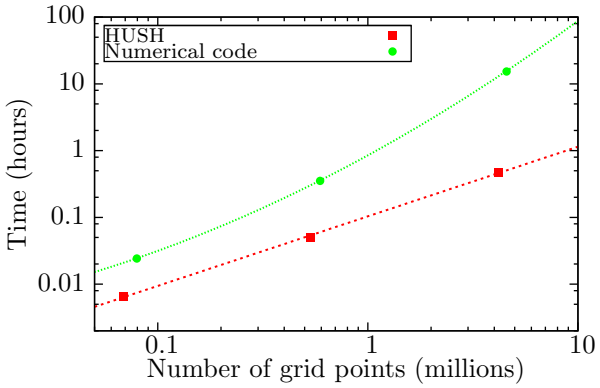


Figure 6.5: Comparison between our code HUSH and another code that calculates the derivatives numerically. Dependence of the computational time, required for obtaining the linear system of Eq. (6.24), on the number of grid points contained within the simulation cell.

a single core within the same cluster computer, with the same amount of available memory. It can be seen how our method scales almost linearly with the number of points in the simulation cell whereas the numerical evaluation of the derivatives involved in Eq. (6.24) performed by the numerical code scaled exponentially for the range shown. This non-linear scaling implies that, for large systems with about 10 million grid points, doing the first part of the calculation numerically takes about 100 times longer than the analytical evaluation that we have presented.

The performance achieved during the second part of the calculation is highly dependent on the particular iterative method chosen and the convergence criteria, as well as on the size of A, and will not be reviewed here. During our testing both codes achieved similar performances (for similar accuracies) in this aspect, as the size of the sparse matrix A tends to be the same in both cases. In general we found that whereas for HUSH this second step accounted for the greater portion of the total computational time, for the other code the calculation of A took much longer than the iterative solving process.

6.5 Summary

We have presented a finite difference staggered grid formulation of the strain energy minimization problem, followed by an analytical evaluation of the condition of minimum for the elastic energy of conventional zinc-blende and wurtzite systems in the [001] and [0001] representations, respectively, as well as the special

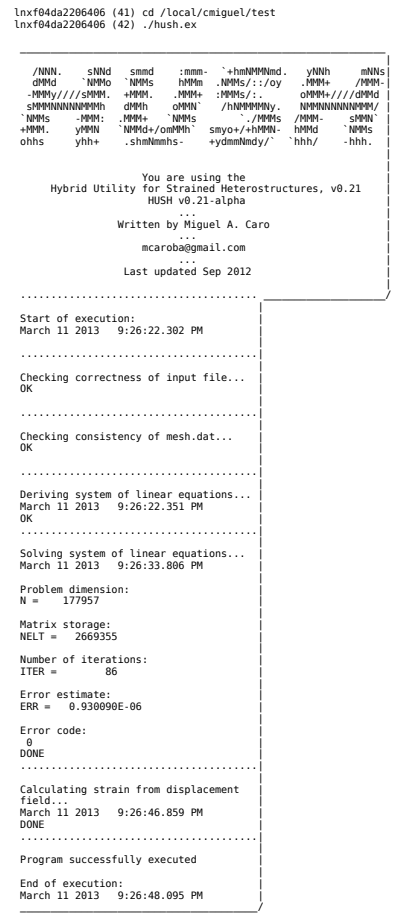


Figure 6.6: Example execution of HUSH on a Linux terminal for a test system with ∼178k grid elements.

case of (111)-oriented ZB systems. The staggered grid formulation accounts for an elegant way to overcome problems associated with conventional finite differences leading to averaging and ill-defined boundaries, while the analytical evaluation of the derivatives of the elastic energy with respect to the displacement field provides a computationally very efficient derivation of the system of linear equations to be solved by the minimization procedure. Our method allows an improvement in the computational times for this task, becoming significantly faster for large simulation cells containing millions of grid points, compared to a conventional purely numerical calculation. This is of relevance for alleviating the computational task related to calculation of strain fields, and therefore for improving the overall performance of calculations of the electronic and optical properties of semiconductor nanostructures. This is especially true for calculations that require a high degree of

accuracy, and therefore a high degree of grid discretization. The results presented here have been obtained for the ZB and WZ crystal lattices, commonly found in semiconductor materials. Nevertheless, the method described in Section 6.3 can be applied to any crystal system, in general. For any lattice with a lower symmetry compatible with either ZB or WZ our results are readily applicable. We have implemented the present method in our code HUSH (Fig. 6.6), which is freely available for the scientific community to use, upon request to the author.

7 Conclusions

7.1 Summary

In this thesis we have presented a theoretical study of elasticity and electric polarization in strained tetrahedrally bonded semiconductors. Because of their technological importance in the field of optoelectronic applications, we have focused our attention on group-III nitrides. The approach taken has been that of multiscale modelling, where calculations performed at different levels of theory take input from one another.

In Chapter 2 we have given an introduction to some of the basic theory and methods used throughout this work: elasticity and piezoelectricity in strained crystals, density functional theory (DFT) and Berry-phase theory of electric polarization. We also provide a practical guide to Berry-phase calculations, which seems to be somehow missing from the literature.

As we outlined in Chapter 1, one of the main issues regarding nanostructures incorporating nitride materials is the existence of strong polarization fields. Our aim in this thesis has been to build a complete description of piezoelectricity in binary and alloyed group-III nitrides, starting by understanding the underlying cause: macroscopic and internal strain. We have worked within two different descriptions, continuum theory and atomistic calculations. Continuum calculations take as input "effective" parameters that are obtained as average properties of atomistic calculations. At the same time, atomistic calculations for large systems rely on the construction of simplified phenomenological models that match (as closely as possible) full quantum-mechanical descriptions performed for smaller systems.

The continuum approximation is appropriate when atomistic calculations are computationally unaffordable, when the atomic effects are of secondary importance, to gain first insights into general behaviours or to identify trends. In this

context, Chapter 3 has dealt with the issue of built-in electrostatic fields in group-III nitride heterostructures. We have shown how composition engineering of the barrier and active layers enables built-in field reduction, and even suppression, for a wide range of quantum well (QW) and quantum dot (QD) nitride structures. In that chapter we have also discussed the dependence of the results on the AlInN band gap bowing parameter, whose value remains to date under discussion. The lack of accuracy in the description of electric polarization in nitride alloys encountered during the development of that work inspired and motivated much of the subsequent studies we carried out. We used analytical models for strain that provided extra speed-up for the calculations, at the cost of reducing the accuracy and flexibility (e.g. regarding QD shape) of the procedure. However, purely numerical calculations are computationally expensive, and more so in the case of the optimization procedure we employed for field control. This motivated the development of a hybrid numerical/analytical formulation of the elastic energy minimization problem, which we have presented in Chapter 6. We used a series of staggered grids, in order to also improve the resolution and accuracy with respect to other finite-difference numerical methods. We have used this code in a number of published works [6–8].

Within the atomistic description, we have presented a DFT study of elasticity and internal strain in tetrahedrally bonded compounds in Chapter 4. Additionally, we have reviewed the two first-principles methods commonly employed in the calculation of elastic constants of crystalline solids: based on total energy and stresses. We have identified problems related to basis set inconsistency in the total energy method, and proposed the stress method (or a basis-corrected total energy method) as a safer route towards determination of elastic constants. We have also presented the correct internal strain vectors for the wurtzite (WZ) crystal structure and reported for the first time a complete set of WZ internal strain parameters for the nitrides. Our treatment of internal strain has enabled us to develop a theory of local electric polarization in Chapter 5, which is possibly this thesis's main contribution to solid-state physics. This theory allows access to the spatial variation of electric polarization within the supercell used in the calculation, that is not accessible through a formal Berry-phase calculation. Our theory is exact in the linear limit and has been validated against Berry-phase calculations for the average polarization of nitride alloy supercells, showing excellent agreement for finite-strain calculations. Our formalism has allowed us to establish the local

piezoelectric tensor of the WZ crystal structure, which bears resemblance with that of rotated [111] zinc-blende (ZB) systems. Within this theory, it is possible to build a spatial map of polarization in semiconductor nanostructures, obtaining far more accuracy *and* specificity than continuum calculations because the statistical alloy fluctuations are explicitly taken into account, and not simply averaged out. In this context, we have developed a "point dipole" technique to compute the local polarization potential in the irregular atomic grid. Using atomistic valence force field (VFF) interatomic potentials derived from *ab initio* elasticity calculations, the local polarization theory and the point dipole technique have enabled us to calculate the local polarization potential in bulk nitride alloys and nanostructures. We are only starting to understand the effects of this local polarization on the electronic structure of nitride materials, which have not been reported in this thesis. Our most recent findings are summarized in the next section.

7.2 Ongoing and future work

There are a number of works that because of time constraints we have not been able to include in this thesis. In particular, we have carried out comprehensive work on VFF models, which has led to the prediction of the correct internal strain vectors and the five independent internal strain parameters of the WZ crystal structure presented in Chapter 4. We have also constructed accurate VFFs for the calculation of local strain and local polarization, as shown in Chapter 5 for some nitride QW examples. The most immediate developments will precisely concern the publication of our local polarization theory, together with a number of studies on how the local polarization associated to alloy fluctuations affects the electronic properties of group-III nitrides. In this regard, we have used the tight-binding approximation to calculate the electronic structure of bulk InGaN. Our results indicate that the local polarization potential plays a significant role in the localization of hole states, leading to a pronounced upward bowing of the valence band edge of InGaN. These results are also significant in the context of research on AlInN, for which the difference in band gap and bond lengths between the binary constituents is even larger than for InGaN. As we showed in Chapter 3, AlInN is of great interest regarding its use for UV devices, such as UV LEDs. However, it might significantly suffer from carrier localization effects and future studies should aim at exploring the effects of the local polarization potential on the optoelectronic

properties of these devices.

