



UNIVERSITY of L'AQUILA

Faculty of Engineering
Department of Pure and Applied Mathematics

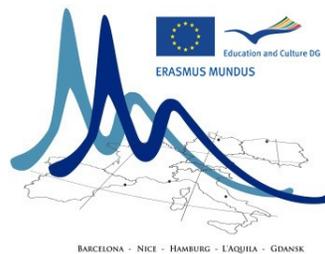
Master Thesis
in
Mathematical Engineering

**Simulation of GaN/AlGaN heterostructures
for a HEMT simulator**

Advisor
Prof. G. Leuzzi

Candidate
P. Ellinghaus

Co-advisor
Dr Vincenzo Stornelli



MathMods Master Programme

With the support of the Erasmus Mundus Master Programme of the European Union

Academic Year 2010/2011

Contents

1	Introduction	1
2	Gallium nitride material system	2
2.1	Group III nitrides	2
2.2	Research history of gallium nitride	2
2.3	Crystal structure	3
2.4	Substrates	4
2.5	Polarization	5
2.5.1	Piezoelectric polarization	6
2.5.2	Spontaneous polarization	6
2.5.3	Pyroelectric effect	7
2.6	Applications	7
2.6.1	Optoelectronics	7
2.6.2	Radio Frequency electronics	8
2.6.3	Power electronics	8
3	High electron mobility transistors	9
3.1	Origin of HEMT structure	9
3.2	Operation principle of a HEMT	9
3.3	Structure	10
3.3.1	Breakdown	11
3.3.2	Passivation	11
3.3.3	Schottky barrier gate contact	12
4	Physical and material models	13
4.1	Material models and parameters	13
4.1.1	Bandgap	14
4.1.2	Effective carrier masses	14
4.1.3	Dielectric constant	15
4.1.4	Lattice constant	15
4.1.5	Crystal strain	15

4.1.6	Polarization	16
4.1.6.1	Piezoelectric polarization	16
4.1.6.2	Spontaneous polarization	17
4.1.6.3	Multi-layer heterostructure	18
4.1.6.4	Thermally-induced strain	18
4.1.7	Schottky barrier	18
4.1.8	Surface states	20
4.2	Semiconductor equations	20
4.2.1	Poisson's equation	20
4.2.2	Schrodinger's equation	21
4.2.3	Calculation of charge carrier densities	22
4.2.3.1	Ionization of dopants	22
4.2.3.2	Mobile charge carriers	22
4.2.3.3	Quantum carrier density	24
4.2.4	Continuity equation	24
4.3	Quasi-2D charge transport model	25
5	Numerical simulation	26
5.1	Solving the system of equations	26
5.1.1	Self-consistent solution of Schrodinger-Poisson system	26
5.1.1.1	Under-relaxation	27
5.1.1.2	Predictor-corrector	27
5.1.2	Nonequilibrium case	28
5.2	Numerical schemes	29
5.2.1	Spatial discretization	29
5.2.2	Poisson's equation	30
5.2.2.1	Discretization by finite volumes	30
5.2.2.2	Boundary conditions	31
5.2.2.3	Newton-Raphson method	32
5.2.2.4	Implementation	33
5.2.3	Schrodinger's equation	34
5.2.3.1	Discretization by finite differences	34
5.2.3.2	Boundary conditions	35
5.2.3.3	Implicitly restarted Arnoldi method	35
5.2.3.4	Inverse iteration	35
5.2.3.5	Implementation	36
5.2.4	Continuity equation	36
5.2.4.1	Discretization by finite differences	37
5.2.4.2	Discretization by finite volumes	38

5.2.4.3	Newton-Raphson method	38
5.2.4.4	Boundary conditions	40
5.2.4.5	Implementation	40
5.3	Code implementation	40
5.3.1	Additions and modifications to AQUILA code	41
5.3.1.1	Material parameters	41
5.3.1.2	Polarization effects	41
5.3.1.3	Continuity equation	41
5.3.1.4	Simplifications for 1D problem	41
5.3.1.5	Plotting functions	42
5.3.2	Assumptions and simplifications	42
5.3.2.1	Polarization-induced sheet charge	42
5.3.2.2	Surface states	42
5.3.2.3	Carrier types considered	42
5.3.3	Units and scaling of equations	42
5.3.4	Initial values of variables	43
5.3.5	Consideration of the quantum region	43
5.3.5.1	Extent of the quantum region	43
5.3.5.2	Coupling of quantum and classical regions	43
5.3.5.3	Classical correction to quantum regions	44
5.3.5.4	The appropriate number of subbands	45
5.3.5.5	Boundary condition	45
5.3.6	Fixes for nonequilibrium calculation	45
5.4	Alternative simulators	46
5.4.1	BandEng	46
5.4.1.1	Using BandEng for validation	46
5.4.2	nextnano	46
5.4.3	1D Poisson/Schrodinger	47
6	Results	48
7	Conclusion	61
7.1	Summary	61
7.2	Recommendations	61