We are also preparing a manuscript on second-order polarization in the III-Vs, and future work shall extend our local theory to second order. This will allow us to assess the importance of local polarization effects in nanostructures where the piezoelectric compounds employed are substantially less ionic than the III-N, for instance InGaAs QDs.

8 Bibliography

[1] M. A. Caro, S. Schulz, S. B. Healy, and E. P. O'Reilly. Built-in field control in alloyed *c*-plane III-N quantum dots and wells. *J. Appl. Phys.*, 109:084110, 2011.

[2] M. A. Caro, S. Schulz, S. B. Healy, and E. P. O'Reilly. Built-in field control in nitride nanostructures operating in the UV. *Phys. Stat. Sol. C*, 9:838, 2012.

[3] M. A. Caro, S. Schulz, and E. P. O'Reilly. Hybrid functional study of the elastic and structural properties of wurtzite and zinc-blende group-III nitrides. *Phys. Rev. B*, 86:014117, 2012.

[4] M. A. Caro, S. Schulz, and E. P. O'Reilly. Comparison of stress and total energy methods for calculation of elastic properties of semiconductors. *J. Phys.: Condens. Matter*, 25:025803, 2013.

[5] M. A. Caro, S. Schulz, and E. P. O'Reilly. Effect of alloy fluctuations on the local polarization in nitride nanostructures. *Phys. Stat. Sol. B*, 249:526, 2012.

[6] S. Schulz, M. A. Caro, E. P. O'Reilly, and O. Marquardt. Symmetry-adapted calculations of strain and polarization fields in (111)-oriented zinc-blende quantum dots. *Phys. Rev. B*, 84:125312, 2011.

[7] S. Schulz, M. A. Caro, and E. P. O'Reilly. Prediction of strong ground state electron and hole wave function spatial overlap in nonpolar GaN/AlN quantum dots. *Appl. Phys. Lett.*, 101:113107, 2012.

[8] S. Schulz, M. A. Caro, E. P. O'Reilly, and O. Marquardt. Piezoelectric properties of zinc blende quantum dots. *Phys. Stat. Sol. B*, 249:521, 2012.

[9] H. A. Al-Brithen, A. R. Smith, and D. Gall. Surface and bulk electronic structure of ScN (001) investigated by scanning tunneling microscopy/spectroscopy and optical absorption spectroscopy. *Phys. Rev. B*, 70:045303, 2004.

[10] M. A. Moram, Z. H. Barber, C. J. Humphreys, T. B. Joyce, and P. R. Chalker. Young's modulus, Poisson's ratio, and residual stress and strain in (111)-oriented scandium nitride thin films on silicon. *J. Appl. Phys.*, 100:023514, 2006.

[11] V. Ranjan, L. Bellaiche, and E. J. Walter. Strained hexagonal ScN: A material with unusual structural and optical properties. *Phys. Rev. Lett.*, 90:257602, 2003.

[12] C. Constantin, H. Al-Brithen, M. B. Haider, D. Ingram, and A. R. Smith. ScGaN alloy growth by molecular beam epitaxy: Evidence for a metastable layered hexagonal phase. *Phys. Rev. B*, 70:193309, 2004.

[13] C. Constantin, M. B. Haider, D. Ingram, A. R. Smith, N. Sandler, K. Sun, and P. Ordejón. Composition-dependent structural properties in ScGaN alloy films: A combined experimental and theoretical study. *J. Appl. Phys.*, 98:123501, 2005.

[14] M. Akiyama, T. Kamohara, K. Kano, A. Teshigahara, Y. Takeuchi, and N. Kawahara. Enhancement of piezoelectric response in scandium aluminum nitride alloy thin films prepared by dual reactive cosputtering. *Adv. Mater.*, 21:593, 2009.

[15] A. Rubio, J. L. Corkill, and M. L. Cohen. Theory of graphitic boron nitride nanotubes. *Phys. Rev. B*, 49:5081, 1994.

[16] N. G. Chopra, R. J. Luyken, K. Cherrey, V. H. Crespi, M. L. Cohen, S. G. Louie, and A. Zettl. Boron nitride nanotubes. *Science*, 269:966, 1995.

[17] A. Zaoui. Plane wave pseudopotential study of ground state properties and electrochemical description of thallium nitride. *Mat. Sci. Eng. B*, 103:258, 2003.

[18] J. P. Mulkey and F. W. Oehme. A review of thallium toxicity. *Vet. Hum. Toxicol.*, 35:445, 1993.

[19] I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan. Band parameters for III-V compound semiconductors and their alloys. *J. Appl. Phys.*, 89:5815, 2001.

[20] I. Vurgaftman and J. R. Meyer. Band parameters for nitrogen-containing semiconductors. *J. Appl. Phys.*, 94:3675, 2003.

[21] J. Wu. When group-III nitrides go infrared: New properties and perspectives. *J. Appl. Phys.*, 106:011101, 2009.

[22] S. V. Novikov, N. M. Stanton, R. P. Campion, R. D. Morris, H. L. Geen, C. T. Foxon, and A. J. Kent. Growth and characterization of free-standing zinc-blende (cubic) GaN layers and substrates. *Semicond. Sci. Technol.*, 23:015018, 2008.

[23] M. Yamada, T. Mitani, Y. Narukawa, S. Shioji, I. Niki, S. Sonobe, K. Deguchi, M. Sano, and T. Mukai. InGaN-based near-ultraviolet and blue-light-emitting diodes with high external quantum efficiency using a patterned sapphire substrate and a mesh electrode. *Jpn. J. Appl. Phys.*, 41:L1431, 2002.

[24] M. R. Krames, O. B. Shchekin, R. Mueller-Mach, G. O. Mueller, L. Zhou, G. Harbers, and M. G. Craford. Status and future of high-power light-emitting diodes for solid-state lighting. *J. Disp. Technol.*, 3:160, 2007.

[25] S. Nakamura, S. J. Pearton, and G. Fasol. *The blue laser diode: the complete story.* Springer Verlag, 2000.

[26] C. J. Humphreys. Solid-state lighting. *MRS Bull.*, 33:459, 2008.

[27] M. A. Khan, M. Shatalov, H. P. Maruska, H. M. Wang, and E. Kuokstis. III-nitride UV devices. *Jpn. J. Appl. Phys.*, 44:7191, 2005.

[28] U. K. Mishra, S. Likun, T. E. Kazior, and Y. F. Wu. GaN-based RF power devices and amplifiers. *P. IEEE*, 96:2875, 2008.

[29] Y. Kuwahara, T. Fujii, T. Sugiyama, D. Iida, Y. Isobe, Y. Fujiyama, Y. Morita, M. Iwaya, T. Takeuchi, S. Kamiyama, I. Akasaki, and H. Amano.

GaInN-based solar cells using strained-layer GaInN/GaInN superlattice active layer on a freestanding GaN substrate. *Appl. Phys. Express*, 4:021001, 2011.

[30] H. K. Cho, J. Y. Lee, G. M. Yang, and C. S. Kim. Formation mechanism of V defects in the InGaN/GaN multiple quantum wells grown on GaN layers with low threading dislocation density. *Appl. Phys. Lett.*, 79:215, 2001.

[31] B. Beaumont, Ph. Vennéguès, and P. Gibart. Epitaxial lateral overgrowth of GaN. *Phys. Stat. Sol. B*, 227:1, 2001.

[32] C. J. Humphreys. Does in form In-rich clusters in InGaN quantum wells? *Philos. Mag.*, 87:1971, 2007.

[33] S. Chichibu, K. Wada, and S. Nakamura. Spatially resolved cathodoluminescence spectra of InGaN quantum wells. *Appl. Phys. Lett.*, 71:2346, 1997.

[34] H. Schömig, S. Halm, A. Forchel, G. Bacher, J. Off, and F. Scholz. Probing individual localization centers in an InGaN/GaN quantum well. *Phys. Rev. Lett.*, 92:106802, 2004.

[35] P. T. Barletta, E. A. Berkman, B. F. Moody, N. A. El-Masry, A. M. Emara, M. J. Reed, and S. M. Bedair. Development of green, yellow, and amber light emitting diodes using InGaN multiple quantum well structures. *Appl. Phys. Lett.*, 90:151109, 2007.

[36] C. Hums, J. Bläsing, A. Dadgar, A. Diez, T. Hempel, J. Christen, A. Krost, K. Lorenz, and E. Alves. Metal-organic vapor phase epitaxy and properties of AlInN in the whole compositional range. *Appl. Phys. Lett.*, 90:022105, 2007.

[37] T. Matsuoka. Progress in nitride semiconductors from GaN to InN—MOVPE growth and characteristics. *Superlattices and Microstructures*, 37:19, 2005.

[38] C. Skierbiszewski, Z. R. Wasilewski, I. Grzegory, and S. Porowski. Nitride-based laser diodes by plasma-assisted MBE – from violet to green emission. *J. Cryst. Growth*, 311:1632, 2009.

[39] H. Ibach and H. Lüth. *Solid-State Physics: An Introduction to Principles of Materials Science*. Springer-Verlag, Berlin Heidelberg, 3rd edition, 2003.

[40] L. D. Landau, L. P. Pitaevskii, A. M. Kosevich, and E. M. Lifshitz. *Theory of elasticity*, volume 7. Elsevier Butterworth-Heinemann, 3rd edition, 1986.

[41] J. F. Nye. *Physical properties of crystals: their representation by tensors and matrices*. Oxford University Press, 1985.

[42] H. Grimmer. The piezoelectric effect of second order in stress or strain: its form for crystals and quasicrystals of any symmetry. *Acta Crystallogr. Sect. A: Foundations of Crystallography*, 63:441, 2007.

[43] R. M. Martin. *Electronic Structure*. Cambridge University Press, 2004.

[44] J. P. Perdew and M. Levy. Physical content of the exact Kohn-Sham orbital energies: Band gaps and derivative discontinuities. *Phys. Rev. Lett.*, 51:1884, 1983.

[45] J. D. Jackson. *Classical Electrodynamics*. John Wiley & Sons, New York, third edition, 1999.

[46] M. Born and R. Oppenheimer. Zur quantentheorie der molekeln. *Annalen der Physik*, 389:457, 1927.

[47] L. H. Thomas. The calculation of atomic fields. *Math. Proc. Cambridge Philos. Soc.*, 23:542, 1927.

[48] E. Fermi. Un metodo statistico per la determinazione di alcune priorietà dell'atome. *Rend. Accad. Naz. Lincei*, 6:32, 1927.

[49] P. A. M. Dirac. Note on exchange phenomena in the Thomas atom. *Proc. Cambridge Philos. Soc.*, 26:376, 1930.

[50] M. E. Casida and T. A. Wesołowski. Generalization of the Kohn-Sham equations with constrained electron density formalism and its time-dependent response theory formulation. *Int. J. Quantum Chem.*, 96:577, 2004.

[51] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133, 1965.

[52] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136:B864, 1964.

[53] J. P. Perdew and A. Zunger. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B*, 23:5048, 1981.

[54] D. M. Ceperley and B. J. Alder. Ground state of the electron gas by a stochastic method. *Phys. Rev. Lett.*, 45:566, 1980.

[55] J. P. Perdew, K. Burke, and M. Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865, 1996.

[56] E. H. Lieb and S. Oxford. Improved lower bound on the indirect Coulomb energy. *Int. J. Quantum Chem.*, 19:427, 1981.

[57] A. Zoroddu, F. Bernardini, P. Ruggerone, and V. Fiorentini. First-principles prediction of structure, energetics, formation enthalpy, elastic constants, polarization, and piezoelectric constants of AlN, GaN, and InN: Comparison of local and gradient-corrected density-functional theory. *Phys. Rev. B*, 64:045208, 2001.

[58] A. Beya-Wakata, P.-Y. Prodhomme, and G. Bester. First- and second-order piezoelectricity in III-V semiconductors. *Phys. Rev. B*, 84:195207, 2011.

[59] L. J. Sham and M. Schlüter. Density-functional theory of the energy gap. *Phys. Rev. Lett.*, 51:1888, 1983.

[60] L. J. Sham and M. Schlüter. Density-functional theory of the band gap. *Phys. Rev. B*, 32:3883, 1985.

[61] A. Seidl, A. Görling, P. Vogl, J. A. Majewski, and M. Levy. Generalized Kohn-Sham schemes and the band-gap problem. *Phys. Rev. B*, 53:3764, 1996.

[62] J. Heyd, G. E. Scuseria, and M. Ernzerhof. Hybrid functionals based on a screened Coulomb potential. *J. Chem. Phys.*, 118:8207, 2003.

[63] J. Heyd and G. E. Scuseria. Efficient hybrid density functional calculations in solids: assessment of the Heyd-Scuseria-Ernzerhof screened Coulomb hybrid functional. *J. Chem. Phys.*, 121:1187, 2004.

[64] T. M. Henderson, J. Paier, and G. E. Scuseria. Accurate treatment of solids with the HSE screened hybrid. *Phys. Stat. Sol. B*, 248:767, 2011.

[65] J. Paier, M. Marsman, K. Hummer, G. Kresse, I. C. Gerber, and J. G. Angyan. Screened hybrid density functionals applied to solids. *J. Chem. Phys.*, 124:154709, 2006.

[66] J. Paier, M. Marsman, K. Hummer, G. Kresse, I. C. Gerber, and J. G. Angyan. Erratum: "Screened hybrid density functionals applied to solids" [J. Chem. Phys. 124, 154709 (2006)]. *J. Chem. Phys.*, 125:249901, 2006.

[67] M. A. Caro, S. Schulz, and E. P. O'Reilly. Erratum: Hybrid functional study of the elastic and structural properties of wurtzite and zinc-blende group-III nitrides [Phys. Rev. B 86, 014117 (2012)]. *Phys. Rev. B*, 86:099901, 2012.

[68] P. E. Blöchl. Projector augmented-wave method. *Phys. Rev. B*, 50:17953, 1994.

[69] Quantum Espresso, available online at [http://www.quantum-espresso.org].

[70] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys.: Condens. Matter*, 21:395502, 2009.

[71] V. Fiorentini, M. Methfessel, and M. Scheffler. Electronic and structural properties of GaN by the full-potential linear muffin-tin orbitals method: The role of the *d* electrons. *Phys. Rev. B*, 47:13353, 1993.

[72] Vienna *Ab initio* Simulation Package (VASP), [http://www.vasp.at], see also on line documentation.

[73] D. Vanderbilt. Soft self-consistent pseudopotentials in a generalized eigenvalue formalism. *Phys. Rev. B*, 41:7892, 1990.

[74] G. Kresse and D. Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B*, 59:1758, 1999.

[75] G. Kresse and J. Furthmüller. Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set. *Phys. Rev. B*, 54:11169, 1996.

[76] C.-L. Fu and K.-M. Ho. First-principles calculation of the equilibrium ground-state properties of transition metals: Applications to Nb and Mo. *Phys. Rev. B*, 28:5480, 1983.

[77] R. Resta and D. Vanderbilt. *Physics of Ferroelectrics: a Modern Perspective*, chapter Theory of Polarization: A Modern Approach. Springer-Verlag, Berlin Heidelberg, 2007.

[78] R. D. King-Smith and D. Vanderbilt. Theory of polarization of crystalline solids. *Phys. Rev. B*, 47:1651, 1993.

[79] D. Vanderbilt and R. D. King-Smith. Electric polarization as a bulk quantity and its relation to surface charge. *Phys. Rev. B*, 48:4442, 1993.

[80] R. Resta. Theory of the electric polarization in crystals. *Ferroelectrics*, 136:51, 1992.

[81] R. Resta. Macroscopic polarization in crystalline dielectrics: the geometric phase approach. *Rev. Mod. Phys.*, 66:899, 1994.

[82] F. Bernardini, V. Fiorentini, and D. Vanderbilt. Spontaneous polarization and piezoelectric constants of III-V nitrides. *Phys. Rev. B*, 56:10024, 1997.

[83] M. V. Berry. Quantal phase factors accompanying adiabatic changes. *Proc. R. Soc. Lond. A*, 392:45, 1984.

[84] M. Zhu, S. You, T. Detchprohm, T. Paskova, E. A. Preble, and C. Wetzel. Various misfit dislocations in green and yellow GaInN/GaN light emitting diodes. *phys. stat. sol. (a)*, 207:1305, 2010.

[85] Y. Taniyasu, M. Kasu, and T. Makimoto. An aluminium nitride light-emitting diode with a wavelength of 210 nanometres. *Nature*, 441:325, 2006.

[86] J. R. Grandusky, J. A. Smart, M. C. Mendrick, L. J. Schowalter, K. X. Chen, and E. F. Schubert. Pseudomorphic growth of thick n-type $Al_xGa_{1-x}N$ layers on low-defect-density bulk AlN substrates for UV LED applications. *J. Crystal Growth*, 311:2864, 2009.

[87] S. V. Novikov, C. T. Foxon, T. S. Cheng, T. L. Tansley, J. W. Orton, D. E. Lacklison, D. Johnston, N. Baba-Ali, S. E. Hooper, L. C. Jenkins, and L. Eaves. Auger investigation of group III nitride films grown by molecular beam epitaxy. *J. Crystal Growth*, 146:340, 1995.

[88] K. Kim, W. R. L. Lambrecht, and B. Segall. Elastic constants and related properties of tetrahedrally bonded BN, AlN, GaN, and InN. *Phys. Rev. B*, 53:16310, 1996.

[89] O. Ambacher, J. Majewski, C. Miskys, A. Link, M. Hermann, M. Eickhoff, M. Stutzmann, F. Bernardini, V. Fiorentini, V. Tilak, B. Schaff, and L. F. Eastman. Pyroelectric properties of Al(In)GaN/GaN hetero- and quantum well structures. *J. Phys.: Condens. Matter*, 14:3399, 2002.

[90] T. Paskova. Development and prospects of nitride materials and devices with nonpolar surfaces. *phys. stat. sol. (b)*, 245:1011, 2008.

[91] T. Takeuchi, S. Sota, M. Katsuragawa, M. Komori, H. Takeuchi, H. Amano, and I. Akasaki. Quantum-confined Stark effect due to piezoelectric fields in GaInN strained quantum wells. *Jpn. J. Appl. Phys.*, 36:L382, 1997.

[92] Y. C. Shen, G. O. Mueller, S. Watanabe, N. F. Gardner, A. Munkholm, and M. R. Krames. Auger recombination in InGaN measured by photoluminescence. *Appl. Phys. Lett.*, 91:141101, 2007.

[93] B. Monemar and B. E. Sernelius. Defect related issues in the "current roll-off" in InGaN based light emitting diodes. *Appl. Phys. Lett.*, 91:181103, 2007.

[94] M.-H. Kim, M. F. Schubert, Q. Dai, J. K. Kim, E. F. Schubert, J. Piprek, and Y. Park. Origin of efficiency droop in GaN-based light-emitting diodes. *Appl. Phys. Lett.*, 91:183507, 2007.

[95] M. F. Schubert and E. F. Schubert. Effect of heterointerface polarization charges and well width upon capture and dwell time for electrons and holes above GaInN/GaN quantum wells. *Appl. Phys. Lett.*, 96:131102, 2010.

[96] A. D. Andreev and E. P. O'Reilly. Optical transitions and radiative lifetime in GaN/AlN self-organized quantum dots. *Appl. Phys. Lett.*, 79:521, 2001.

[97] N. Baer, S. Schulz, S. Schumacher, P. Gartner, G. Czycholl, and F. Jahnke. Optical properties of self-organized wurtzite InN/GaN quantum dots: A combined atomistic tight-binding and full configuration interaction calculation. *Appl. Phys. Lett.*, 87:231114, 2005.