Chapter 1

Introduction

The interest in group III nitride semiconductors has flared in the last two decades as the inherent material properties offer attractive advantages for optoelectronic and high-power devices in diverse application fields ranging from consumer electronics to military radars. Gallium nitride (GaN), and its associated alloys, has emerged as the most promising nitride semiconductor for commercial applications.

The group III nitride materials lack the technological maturity of other semiconductors like silicon - the mainstay of the semiconductor industry - or gallium arsenide, as it still presents many technological challenges e.g. high quality substrate production. The intricate material properties of the gallium nitride material system must be well understood to fully exploit the advantages offered by this semiconductor in devices. The physics-based simulation of semiconductor devices assists in this cause by exposing a quantitative relationship between underlying material properties and device behaviour. Device optimization in an experimental environment is very costly and time consuming and should therefore also be complemented by simulations.

High electron mobility transistors (HEMTs) have proliferated themselves as the preferred transistor type for RF and microwave frequency applications, especially when high power outputs are required. This thesis investigates the physical simulation of the charge carrier distribution in GaN/AlGaN heterostructures found in such HEMTs for the incorporation in a complete simulator for GaN HEMT devices. This is done through the numerical solution of a self-consistent Poisson-Schrodinger system incorporating the polarization effects, evident in wurtzite GaN, and the solution of the continuity equation for the nonequilibrium case is also attempted.

This report discusses the GaN material system (Chapter 2) followed by its implementation in HEMT devices (Chapter 3). The material and physical models used are given in Chapter 4 followed by a discussion of their implementation using numerical simulation (Chapter 5). The subsequent results of the implementation are shown in Chapter 6. This report is concluded (Chapter 7) by giving an outlook and recommendations for future research efforts based on the conclusions reached in this work.

Chapter 2

Gallium nitride material system

This chapter reviews the basic material properties particular to group III nitrides, specifically $\text{Al}_x\text{Ga}_{1-x}\text{N}$. The research history of material system is briefly reviewed and its implementation in various application fields is outlined.

2.1 Group III nitrides

Group III nitrides refer to the compound semiconductors composed of a group III element(s) (Al, Ga, In) and Nitrogen (N) and their associated alloys. These materials' characteristics of distinguishable merit are their wide energy bandgaps and a wurtzite crystal structure which bring about significant polarization effects. GaN and AlGaN have proliferated themselves amongst the III-nitrides for technological reasons but other III-nitrides like InGaN may prove superior in FET¹ applications if the technological issues can be cured that arise due to the big difference in ionic size of In and N [1]. Quaternary alloys, like InAlGaN, provide an extra degree of freedom that allow manipulating the energy bandgap and lattice constants of the material independently but these alloys are also plagued by the high ionicity of Indium.

Table 2.1 provides a comparison of key material properties between prevalent semiconductors. It can be said that GaN consolidates the best properties of GaAs and SiC - materials used in RF and power electronics, respectively - making it suitable for a wide range of applications.

Table 2.1: Comparison of material properties at 300 K [2]

Property	Si	GaAs	SiC	GaN
Bandgap E_g (eV)	1.12	1.42	3.25	3.40
Breakdown field (MV cm^{-1})	0.25	0.4	3.0	4.0
Electron mobility μ ($\text{cm}^2 \text{V}^{-1} \text{s}^{-1}$)	1350	6000	800	1300
Maximum drift velocity v_d (10^7 cm s^{-1})	1.0	2.0	2.0	3.0
Thermal conductivity κ ($\text{W cm}^{-1} \text{K}$)	1.5	0.5	4.9	1.3
Relative dielectric constant ϵ_r	11.8	12.8	9.7	9.0

2.2 Research history of gallium nitride

The first utilisation of gallium nitride in electronic devices started in the 1970's with Pankove *et al.* [3] reporting a GaN-based blue light emitting diode (LED). A commercial

¹Field-effect transistor

realization was hindered by difficulties with producing highly p-doped GaN. A widespread research interest in GaN-based devices withered due to the lack of high quality bulk substrates at the time. To circumvent this problem, GaN epitaxial layers were grown on substrates with big lattice mismatches which led to very high defect densities which seriously diminished the quantum efficiency, power capability and lifetimes of light emitting devices. Research efforts were still limited to optoelectronic devices, at that time, as silicon technology was well established and adequately served the needs for most electronic device application fields.