[98] H. J. Chung, R. J. Choi, M. H. Kim, J. W. Han, Y. M. Park, Y. S. Kim, H. S. Paek, C. S. Sone, Y. J. Park, J. K. Kim, and E. F. Schubert. Improved performance of GaN-based blue light emitting diodes with InGaN/GaN multilayer barriers. *Appl. Phys. Lett.*, 95:241109, 2009.

[99] J. Xu, M. F. Schubert, A. N. Noemaun, D. Zhu, J. K. Kim, E. F. Schubert, M. H. Kim, H. J. Chung, S. Yoon, C. Sone, and Y. Park. Reduction in efficiency droop, forward voltage, ideality factor, and wavelength shift in polarization-matched GaInN/GaInN multi-quantum-well light-emitting diodes. *Appl. Phys. Lett.*, 94:011113, 2009.

[100] W. Lee, M.-H. Kim, D. Zhu, A. N. Noemaun, J. K. Kim, and E. F. Schubert. Growth and characteristics of GaInN/GaInN multiple quantum well light-emitting diodes. *J. Appl. Phys.*, 107:063102, 2010.

[101] D. P. Williams, A. D. Andreev, E. P. O'Reilly, and D. A. Faux. Derivation of built-in polarization potentials in nitride-based semiconductor quantum dots. *Phys. Rev. B*, 72:235318, 2005.

[102] S. Schulz, A. Berube, and E. P. O'Reilly. Polarization fields in nitride-based quantum dots grown on nonpolar substrates. *Phys. Rev. B*, 79:081401(R), 2009.

[103] H. Shen, M. Wraback, H. Zhong, A. Tyagi, S. P. DenBaars, S. Nakamura, and J. S. Speck. Unambiguous evidence of the existence of polarization field crossover in a semipolar InGaN/GaN single quantum well. *Appl. Phys. Lett.*, 95:033503, 2009.

[104] K. Shimada. First-principles determination of piezoelectric stress and strain constants of wurtzite III-V nitrides. *Jpn. J. Appl. Phys.*, 45:L358, 2006.

[105] J.-M. Wagner and F. Bechstedt. Properties of strained wurtzite GaN and AlN: *Ab initio* studies. *Phys. Rev. B*, 66:115202, 2002.

[106] J. Furthmüller, P. H. Hahn, F. Fuchs, and F. Bechstedt. Band structures and optical spectra of InN polymorphs: Influence of quasiparticle and excitonic effects. *Phys. Rev. B*, 72:205106, 2005.

[107] F. Bernardini and V. Fiorentini. Nonlinear behavior of spontaneous and piezoelectric polarization in III-V nitride alloys. *phys. stat. sol. (a)*, 190:65, 2002.

[108] M. Winkelnkemper, A. Schliwa, and D. Bimberg. Interrelation of structural and electronic properties in $In_xGa_{1-x}N$/GaN quantum dots using an eight-band k·p model. *Phys. Rev. B*, 74:155322, 2006.

[109] F. Bernardini and V. Fiorentini. Nonlinear macroscopic polarization in III-V nitride alloys. *Phys. Rev. B*, 64:085207, 2001.

[110] T. H. Glisson, J. R. Hauser, M. A. Littlejohn, and C. K. Williams. Energy bandgap and lattice constant contours of III-V quaternary alloys. *J. Electron. Mater.*, 7:1, 1978.

[111] J.-R. Chen, S.-C. Ling, H.-M. Huang, P.-Y. Su, T.-S. Ko, T.-C. Lu, H.-C. Kuo, Y.-K. Kuo, and S.-C. Wang. Numerical study of optical properties of InGaN multi-quantum-well laser diodes with polarization-matched AlInGaN barrier layers. *Appl. Phys. B*, 95:145, 2009.

[112] V. Fiorentini, F. Bernardini, and O. Ambacher. Evidence for nonlinear macroscopic polarization in III-V nitride alloy heterostructures. *Appl. Phys. Lett.*, 80:1204, 2002.

[113] Q. Yan, P. Rinke, M. Scheffler, and C. G. Van de Walle. Strain effects in group-III nitrides: Deformation potentials for AlN, GaN, and InN. *Appl. Phys. Lett.*, 95:121111, 2009.

[114] S. Schulz, S. Schumacher, and G. Czycholl. Spin-orbit coupling and crystal-field splitting in the electronic and optical properties of nitride quantum dots with a wurtzite crystal structure. *Eur. Phys. J. B*, 64:51, 2008.

[115] J. R. Downes, D. A. Faux, and E. P. O'Reilly. A simple method for calculating strain distributions in quantum dot structures. *J. Appl. Phys.*, 81:6700, 1997.

[116] J. H. Davies. Elastic and piezoelectric fields around a buried quantum dot: A simple picture. *J. Appl. Phys.*, 84:1358, 1998.

[117] H. Zhang, E. J. Miller, E. T. Yu, C. Poblenz, and J. S. Speck. Measurement of polarization charge and conduction-band offset at $In_xGa_{1-x}N$/GaN heterojunction interfaces. *Appl. Phys. Lett.*, 84:4644, 2004.

[118] S. Nicolay, J.-F. Carlin, E. Feltin, R. Butté, M. Mosca, N. Grandjean, M. Ilegems, M. Tchernycheva, L. Nevou, and F. H. Julien. Midinfrared intersubband absorption in lattice-matched AlInN/GaN multiple quantum wells. *Appl. Phys. Lett.*, 87:111106, 2005.

[119] P. D. C. King, T. D. Veal, C. E. Kendrick, L. R. Bailey, S. M. Durbin, and C. F. McConville. InN/GaN valence band offset: High-resolution x-ray photoemission spectroscopy measurements. *Phys. Rev. B*, 78:033308, 2008.

[120] S. Schulz, T. J. Badcock, M. A. Moram, P. Dawson, M. J. Kappers, C. J. Humphreys, and E. P. O'Reilly. Electronic and optical properties of nonpolar *a*-plane GaN quantum wells. *Phys. Rev. B*, 82:125318, 2010.

[121] P. G. Moses and C. G. Van de Walle. Band bowing and band alignment in InGaN alloys. *Appl. Phys. Lett.*, 96:021908, 2010.

[122] S. L. Chuang and C. S. Chang. A band-structure model of strained quantum-well wurtzite semiconductors. *Semicond. Sci. Technol.*, 12:252, 1997.

[123] C. Hums, A. Gadanecz, A. Dadgar, J. Bläsing, P. Lorenz, S. Krischok, F. Bertram, A. Franke, J. A. Schaefer, J. Christen, and A. Krost. AlIn/GaN based multi quantum well structures - growth and optical properties. *phys. stat. sol. (c)*, 6:S451, 2009.

[124] D. P. Williams, S. Schulz, A. D. Andreev, and E. P. O'Reilly. Theory of GaN quantum dots for optical applications. *IEEE J. Sel. Top. Quantum Electron.*, 15:1092, 2009.

[125] S. Schulz and E. P. O'Reilly. Theory of reduced built-in polarization field in nitride-based quantum dots. *Phys. Rev. B*, 82:033411, 2010.

[126] Y.-R. Wu, Y.-Y. Lin, H.-H. Huang, and J. Singh. Electronic and optical properties of InGaN quantum dot based light emitters for solid state lighting. *J. Appl. Phys.*, 105:013117, 2009.

[127] M. Zhang, P. Bhattacharya, and W. Guo. InGaN/GaN self-organized quantum dot green light emitting diodes with reduced efficiency droop. *Appl. Phys. Lett.*, 97:011103, 2010.

[128] Y.-K. Kuo, S.-H. Yen, and Y.-W. Wang. Simulation of deep ultraviolet light-emitting diodes. *Proc. SPIE*, 6669:66691J, 2007.

[129] M. S. Shur and R. Gaska. III-nitride based deep ultraviolet light sources. *Proc. SPIE*, 6894:689419, 2008.

[130] H. Zhao, G. Liu, X.-H. Li, G. S. Huang, J. D. Poplawsky, S. T. Penn, V. Dierolf, and N. Tansu. Growth of staggered InGaN quantum wells light-emitting diodes emitting at 520-525 nm employing graded growth-temperature profile. *Appl. Phys. Lett.*, 95:061104, 2009.

[131] T.-T. Kang, M. Yamamoto, M. Tanaka, A. Hashimoto, and A. Yamamoto. Effect of gas flow on the growth of In-rich AlInN films by metal-organic chemical vapor deposition. *J. Appl. Phys.*, 106:053525, 2009.

[132] L. Kirste, T. Lim, R. Aidam, S. Müller, P. Waltereit, and O. Ambacher. Structural properties of MBE AlInN and AlGaInN barrier layers for GaN-HEMT structures. *phys. stat. sol. (a)*, 207:1338, 2010.

[133] K. Lorenz, S. Magalhães, N. Franco, N. P. Barradas, V. Darakchieva, E. Alves, S. Pereira, M. R. Correia, F. Munnik, R. W. Martin, K. P. O'Donnell, and I. M. Watson. $Al_{1-x}In_x$/GaN bilayers: Structure, morphology, and optical properties. *phys. stat. sol. (b)*, 247:1740, 2010.

[134] C. Adelmann, J. Simon, G. Feuillet, N. T. Pelekanos, B. Daudin, and G. Fishman. Self-assembled InGaN quantum dots grown by molecular-beam epitaxy. *Appl. Phys. Lett.*, 76:1570, 2000.

[135] O. Moriwaki, T. Someya, K. Tachibana, S. Ishida, and Y. Arakawa. Narrow photoluminescence peaks from localized states in InGaN quantum dot structures. *Appl. Phys. Lett.*, 76:2361, 2000.

[136] T. Aschenbrenner, H. Dartsch, C. Kruse, M. Anastasescu, M. Stoica, M. Gartner, A. Pretorius, A. Rosenauer, T. Wagner, and D. Hommel. Optical and structural characterization of AlInN layers for optoelectronic applications. *J. Appl. Phys.*, 108:063533, 2010.

[137] S. Marcinkevičius, V. Liuolia, D. Billingsley, M. Shatalov, J. Yang, R. Gaska, and M. S. Shur. Carrier dynamics and localization in AlInN/GaN heterostructures. *Phys. Stat. Sol. C*, 10:853, 2013.

[138] O. H. Nielsen and R. M. Martin. Stresses in semiconductors: *Ab initio* calculations on Si, Ge, and GaAs. *Phys. Rev. B*, 32:3792, 1985.

[139] M. González-Díaz, P. Rodríguez-Hernández, and A. Muñoz. Elastic constants and electronic structure of beryllium chalcogenides BeS, BeSe, and BeTe from first-principles calculations. *Phys. Rev. B*, 55:14043, 1997.

[140] S. P. Łepkowski, J. A. Majewski, and G. Jurczak. Nonlinear elasticity in III-N compounds: *Ab initio* calculations. *Phys. Rev. B*, 72:245201, 2005.

[141] D. Holec, M. Friák, J. Neugebauer, and P. H. Mayrhofer. Trends in the elastic response of binary early transition metal nitrides. *Phys. Rev. B*, 85:064101, 2012.

[142] M. Łopuszyński and J. A. Majewski. *Ab initio* calculations of third-order elastic constants and related properties for selected semiconductors. *Phys. Rev. B*, 76:045202, 2007.

[143] A. F. Wright. Elastic properties of zinc-blende and wurtzite AlN, GaN, and InN. *J. Appl. Phys.*, 82:2833, 1997.

[144] C. Kittel. *Introduction to Solid State Physics*. John Wiley & Sons, New York, 7 edition, 1996.

[145] L. Kleinman. Deformation potentials in Silicon. I. Uniaxial strain. *Phys. Rev.*, 128:2614, 1962.

[146] P. N. Keating. Effect of invariance requirements on the elastic strain energy of crystals with application to the diamond structure. *Phys. Rev.*, 145:637, 1966.

[147] D. Camacho and Y. M. Niquet. Application of Keating's valence force field model to non-ideal wurtzite materials. *Physica E*, 42:1361, 2010.

[148] M. A. Nusimovici and J. L. Birman. Lattice dynamics of wurtzite: CdS. *Phys. Rev.*, 156:925, 1967.

[149] J. Ihm, A. Zunger, and M. L. Cohen. Momentum-space formalism for the total energy of solids. *J. Phys. C: Solid State Phys.*, 12:4409, 1979.

[150] O. H. Nielsen and R. M. Martin. Quantum-mechanical theory of stress and force. *Phys. Rev. B*, 32:3780, 1985.

[151] T. Schupp, K. Lischka, and D. J. As. MBE growth of atomically smooth non-polar cubic AlN. *J. Cryst. Growth*, 312:1500, 2010.

[152] R. Yu, J. Zhu, and HQ Ye. Calculations of single-crystal elastic constants made simple. *Comp. Phys. Commun.*, 181:671, 2010.

[153] A. Polian, M. Grimsditch, and I. Grzegory. Elastic constants of gallium nitride. *J. Appl. Phys.*, 79:3343, 1996.

[154] Q. Yan, P. Rinke, M. Winkelnkemper, A. Qteish, D. Bimberg, M. Scheffler, and C. G. Van de Walle. Band parameters and strain effects in ZnO and group-III nitrides. *Semicond. Sci. Technol.*, 26:014037, 2011.

[155] L. E. McNeil, M. Grimsditch, and R. H. French. Vibrational spectroscopy of aluminum nitride. *J. Am. Ceram. Soc.*, 76:1132, 1993.

[156] J. Serrano, A. Bosak, M. Krisch, F. J. Manjón, A. H. Romero, N. Garro, X. Wang, A. Yoshikawa, and M. Kuball. InN thin film lattice dynamics by grazing incidence inelastic x-ray scattering. *Phys. Rev. Lett.*, 106:205501, 2011.

[157] D. Moss, A. V. Akimov, S. V. Novikov, R. P. Campion, C. R. Staddon, N. Zainal, C. T. Foxon, and A. J. Kent. Elasto-optical properties of zinc-blende (cubic) GaN measured by picosecond acoustics. *J. Phys. D: Appl. Phys.*, 42:115412, 2009.

[158] R. B. Schwarz, K. Khachaturyan, and E. R. Weber. Elastic moduli of gallium nitride. *Appl. Phys. Lett.*, 70:1122, 1997.

[159] M. A. Hopcroft, W. D. Nix, and T. W. Kenny. What is the Young's modulus of silicon? *J Microelectromech. S.*, 19:229, 2010.

[160] G. Bester, X. Wu, D. Vanderbilt, and A. Zunger. Importance of second-order piezoelectric effects in zinc-blende semiconductors. *Phys. Rev. Lett.*, 96:187602, 2006.

[161] M. Grundmann, O. Stier, and D. Bimberg. InAs/GaAs pyramidal quantum dots: Strain distribution, optical phonons, and electronic structure. *Phys. Rev. B*, 52:11969, 1995.

[162] G. Bester and A. Zunger. Cylindrically shaped zinc-blende semiconductor quantum dots do not have cylindrical symmetry: Atomistic symmetry, atomic relaxation, and piezoelectric effects. *Phys. Rev. B*, 71:045318, 2005.

[163] J. Pal, G. Tse, V. Haxha, M.A. Migliorato, and S. Tomić. Second-order piezoelectricity in wurtzite III-N semiconductors. *Phys. Rev. B*, 84:085211, 2011.

[164] S.-H. Wei, L. G. Ferreira, J. E. Bernard, and A. Zunger. Electronic properties of random alloys: Special quasirandom structures. *Phys. Rev. B*, 42:9622, 1990.

[165] R. G. Barrera, O. Guzman, and B. Balaguer. Point charge in a three-dielectric medium with planar interfaces. *Am. J. Phys*, 46:11, 1978.

[166] C. Pryor, J. Kim, L. W. Wang, A. J. Williamson, and A. Zunger. Comparison of two methods for describing the strain profiles in quantum dots. *J. Appl. Phys.*, 83:2546, 1998.

[167] R. M. Martin. Relation between elastic tensors of wurtzite and zinc-blende structure materials. *Phys. Rev. B*, 6:4546, 1972.

[168] J. R. Petta, A. C. Johnson, J. M. Taylor, E. A. Laird, A. Yacoby, M. D. Lukin, C. M. Marcus, M. P. Hanson, and A. C. Gossard. Coherent manipulation of coupled electron spins in semiconductor quantum dots. *Science*, 309:2180, 2005.

[169] A. C. Hindmarsh, P. N. Brown, K. E. Grant, S. L. Lee, R. Serban, D. E. Shumaker, and C. S. Woodward. SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers. *ACM Trans. Math. Softw.*, 31:363, 2005.

Appendices

A Additional material on local polarization and the point dipole method

A.1 Example of the calculation of a local piezoelectric coefficient

To illustrate how the calculation of the local piezoelectric tensor of Eq. (5.13) is done, we give here the details of the calculation for $e_{15}^{*,\mathrm{A}}$. Equation (5.12) for $e_{15}^{*,\mathrm{A}}$ is simplified to

$$e_{15}^{*,\mathrm{A}} = e_{15}^{(0)} - \frac{e\mathcal{Z}_1^{\mathrm{A}}}{\sqrt{3}{a_0}^2 c_0}\left(\frac{\partial \mu_1^{\mathrm{A}}}{\partial \epsilon_5} - \frac{1}{2}\mu_{3,0}^{\mathrm{A}}\right), \tag{A.1}$$

where we have made use of the Voigt relation $\partial\epsilon_{13}/\partial\epsilon_5 = 1/2$. $\mu_{3,0}^{\mathrm{A}} = 4(u_0 - 3/8)c_0$ is given by the WZ internal parameter, and $\mathcal{Z}_1^{\mathrm{A}} \equiv \mathcal{Z}_1$ for A being a cation. We need to calculate μ_1^{A}. Looking at Fig. 5.4, it is clear that the nearest neighbours of A are B, which we label 1, and three periodic replicas of D contained in a plane below A, which we label 2–4. If A is fixed at the origin, $\mathbf{r}_{\mathrm{A}} = (0,0,0)$, then the distances of the different nearest neighbours from A are given by:

$$\begin{aligned}
\boldsymbol{\ell}^1 &= \mathbf{r}_{\mathrm{B}}, \\
\boldsymbol{\ell}^2 &= \mathbf{r}_{\mathrm{D}} - \mathbf{c}, \\
\boldsymbol{\ell}^3 &= \mathbf{r}_{\mathrm{D}} - \mathbf{c} - \mathbf{a}, \\
\boldsymbol{\ell}^4 &= \mathbf{r}_{\mathrm{D}} - \mathbf{c} - \mathbf{b},
\end{aligned} \tag{A.2}$$

where $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ are the (strained) lattice vectors of the unit cell. Since for this example we are interested in $e_{15}^{*,\mathrm{A}}$ only, we set all the strain components to zero except for $\epsilon_5 = 2\epsilon_{13}$. Following all the definitions given in Ref. [3] and Chapter 4

(with exchanged notation $\epsilon_{13} \leftrightarrow \epsilon_{xz}$), we can write:

$$\begin{aligned}
\mathbf{r}_\mathrm{B} &= [u_0 c_0 \epsilon_{13}, 0, u_0 c_0] + [\zeta_1 c_0 \epsilon_{13}, 0, 0]\,, \\
\mathbf{r}_\mathrm{D} &= \left[\frac{a_0}{2} + \left(\frac{1}{2} + u_0\right) c_0 \epsilon_{13}, \frac{\sqrt{3} a_0}{6}, \left(\frac{1}{2} + u_0\right) c_0 + \frac{a_0}{2}\epsilon_{13}\right] + [\zeta_1 c_0 \epsilon_{13}, 0, 0]\,, \\
\mathbf{a} &= [a_0, 0, a_0 \epsilon_{13}]\,, \\
\mathbf{b} &= \left[\frac{a_0}{2}, \frac{\sqrt{3} a_0}{2}, \frac{a_0}{2}\epsilon_{13}\right], \\
\mathbf{c} &= [c_0 \epsilon_{13}, 0, c_0]\,.
\end{aligned} \tag{A.3}$$