The late 1980's saw a renewed interest in GaN-based devices with several fundamental breakthroughs, made by Shuji Nakamura, that facilitated the growth of high quality GaN epitaxial layers on sapphire substrates (Al_2O_3) through the use of AlN or GaN nucleation layers [4] and metalorganic chemical vapour deposition (MOCVD). Furthermore, the problem of strongly p-doping GaN - found to be caused by hydrogen passivation [5] - was remedied which made marketable efficient blue LEDs a reality in 1992 [6]. This set the stage for all further developments in GaN-based optoelectronics; a GaN-based laser was realised in 1999.

The first reports of GaN-based transistors appeared in 1993 by Khan *et al.*, showing both HEMT [7] and MESFET [8] structures. This was shortly followed by reports of the first GaN HBT, touted for high-temperature applications [9]. GaN can help attain high speeds and high breakdown voltages in HBTs, however material quality issues, like low p-type doping and short diffusion lengths in epitaxial layers, must first be solved [10]. A transition frequency for GaN HEMTs exceeding 100 GHz was reported in 2000 for the first time [11]. Research efforts continue with the various transistor structures to improve speed and reliability.

2.3 Crystal structure

The group III nitrides may occur in either a wurtzite (hexagonal) or a zincblende (cubic) crystal structure², the former being the thermodynamically stable form at room temperature [1].

The wurtzite crystal structure has a hexagonal unit cell, with lattice constants a and c , formed by two interpenetrating hexagonal close-packed (hcp) sublattices offset by $5c/8$ along the c -axis (figure 2.1). Each sublattice consists of either group III element(s) or Nitrogen forming cations and anions, respectively. Each Nitrogen (group III) atom is coordinated by four group III (Nitrogen) atoms, figure 2.1 shows this tetrahedral arrangement with the cation-anion bonds. The nonsymmetric offset between the sublattices along the c -axis makes the GaN epitaxial layers - commonly grown along the $\{0001\}$ basal plane - either Ga-faced or N-faced giving a polarity to the surface of the material. Each polar surface has unique chemical and physical properties. The polarity of the epitaxial layers are not predicted easily for a specific deposition technique and should be verified with experiment for every structure [12].

A thermodynamically stabilised zincblende form of group III nitrides can be epitaxially grown on specific substrate planes but requires an increased technological effort compared to growing wurtzite structures. The zincblende form of group III nitrides, the same cubic form as Si or GaAs, has gathered interest for applications in optoelectronic devices where the lack of an internal electric field (as opposed to the wurtzite form) supports the light-producing carrier recombination process. This promises an increased quantum efficiency while retaining the wide bandgap property of GaN.

In ternary alloys of III-nitride the specific ordering of the group III atoms with respect to the N atoms may have an effect on the value of the bowing parameters but this effect is insignificant for $\text{Al}_x\text{Ga}_{1-x}\text{N}$ since there is not such a big mismatch in lattice constants between the composing binaries (unlike InGaN or AlInN, for example) [13].

²The rock salt structure does not occur naturally and can only be formed under high pressure

Henceforth, the wurtzite crystal structure along the [0001] direction with Ga-polarity will be assumed in this report, unless stated otherwise.

2.4 Substrates

The growth of pure bulk GaN substrates on commercial scale with wafer diameters greater than 2 inches still eludes the semiconductor industry. However, recent advancements look promising for a cost effective commercial production of GaN wafers up to 4 inches in diameter with low defect densities [?]. The limited supply of pure bulk GaN substrates is currently completely accounted for by the production of blue laser devices to satisfy the great market pull for BluRay™ players and game stations [14]. Most GaN-based applications in RF, power electronics and LEDs must still rely on so-called epiwafers (GaN epitaxially grown on substrates).