To obtain μ_1^A we sum over nearest-neighbour distances:

$$\mu_1^\mathrm{A} = \sum_{\alpha=1}^{4} \ell_1^\alpha = 4 u_0 c_0 \epsilon_{13} - \frac{3}{2} c_0 \epsilon_{13} + 4 \zeta_1 c_0 \epsilon_{13}. \tag{A.4}$$

The last term of Eq. (A.1) therefore reduces to

$$\frac{\partial \mu_1^\mathrm{A}}{\partial \epsilon_5} - \frac{1}{2}\mu_{3,0}^\mathrm{A} = 2\left(u_0 - \frac{3}{8}\right) c_0 + 2\zeta_1 c_0 - 2\left(u_0 - \frac{3}{8}\right) c_0 = 2\zeta_1 c_0, \tag{A.5}$$

which leads to the final result:

$$e_{15}^{*,\mathrm{A}} = e_{15}^{(0)} - \frac{2 e \mathcal{Z}_1}{\sqrt{3} {a_0}^2} \zeta_1. \tag{A.6}$$

A.2 Point dipole solution for the three-dielectric problem

Building on the description made by Barrera for point charges in a three-dielectric configuration [165], we give here the analogous solution for point dipoles. The reflections necessary to construct the image point dipoles are illustrated in Fig. A.1. Following the convention of Fig. A.1, where d is the distance from the dipole to the left side interface and h is the distance between the two interfaces, we can obtain some rules for the form of the image charges $\mathbf{p}'_{(n)}$ and $\mathbf{p}''_{(n)}$, being the nth reflections of $\mathbf{p}$ starting at left and right, respectively:

1. Each reflection to the left adds $2d$ to the distance between the original dipole

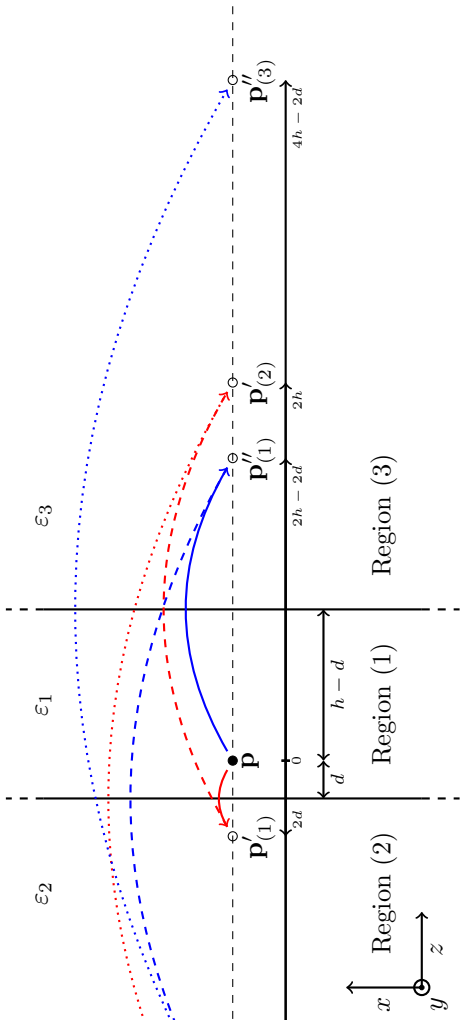


Figure A.1: Reflection of image dipoles up to third order in a three-dielectric set up. Each of the reflection sequences is denoted by a different color: the sequence starting to the left and originating the series $\mathbf{p}'_{(n)}$ is coloured in red whereas the sequence starting to the right and originating the series $\mathbf{p}''_{(n)}$ is in blue. Solid lines indicate first order reflections, dashed lines indicate second order reflections and dotted lines indicate third order reflections.

p and the reflected dipole.

2. Each reflection to the right adds $2h-2d$ to the distance between the original dipole **p** and the reflected dipole.

3. Each reflection to the left adds a factor of $\frac{\varepsilon_1-\varepsilon_2}{\varepsilon_1+\varepsilon_2}$ to each of the components of the image dipole and inverts the sign of the z component.

4. Each reflection to the right adds a factor of $\frac{\varepsilon_1-\varepsilon_3}{\varepsilon_1+\varepsilon_3}$ to each of the components of the image dipole and inverts the sign of the z component.

These rules can be written as the following expressions. For the position of the image dipoles:

$$\begin{aligned}
z_{\mathbf{p}'_{(2n-1)}} &= z_{\mathbf{p}} - \left[n\,2d + (n-1)\,(2h-2d)\right], \\
z_{\mathbf{p}'_{(2n)}} &= z_{\mathbf{p}} + \left[n\,2d + (n-1)\,(2h-2d)\right], \\
z_{\mathbf{p}''_{(2n-1)}} &= z_{\mathbf{p}} + \left[n\,(2h-2d) + (n-1)\,2d\right], \\
z_{\mathbf{p}''_{(2n)}} &= z_{\mathbf{p}} - \left[n\,(2h-2d) + (n-1)\,2d\right],
\end{aligned} \tag{A.7}$$

and for the value of the image dipoles:

$$\begin{aligned}
\mathbf{p}'_{(2n-1)} &= [p_x, p_y, -p_z] \left(\frac{\varepsilon_1-\varepsilon_2}{\varepsilon_1+\varepsilon_2}\right)^n \left(\frac{\varepsilon_1-\varepsilon_3}{\varepsilon_1+\varepsilon_3}\right)^{n-1}, \\
\mathbf{p}'_{(2n)} &= [p_x, p_y, p_z] \left(\frac{\varepsilon_1-\varepsilon_2}{\varepsilon_1+\varepsilon_2}\right)^n \left(\frac{\varepsilon_1-\varepsilon_3}{\varepsilon_1+\varepsilon_3}\right)^{n}, \\
\mathbf{p}''_{(2n-1)} &= [p_x, p_y, -p_z] \left(\frac{\varepsilon_1-\varepsilon_2}{\varepsilon_1+\varepsilon_2}\right)^{n-1} \left(\frac{\varepsilon_1-\varepsilon_3}{\varepsilon_1+\varepsilon_3}\right)^{n}, \\
\mathbf{p}''_{(2n)} &= \mathbf{p}'_{(2n)},
\end{aligned} \tag{A.8}$$

with $n \in \mathbb{N}$. Finally the expression of the potential in all three regions can be

written as

$$\phi_{\mathbf{p}}^{(1)}(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0\varepsilon_1}\left\{\frac{\mathbf{p}\cdot(\mathbf{r}-\mathbf{r}_{\mathbf{p}})}{|\mathbf{r}-\mathbf{r}_{\mathbf{p}}|^3} + \sum_{n=1}^{\infty}\left(\frac{\mathbf{p}'_{(n)}\cdot\left(\mathbf{r}-\mathbf{r}_{\mathbf{p}'_{(n)}}\right)}{|\mathbf{r}-\mathbf{r}_{\mathbf{p}'_{(n)}}|^3} + \frac{\mathbf{p}''_{(n)}\cdot\left(\mathbf{r}-\mathbf{r}_{\mathbf{p}''_{(n)}}\right)}{|\mathbf{r}-\mathbf{r}_{\mathbf{p}''_{(n)}}|^3}\right)\right\},$$

$$\phi_{\mathbf{p}}^{(2)}(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0\varepsilon_2}\left\{\frac{\mathbf{p}\cdot(\mathbf{r}-\mathbf{r}_{\mathbf{p}})}{|\mathbf{r}-\mathbf{r}_{\mathbf{p}}|^3} + \sum_{n=1}^{\infty}\left(\frac{\mathbf{p}'_{(2n)}\cdot\left(\mathbf{r}-\mathbf{r}_{\mathbf{p}'_{(2n)}}\right)}{|\mathbf{r}-\mathbf{r}_{\mathbf{p}'_{(2n)}}|^3} + \frac{\mathbf{p}''_{(2n-1)}\cdot\left(\mathbf{r}-\mathbf{r}_{\mathbf{p}''_{(2n-1)}}\right)}{|\mathbf{r}-\mathbf{r}_{\mathbf{p}''_{(2n-1)}}|^3}\right)\right\}\frac{2\varepsilon_2}{\varepsilon_1+\varepsilon_2},$$

$$\phi_{\mathbf{p}}^{(3)}(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0\varepsilon_3}\left\{\frac{\mathbf{p}\cdot(\mathbf{r}-\mathbf{r}_{\mathbf{p}})}{|\mathbf{r}-\mathbf{r}_{\mathbf{p}}|^3} + \sum_{n=1}^{\infty}\left(\frac{\mathbf{p}'_{(2n-1)}\cdot\left(\mathbf{r}-\mathbf{r}_{\mathbf{p}'_{(2n-1)}}\right)}{|\mathbf{r}-\mathbf{r}_{\mathbf{p}'_{(2n-1)}}|^3} + \frac{\mathbf{p}''_{(2n)}\cdot\left(\mathbf{r}-\mathbf{r}_{\mathbf{p}''_{(2n)}}\right)}{|\mathbf{r}-\mathbf{r}_{\mathbf{p}''_{(2n)}}|^3}\right)\right\}\frac{2\varepsilon_3}{\varepsilon_1+\varepsilon_3}. \tag{A.9}$$

It is implicit in Eq. (A.9) that for the calculation of the potential $\phi_{\mathbf{p}}^{(2)}$ in region (2) only the image dipoles in region (3) (together with the original dipole) are taken into account, and vice-versa. Given the form of Eq. (A.9) it is clear that an exact solution to the problem of three media cannot be obtained for a finite number of terms in the summation. However, approximate solutions can be obtained whose accuracy will depend mostly on the difference in the values of the dielectric constants of the different materials. In Fig. A.2 we show approximations up to third order reflections, for different orientations of the dipole, in the case of three materials for which $\varepsilon_2 = 2\varepsilon_1$ and $\varepsilon_3 = 3\varepsilon_1$. This is an extreme case in the context of III-V compounds, for which the differences in ε_r between materials do not usually go beyond 50%. For clarity of interpretation, the potential isolines shown decay as a power of 2, which allows to visualize the fine effects of the interfaces far from the dipole. As can be seen, the second order correction [Figs. A.2(d–f)] is already very well converged for this extreme case and we expect first order corrections to be sufficient for the materials of interest, group-III nitrides in particular.

A.3 Gaussian smearing of point dipoles

As mentioned in Chapter 5, the potential solution for a point dipole is an approximation to the potential due to the dipole moment of a charge distribution [45]. This approximation is only valid in the limit when the potential is calculated sufficiently far away from the charge distribution. How far is "sufficiently far" depends on the particular problem at hand, basically on the value of the dipole and the volume over which the charge density giving rise to the dipole moment spread originally. A Gaussian smearing of the conflicting dipoles, i.e. those that are close

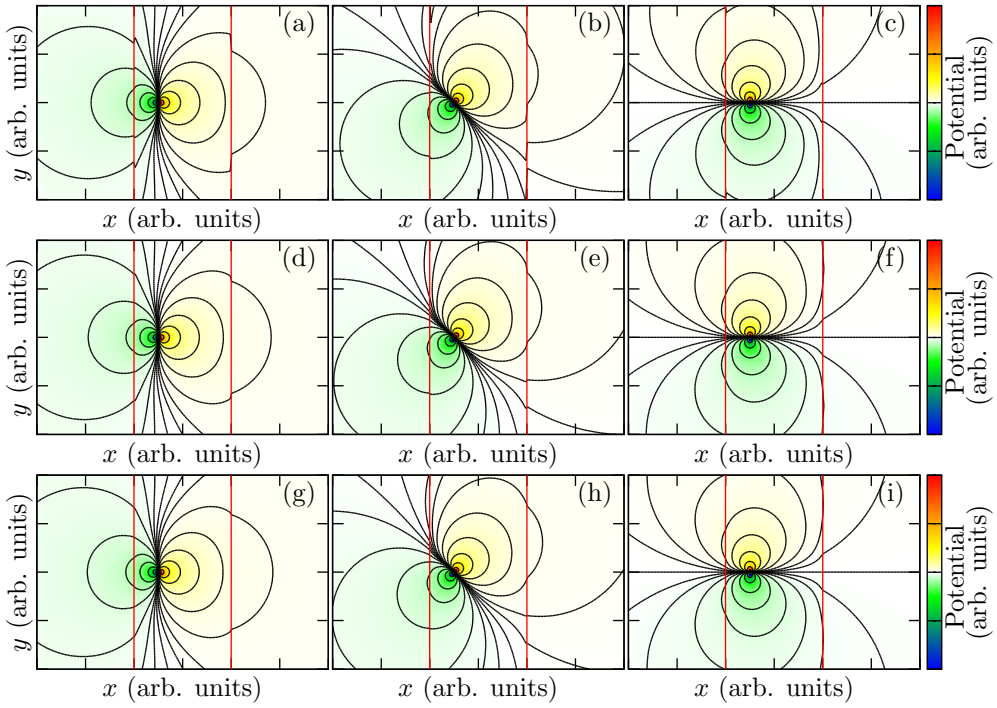


Figure A.2: Different approximations for the three media problem shown schematically in Fig. A.1 for different orientations of the dipole. (a–c) include up to first order reflections, (d–f) second order reflections and (g–i) include up to third order reflections. The red lines indicate the interfaces between different materials: the central material has an (arbitrary) permittivity of $\varepsilon = 1$, the material on the left has $\varepsilon = 2$ and the material on the right has $\varepsilon = 3$. It can be seen that third order reflections are sufficient to converge the potential for that particular set of relative values of ε.

to the position where the potential is calculated, is a straightforward manner to deal with this problem, as the parameters controlling the smearing can be tuned easily at need. We propose the implementation of this smearing controlled by two parameters:

1. r_{smear} is the cut-off radius for which all the dipoles that obey $|\mathbf{r} - \mathbf{r_p}| < r_{\text{smear}}$ are smeared, where $\mathbf{r}$ is the position where the potential is calculated and $\mathbf{r_p}$ is the position of the dipole under consideration.

2. σ is the standard deviation of the Gaussian function that produces the smearing. It gives a measure of the volume over which the dipole is smeared.

Therefore, the expression for the potential $\phi_{\mathbf{p}}(\mathbf{r})$ at $\mathbf{r}$ due to a dipole $\mathbf{p}$ located at $\mathbf{r_p}$ can be rewritten, in spherical coordinates, as

$$\phi_{\mathbf{p}}(\mathbf{r}) = \begin{cases} \dfrac{1}{4\pi\varepsilon_0\varepsilon_r}\dfrac{\mathbf{p}\cdot(\mathbf{r}-\mathbf{r_p})}{|\mathbf{r}-\mathbf{r_p}|^3} & \text{for } |\mathbf{r}-\mathbf{r_p}| \geq r_{\text{smear}} \\ \dfrac{1}{4\pi\varepsilon_0\varepsilon_r}\dfrac{1}{(2\pi\sigma^2)^{\frac{3}{2}}}\int_0^{2\pi}\mathrm{d}\varphi\int_0^{\pi}\mathrm{d}\theta\int_0^{\infty}\mathrm{d}\rho\sin\theta\rho^2 e^{-\frac{|\mathbf{r}-\mathbf{r'_p}|^2}{2\sigma^2}}\dfrac{\mathbf{p}\cdot(\mathbf{r}-\mathbf{r'_p})}{|\mathbf{r}-\mathbf{r'_p}|^3} & \text{for } |\mathbf{r}-\mathbf{r_p}| < r_{\text{smear}} \end{cases} \tag{A.10}$$

where $\mathbf{r'_p}$ is given by

$$\mathbf{r'_p} = \mathbf{r_p} + [\rho\sin\theta\cos\varphi,\ \rho\sin\theta\sin\varphi,\ \rho\cos\theta]\,. \tag{A.11}$$

Typically, the integration in ρ can be done up to a certain cutoff since the value of the integrant will decay rapidly. For example, our current implementation sets 3.4σ as the upper limit for the integration, which comprises a volume that contains about 99% of the total original dipole moment $\mathbf{p}$. The extension of Eq. (A.10) to the case in which different dielectric constants are present is straightforward and done in the same way as explained in Chapter 5 and Section A.2 of this appendix.

A.4 Computational tricks

Method of layers

In the same way that a point dipole is an approximation for a charge density distribution valid far away from the location of the dipole, it can be shown that a point dipole can be a valid approximation for a given *ensemble* of neighbouring point dipoles at a certain distance from the ensemble. Figure A.3(a) shows the potential due to an ensemble of N dipoles $\mathbf{p}_i$, of typical magnitude in nitride QWs, that are localized in a restricted region in space, each at position $\mathbf{r}_{\mathbf{p}_i}$. This ensemble can be approximated by a single dipole $\mathbf{P}$ whose magnitude equals the summation of all the original dipoles and whose position $\mathbf{r_P}$ is given by the weighted average of the dipoles in the ensemble [Fig. A.3(b)]:

$$\mathbf{P} = \sum_{i=1}^{N}\mathbf{p}_i, \qquad \mathbf{r_P} = \frac{1}{|\mathbf{P}|}\sum_{i=1}^{N}\mathbf{r}_{\mathbf{p}_i}\,|\mathbf{p}_i|. \tag{A.12}$$

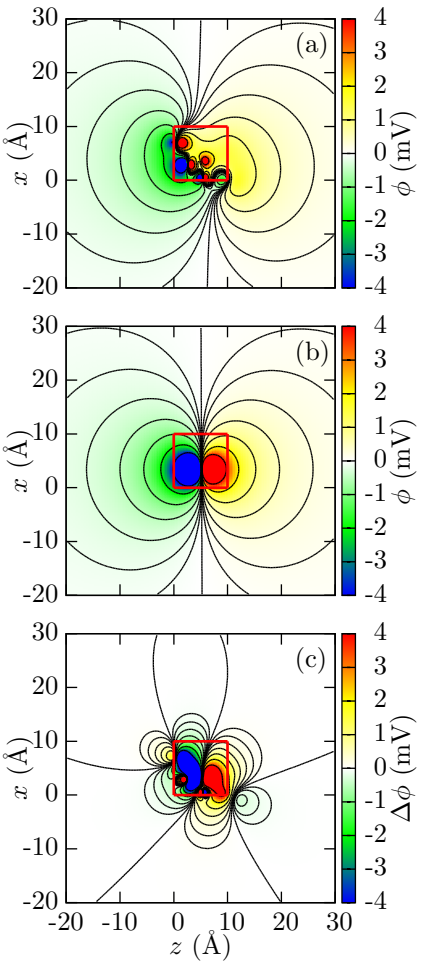


Figure A.3: Section in the xz plane of: (a) The polarization potential due to an ensemble of 10 dipoles randomly placed inside a cube of side 10 Å with origin at $(0, 0, 0)$. The boundaries of the cube within the xz plane is indicated by the red line. The dipoles components are given by a Gaussian probability distribution with p_x centered at 10^{-31} Cm, p_y and p_z centered at zero, and standard deviation 5×10^{-32} Cm for all three components; (b) The polarization potential due to a single dipole obtained from the ensemble in (a) calculated by means of Eq. (A.12); And (c) difference between the potentials shown in (b) and (a). Note that the potential isolines shown follow an exponential behaviour to exaggerate the results: the lines escaping the plots correspond to zero and the outer lobe-shaped isolines indicate $\sim 6\mu$V.