Epitaxy (or epitaxial growth) is a process where a thin layer of material is deposited on a single crystal substrate. This layer (epitaxial film) is monocrystalline and assumes the same orientation and lattice structure as the substrate (seed crystal). The methods of metalorganic chemical vapour deposition (MOCVD) and molecular beam epitaxy (MBE) are commonly employed in the production of GaN epiwafers.

The choice of the substrate has an influence on the cost and performance of a device. There are several factors that must be considered when selecting a substrate for a specific application:

- the lattice mismatch between the substrate and epitaxial material(s);
- the difference in thermal expansion coefficients between the substrate and epitaxial material(s);
- the thermal conductivity of the substrate;
- the overall cost and commercial availability of the substrate.

The important material properties of commonly used substrates are compared to GaN and AlN in table 2.2. Sapphire (Al_2O_3) and silicon carbide (SiC-6H) have emerged as the most widely adopted substrates for GaN epiwafers as they provide a good tradeoff between the mentioned criteria. SiC (or diamond), while more expensive than sapphire, has superior thermal conductivity that may be essential in high-power devices [15]. Silicon has also been adopted as a growth substrate [16], despite the greatly mismatched lattice constants, to exploit its low overall cost to drive down the cost of GaN devices.

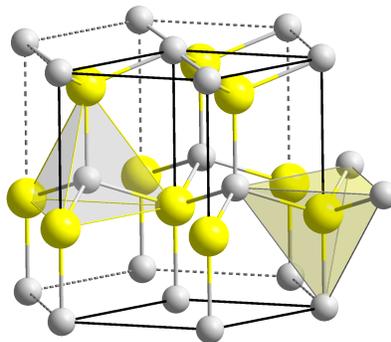


Figure 2.1: Wurtzite unit cell showing interpenetrating *hcp* structures and tetrahedral coordination of constituent elements (Ga: yellow, big; N: grey, small)

The quality of the substrates/epilayer interface is very important as the crystallographic line defects (i.e. edge and screw dislocations) introduced here can erode the performance capability and reliability of a device. Dislocations can propagate through the epitaxial layer to the core of a device where they can significantly decrease the lifetime of a device. Nitride devices suffer from dislocation densities more than five orders of magnitude greater than other compound semiconductors [13]. The strain present in the lattice structure of GaN epiwafers is the major cause (process technological issues aside) of these dislocations.

The two major sources of strain in GaN epiwafers are the mismatches between the lattice constants and the thermal expansion coefficients of the layered materials. The lattice mismatch between two materials induces a compressive or a tensile strain which causes increased defect densities or even substrate cracking. Figure 2.2 clearly shows the mismatch in lattice constants at a GaN-sapphire interface under an electron microscope. The mismatched thermal expansion coefficients provide another source of strain: as two materials cool down from their growth temperature they contract at different rates introducing further material defects during this cool-down phase.

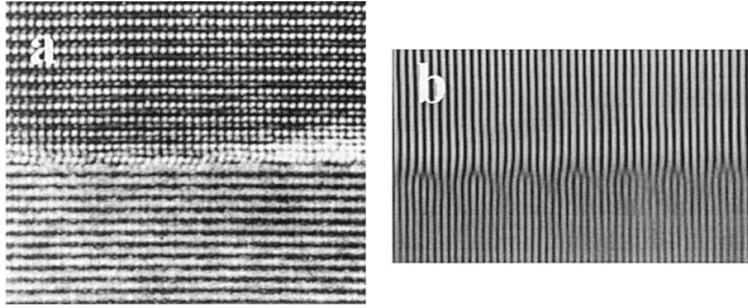


Figure 2.2: HREM image of a GaN sapphire interface (a) and its Fourier filtered version (b)[17]

To reduce the strain in the epitaxial GaN layer nucleation layers of AlN or GaN (depending on the substrate material) are first grown on the substrate to serve as an intermediate layer to better match the lattice constants thereby relieving the strain in the GaN epitaxial layer. The nucleation layer is critical for producing epiwafers of acceptable quality.

2.5 Polarization

Polarization here refers to the polarization density of a dielectric material. It is defined as a dipole moment per unit volume with an SI unit of C m^{-2} and is denoted by the polarization vector \vec{P} . There are two mechanisms involved in the polarization of wurtzite group III nitride, namely piezoelectric and spontaneous polarization. These mechanisms are discussed below.

Table 2.2: Material properties of commonly used substrates for GaN epitaxy (compiled from [1])

	Lattice constant a [\AA]@300 K	Thermal conductivity κ [$\text{W m}^{-1} \text{K}^{-1}$]	Thermal expansion $\frac{\Delta a}{a} \times 10^{-6}$ [K^{-1}]
Sapphire	4.765	25	5.0
SiC-6H	3.081	380	4.46
Si	5.431	156	2.62
GaN	3.20	160	5.6
AlN	3.11	210	4.2

2.5.1 Piezoelectric polarization

Piezoelectric polarization refers to the polarization (an electric state) induced by a mechanical strain in a material. Group III nitrides, like GaN, show very strong piezoelectric polarization effects - an order greater than other III-V semiconductors [1].

The piezoelectric polarization can be intuitively understood as follows: the bond symmetry in the tetrahedral arrangement of the constituent materials (figure 2.3) is perturbed by a strain in the crystal lattice. If the semiconductor crystal is assumed to be unconstrained in the growth direction (c -axis; $[0001]$) the strain is limited to the orthogonal planes. A strain in the (0001) plane, either compressive or tensile, changes the bond angles of atoms at the base vertices thereby changing their vertical components - a dipole moment arises along the $[0001]$ directions.

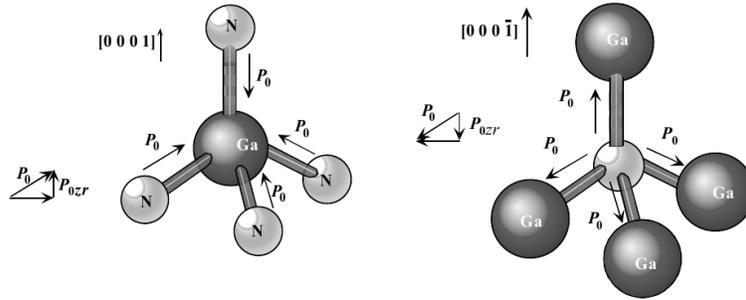


Figure 2.3: Stick-and-ball representation of the tetrahedral arrangement of Ga and N (taken from [1]).

The piezoelectric polarization is related to the strain (both in magnitude and orientation) present in the material. A larger strain causes a larger polarization; compressive strain induces a polarization oppositely directed than a tensile strain. This is portrayed in figure 2.4.

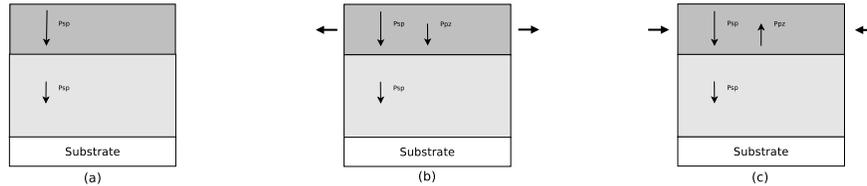


Figure 2.4: Heterostructures showing the magnitude and orientation of the spontaneous and piezoelectric polarization components when the top layer is relaxed (a), under tensile strain (b) and under compressive strain (c).

2.5.2 Spontaneous polarization

Spontaneous polarization can only be measured relative to another 'unstrained' material. Hence, only the difference in spontaneous polarization between two materials (like at a heterointerface) are of significance and not their absolute magnitude. The orientation of the polarization vector (i.e. the 'sign') is determined by the polarity of the material.

The spontaneous polarization (\vec{P}_{sp}) effect occurs in ferroelectrics and pyroelectrics (e.g. III-nitrides). The spontaneous polarization is a direct result of the lacking inversion symmetry along the c -axis of the wurtzite crystal structure influencing the bonding nature of the material and cannot be altered by applying an electric field (unlike in ferroelectrics). The inequivalence between the bonds of nearest neighbours (figure 2.3), where one bond, along the $[0001]$ direction, has a different ionicity (longer bond) compared to the other

three, causes the geometric centres of the negative charge (electrons) and positive charge (nuclei) not to coincide and a dipole moment results. All III-nitrides have a negative value for $P_{sp}^{\vec{}}$ [13] (taking the direction along the c -axis, [0001], as positive).

2.5.3 Pyroelectric effect

The same crystal properties that give III-nitrides their polarization property also lead to the pyroelectric effect. The pyroelectric effect arises in the crystal due to a small displacement of atoms due to a change in temperature leading to a temporary thermally induced electric field. This effect may be significant in high-power amplifiers and lasers. To simulate the pyroelectric effect an energy transport model is required and is not considered here.