As shown in Fig. A.3(c), the difference between an ensemble of dipoles and its correspondent approximation calculated as in Eq. (A.12) decays rapidly away from the ensemble. Applying Eq. (A.12) recurrently, one can construct, around the point $\mathbf{r}$ where the potential ϕ is being calculated, a system of "layers" in which the density of dipoles decreases as one moves away from $\mathbf{r}$.

Simplification for quantum wells

A simplification can be made for QW systems, or even a quantum dot (QD) system, if a constant value for the polarization can be assumed for the greatest part of the system. Since only *differences* in polarization are meaningful for the calculation of polarization potentials, an arbitrary constant shift of the polarization of the *whole* system will not have any effect on the calculated value of the polarization potential. This shift can be chosen in such a way that the resultant polarization, at least on average, is zero in the barrier in the case of a QW, or in the *unstrained* barrier in the case of a QD.[1] In that case, all the dipoles arising from that region, once the discretization described in Chapter 5 is made, will have value zero. Therefore, the dipoles contained within that region can be left out of the calculation.

[1]The strain field will not vanish in the barrier material immediately around a QD, but will effectively be zero far away from it. See for instance Pryor *et al.* [166].

B Expression for the derivative of the total energy in the staggered grid

The following is the explicit form of Eq. (6.23) for the case $\partial E_{\text{total}}/\partial u_x(l+\frac{1}{2},m,n)$:

$$
\begin{aligned}
&\frac{\partial E_{\text{total}}\left(\mathbf{u}\right)}{\partial u_x\left(l+\frac{1}{2},m,n\right)} = \\
&\quad\frac{1}{h_x}\Big(a_{\text{ref}}\left(a_{\text{ref}}-a\left[l,m,n\right]\right)c\left[l,m,n\right]C_{11}\left[l,m,n\right]-a_{\text{ref}}\left(a_{\text{ref}}-a\left[l+1,m,n\right]\right)c\left[l+1,m,n\right]C_{11}\left[l+1,m,n\right] \\
&\qquad+a_{\text{ref}}\left(a_{\text{ref}}-a\left[l,m,n\right]\right)c\left[l,m,n\right]C_{12}\left[l,m,n\right]-a_{\text{ref}}\left(a_{\text{ref}}-a\left[l+1,m,n\right]\right)c\left[l+1,m,n\right]C_{12}\left[l+1,m,n\right] \\
&\qquad+a_{\text{ref}}a\left[l,m,n\right]\left(c_{\text{ref}}-c\left[l,m,n\right]\right)C_{13}\left[l,m,n\right]-a_{\text{ref}}a\left[l+1,m,n\right]\left(c_{\text{ref}}-c\left[l+1,m,n\right]\right)C_{13}\left[l+1,m,n\right]\Big) \\
&\quad-\frac{a_{\text{ref}}^2c\left[l,m,n\right]C_{11}\left[l,m,n\right]u_x\left[l-\frac{1}{2},m,n\right]}{h_x^2}-\frac{a_{\text{ref}}^2c\left[l+\frac{1}{2},m-\frac{1}{2},n\right]C_{66}\left[l+\frac{1}{2},m-\frac{1}{2},n\right]u_x\left[l+\frac{1}{2},m-1,n\right]}{h_y^2} \\
&\quad-\frac{c_{\text{ref}}^2a\left[l+\frac{1}{2},m,n-\frac{1}{2}\right]^2C_{44}\left[l+\frac{1}{2},m,n-\frac{1}{2}\right]u_x\left[l+\frac{1}{2},m,n-1\right]}{h_z^2c\left[l+\frac{1}{2},m,n-\frac{1}{2}\right]} \\
&\quad+\Bigg(\frac{a_{\text{ref}}^2c\left[l,m,n\right]C_{11}\left[l,m,n\right]}{h_x^2}+\frac{a_{\text{ref}}^2c\left[l+1,m,n\right]C_{11}\left[l+1,m,n\right]}{h_x^2}+\frac{c_{\text{ref}}^2a\left[l+\frac{1}{2},m,n-\frac{1}{2}\right]^2C_{44}\left[l+\frac{1}{2},m,n-\frac{1}{2}\right]}{h_z^2c\left[l+\frac{1}{2},m,n-\frac{1}{2}\right]} \\
&\qquad+\frac{c_{\text{ref}}^2a\left[l+\frac{1}{2},m,n+\frac{1}{2}\right]^2C_{44}\left[l+\frac{1}{2},m,n+\frac{1}{2}\right]}{h_z^2c\left[l+\frac{1}{2},m,n+\frac{1}{2}\right]}+\frac{a_{\text{ref}}^2c\left[l+\frac{1}{2},m-\frac{1}{2},n\right]C_{66}\left[l+\frac{1}{2},m-\frac{1}{2},n\right]}{h_y^2} \\
&\qquad+\frac{a_{\text{ref}}^2c\left[l+\frac{1}{2},m+\frac{1}{2},n\right]C_{66}\left[l+\frac{1}{2},m+\frac{1}{2},n\right]}{h_y^2}\Bigg)u_x\left[l+\frac{1}{2},m,n\right] \\
&\quad-\frac{c_{\text{ref}}^2a\left[l+\frac{1}{2},m,n+\frac{1}{2}\right]^2C_{44}\left[l+\frac{1}{2},m,n+\frac{1}{2}\right]u_x\left[l+\frac{1}{2},m,n+1\right]}{h_z^2c\left[l+\frac{1}{2},m,n+\frac{1}{2}\right]} \\
&\quad-\frac{a_{\text{ref}}^2c\left[l+\frac{1}{2},m+\frac{1}{2},n\right]C_{66}\left[l+\frac{1}{2},m+\frac{1}{2},n\right]u_x\left[l+\frac{1}{2},m+1,n\right]}{h_y^2}-\frac{a_{\text{ref}}^2c\left[l+1,m,n\right]C_{11}\left[l+1,m,n\right]u_x\left[l+\frac{3}{2},m,n\right]}{h_x^2} \\
&\quad+\left(-\frac{a_{\text{ref}}^2c\left[l,m,n\right]C_{12}\left[l,m,n\right]}{h_xh_y}-\frac{a_{\text{ref}}^2c\left[l+\frac{1}{2},m-\frac{1}{2},n\right]C_{66}\left[l+\frac{1}{2},m-\frac{1}{2},n\right]}{h_xh_y}\right)u_y\left[l,m-\frac{1}{2},n\right] \\
&\quad+\left(\frac{a_{\text{ref}}^2c\left[l,m,n\right]C_{12}\left[l,m,n\right]}{h_xh_y}+\frac{a_{\text{ref}}^2c\left[l+\frac{1}{2},m+\frac{1}{2},n\right]C_{66}\left[l+\frac{1}{2},m+\frac{1}{2},n\right]}{h_xh_y}\right)u_y\left[l,m+\frac{1}{2},n\right] \\
&\quad+\left(\frac{a_{\text{ref}}^2c\left[l+1,m,n\right]C_{12}\left[l+1,m,n\right]}{h_xh_y}+\frac{a_{\text{ref}}^2c\left[l+\frac{1}{2},m-\frac{1}{2},n\right]C_{66}\left[l+\frac{1}{2},m-\frac{1}{2},n\right]}{h_xh_y}\right)u_y\left[l+1,m-\frac{1}{2},n\right] \\
&\quad+\left(-\frac{a_{\text{ref}}^2c\left[l+1,m,n\right]C_{12}\left[l+1,m,n\right]}{h_xh_y}-\frac{a_{\text{ref}}^2c\left[l+\frac{1}{2},m+\frac{1}{2},n\right]C_{66}\left[l+\frac{1}{2},m+\frac{1}{2},n\right]}{h_xh_y}\right)u_y\left[l+1,m+\frac{1}{2},n\right] \\
&\quad+\left(-\frac{a_{\text{ref}}c_{\text{ref}}a\left[l,m,n\right]C_{13}\left[l,m,n\right]}{h_xh_z}-\frac{a_{\text{ref}}c_{\text{ref}}a\left[l+\frac{1}{2},m,n-\frac{1}{2}\right]C_{44}\left[l+\frac{1}{2},m,n-\frac{1}{2}\right]}{h_xh_z}\right)u_z\left[l,m,n-\frac{1}{2}\right]
\end{aligned}
$$

$$
\begin{aligned}
&+\left(\frac{a_{\mathrm{ref}}c_{\mathrm{ref}}a\left[l,m,n\right]C_{13}\left[l,m,n\right]}{h_x h_z}+\frac{a_{\mathrm{ref}}c_{\mathrm{ref}}a\left[l+\frac{1}{2},m,n+\frac{1}{2}\right]C_{44}\left[l+\frac{1}{2},m,n+\frac{1}{2}\right]}{h_x h_z}\right)u_z\left[l,m,n+\frac{1}{2}\right]\\
&+\left(\frac{a_{\mathrm{ref}}c_{\mathrm{ref}}a\left[l+1,m,n\right]C_{13}\left[l+1,m,n\right]}{h_x h_z}\right.\\
&\quad\left.+\frac{a_{\mathrm{ref}}c_{\mathrm{ref}}a\left[l+\frac{1}{2},m,n-\frac{1}{2}\right]C_{44}\left[l+\frac{1}{2},m,n-\frac{1}{2}\right]}{h_x h_z}\right)u_z\left[l+1,m,n-\frac{1}{2}\right]\\
&+\left(-\frac{a_{\mathrm{ref}}c_{\mathrm{ref}}a\left[l+1,m,n\right]C_{13}\left[l+1,m,n\right]}{h_x h_z}\right.\\
&\quad\left.-\frac{a_{\mathrm{ref}}c_{\mathrm{ref}}a\left[l+\frac{1}{2},m,n+\frac{1}{2}\right]C_{44}\left[l+\frac{1}{2},m,n+\frac{1}{2}\right]}{h_x h_z}\right)u_z\left[l+1,m,n+\frac{1}{2}\right].
\end{aligned}
\tag{B.1}
$$