2.6 Applications

The inherent material properties of III-nitrides - high temperature stability, high operating voltages and high maximum electron drift velocities - give rise to the selling points of GaN devices: handling high frequencies, high power levels and higher operational linearity. These are traits required by technologies that are, or will soon become, economically significant [15]. The GaN device revenues for 2010 were estimated at almost 7000 million US Dollars³ [14]. This revenue is almost exclusively ascribed to high-brightness LEDs and lasers currently, but GaN presents a gamut of other application fields that are set to grow. It is expensive to maintain several technologies in a production line which makes it favourable to have a single technology, like GaN, that can cover a broad range of applications.

Table 2.3 summarises the performance advantages that the inherent material properties of III-nitrides offer. These are discussed in context of the various application fields below.

Table 2.3: Performance advantages of GaN material properties [?]

Need	Enabling feature	Performance advantage
High voltage operation	High breakdown field	Reduce voltage step-down
High temperature operation	Wide energy bandgap	Reliability; reduced cooling
High efficiency	High operating voltage	Power saving; reduced cooling
High frequency	High peak electron velocity	Bandwidth; microwaves

2.6.1 Optoelectronics

Gallium nitride, with its wide direct energy bandgap, is a huge technological enabler for envisaged applications in the green to ultraviolet spectral range.

The advent of efficient blue GaN LEDs paved the way for the realization of 'white' LEDs which are proliferating themselves in efficient general illumination (which accounts for approximately 20% of the world's electricity consumption [18]). Green GaN LEDs also are a key development for somewhat mundane applications like traffic signals.

Blue GaN-based lasers have enabled a new technology step in high-density optical storage (e.g. BluRayTM) found in consumer electronics. GaN-based lasers are also used in medical applications and laser printers.

³approximately 4858 million Euro at an exchange rate of 1 USD = 0.69 EUR

The wide direct bandgap of GaN allows for efficient UV detectors to be made that can be used for flame detectors or pollution monitors. The high temperature stability of III-nitrides allows these sensors to be operated in harsh corrosive environments.

2.6.2 Radio Frequency electronics

The three major application sectors of radio frequency (RF) electronics for GaN transistors are wireless communication, defence and data broadcast; the major markets for GaN devices in RF electronics are in cellular base transceiver stations (BTS) and military radar. The market value for GaN RF transistors was estimated to lie upwards of 100 million US Dollars⁴ in 2010 [?].

Military applications mostly relate to phased array radars and electronic warfare equipment, especially in the X-band⁵. Phased array radars can consist of up to 1000 antennas each of which has to be individually driven by a power amplifier (PA). The performance of a radar strongly depends on its power output. The required power can be facilitated through high voltages instead of high currents, thanks to the high breakdown voltage of GaN, which allows thinner conductors. The high temperature stability of GaN also allows for reduced cooling which adds to a more compact design.

Approximately 60% of the electricity costs of mobile network operators are attributed solely to the PAs in cellular base stations [18]. New switching amplifier architectures (class E/F) promise improvements up to 75% in efficiency [18]. An implementation using Si-LDMOS - the current dominating technology in the 3G BTS market - is only viable at frequencies up to 1 GHz. The performance advantages of GaN are needed for higher frequency implementations [19] needed for next-generation wireless communication protocols e.g. WiMax and LTE advanced. The superior power density of GaN power transistors (10× compared to GaAs devices) allows power amplifiers to be shrunk to the extent that they can be fitted on top of the BTS mast avoiding the cable losses of present setups.

Data broadcast applications include cable television and very small aperture satellite (VSAT) that can benefit from the increased power output capabilities of GaN transistors.

2.6.3 Power electronics

An estimated 40% of the world's energy is consumed in the form of electricity and this is predicted to increase to 60% by 2040 [18]. Power conversion (rectifiers and inverters) plays a key part in many technologies related to electrical energy e.g. wind turbines, photovoltaics, hybrid vehicles. The wide bandgap of GaN allows devices based on this technology to directly handle high voltages, without step-down, thereby improving conversion efficiency and reducing bulk. GaN devices provide a very low thermal impedance (when grown over SiC substrates) and a reliable high temperature operation which eases the burden of the thermal management.

The mentioned advantages of GaN in power electronics equally apply to SiC - an already established competing technology. However, the strong development of GaN technology in optoelectronic and RF devices may sway the economic advantage to GaN as the gained know-how can be shared amongst several application sectors.

⁴approximately 69 million Euro at an exchange rate of 1 USD = 0.69 EUR

⁵the 8.0 to 12 GHz electromagnetic spectrum defined by the IEEE for radar engineering

Chapter 3

High electron mobility transistors

This chapter reviews the basic operation principle of GaN HEMTs giving a description of their geometry and some physical device characteristics.

3.1 Origin of HEMT structure

The operation of field effect transistors, like MOSFETs, rely on the concept of charge control - being able to influence the carrier concentration in a plane with a gate potential. Group III-nitrides do not allow the MOSFET structure to be used for high performance transistors because there is no high quality gate dielectric available, as is conveniently the case in silicon technology with its native oxide, SiO_2 . The MESFET structure in GaN technology has been used but the required doping and gate proximity to the channel were difficult to attain for high-performance (short-channel) devices [10]. The HEMT structure offers the attractive attributes of close gate proximity to the conducting channel (implying high transconductance) and high drain efficiency which, along with the high low-field mobility, high maximum drift velocity and high electron sheet densities, make GaN HEMTs ideally suited for high-power RF and microwave applications. The high power per width of GaN transistors translates into smaller devices that are easier to manufacture and have a higher impedance which simplifies impedance matching [?].

The HEMT transistor structure was inspired through the experimentation with AlGaAs/GaAs quantum wells and superlattices where the spatially confined, high concentration of carriers at heterointerfaces was discovered [20]. The HEMT structure (figure 3.1) exploits this by 'burying' the 2DEG at the heterointerface, away from the 'rough' surface interface, and controls it through band bending by a gate potential.

3.2 Operation principle of a HEMT

A heterointerface between two materials with a sufficient difference in energy bandgaps is required to achieve the necessary energy band bending to form a 2DEG. The electrons of donor impurities in the larger bandgap material (AlGaAs), near the junction to the smaller bandgap material (GaAs), diffuse across the heterointerface to the lower conduction band where they are confined due to the potential barrier formed by the heterointerface. The resultant 2DEG is spatially separated from the ionized donor atoms and thereby benefits from transport without (or little) ionic scattering in the undoped GaAs region thereby increasing the electron mobility in the channel substantially, hence the name high electron mobility transistor, or HEMT¹. This 2DEG formed at the heterointerface

¹also commonly called MODFET or HFET

can be controlled by applying a potential at the gate of the transistor which bends the energy bands.

The strong polarization effects in III-nitrides (of wurtzite structure) induce high electric fields in the structure that further enhance the carrier concentration and spatial confinement of the 2DEG. A redistribution of free and weakly bound electrons takes place in an attempt to shield the polarization-induced charge. For the HEMT structure considered here, a positive sheet charge is induced at the interface between two materials with different polarizations. This positive, fixed charge attracts mobile electrons which tend to accumulate at the interface thereby adding to the electron concentration of the 2DEG. In theory, the polarization differences at the metal/GaN and GaN/substrate interfaces should also be considered. However, the induced charge at the surface interface is completely shielded by carriers on the metal gate contact. The sheet charge of the GaN/substrate interface may reduce the carrier concentration in the channel if the bulk is very thin, however, if a thick bulk is considered the sheet charge is shielded by the background doping in the bulk.

The source of the mobile electrons in the 2DEG is still a point of debate. In the absence of a metallic contact on the surface of the device a sheet charge of surface states supply electrons, otherwise the n-doped AlGa_xN layer, or carrier injection from the gate, supply electrons.

The 2DEG serves as a current channel that facilitates a high current density. The electron concentration is mainly determined by conduction band offset, the doping concentration in the barrier layer(s) and the polarization effects.

Anomalies in the operation of AlGa_xN/GaN HEMTs can occur due to electrons of the channel getting trapped in the AlGa_xN, GaN bulk and unpassivated surface states, causing current collapse, I-V 'kinks' and long-term instability.

3.3 Structure

The structure of a GaN/AlGa_xN HEMT is shown in figure 3.1 with a [0001] growth direction (Ga-polarity). The strain-relieving nucleation layer(s) is followed by a GaN layer. The GaN layer, known as the bulk, is unintentionally n-doped to a concentration between 10^{14} cm^{-3} and 10^{16} cm^{-3} by impurities (contamination, native defects, O, Si) depending on the specific process.

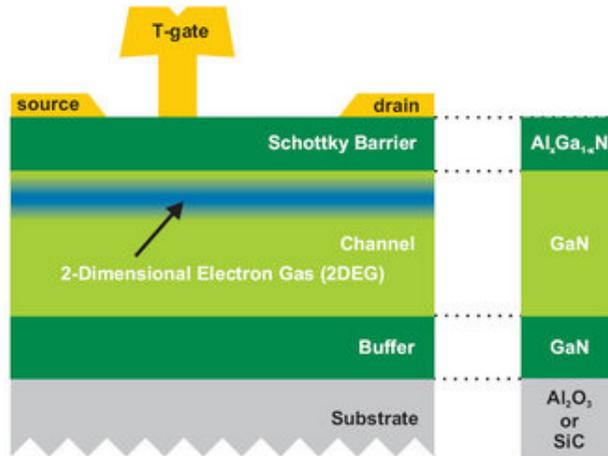


Figure 3.1: Structure of a generic HEMT showing the substrate, nucleation layer, GaN bulk and AlGa_xN layer. The blue field represents the 2DEG that serves as a current channel between drain and source.

The GaN layer is followed by an AlGa_xN layer. The Al_xGa_{1-x}N layer is intentionally n-doped between 10^{17} cm^{-3} and 10^{19} cm^{-3} to supply the mobile electrons for the 2DEG. Increased n-doping in the barrier increases the electron density in the channel but also reduces the mobility and increases the gate leakage of the device. The mole-fraction of Al influences the carrier density in the channel but also the mobility in the 2DEG. An Al-content up to 35% shows an increase in both 2DEG density and mobility. Further increases increase the strain and thereby the defects in the AlGa_xN layer, although Al-content up to 50% has been shown to give increased performance without significant penalties in mobility at room temperature [21]. The doping characteristics with increasing Al-content is not yet well understood but it has been suggested that donor atoms move deeper into the energy bandgap i.e. their ionization energies increase [1] which reduces the achievable n-doping levels.

A thin undoped layer ($\sim 5 \text{ nm}$) of AlGa_xN or AlN is sometimes inserted between the bulk and the barrier layers to shield the 2DEG from the ionized dopants in the adjacent n-doped Al_xGa_{1-x}N. The final epitaxial layer is the cap layer of GaN, its purpose is facilitating ohmic contacts with a low resistance and to modify the effect of surface depletion in the un-gated regions. The gate contact is formed on the cap layer with the drain and source contacts extending down to the 2DEG at the heterointerface to ensure contact with the 2DEG. Since GaN does not form a native oxide, like silicon does, a passivation layer of silicon nitrate (or others) is deposited on the surface of the device.

3.3.1 Breakdown

The breakdown voltage of a semiconductor device is ultimately limited by the inherent material properties of the semiconductor. The breakdown voltage of a semiconductor is related to the energy bandgap of the material which provides III-nitride devices with good credentials for high-voltage operation.

To fully exploit the high-voltage capability of the material in GaN/AlGa_xN HEMTs field plates between the gate and the drain are used to spread the electric field, arising from the gate potential, more uniformly thereby increasing the drain breakdown voltage. The trade-off is more complicated fabrication and increased capacitance. These structural modifications of the 'generic' HEMT structure are not considered here.

The amount of electrons in the current channel (2DEG) limit the maximum current density. The sheet carrier density may be increased by increasing the doping in the AlGa_xN supply layer but the breakdown voltage of the HEMT is thereby decreased [22]. It is suggested in [22] that by spreading the donor doping over several layers and using several heterojunctions a higher current density can be achieved in the channel(s) without decreasing the breakdown voltage of the HEMT since multiple 2DEG channels will be formed. The breakdown effect cannot be handled by the equations considered in this simulator and it is certainly not straight-forward affair to model breakdown effects numerically.

3.3.2 Passivation

Unpassivated surface states can cause current lag through donor-like surface states that capture and release electrons. Passivation reduces the effects of surface states on current lag and increases the breakdown voltage of the device. Surface passivation may improve HEMT characteristics like pinch-off and current lag while deteriorating others like transition frequency and noise figures [23, 24]. The specific effects depend on the passivation layer used and device structure specifics.

3.3.3 Schottky barrier gate contact

The Schottky metal at the surface will deplete all the surface states. The potential at the surface is fixed by the Schottky barrier height and the applied gate voltage. The high defects density in GaN devices introduce many surface states that make it difficult to realize a low-leakage Schottky barrier [10].

Chapter 4

Physical and material models

This chapter reviews the material parameters, material models and the semiconductor equations used to simulate the one dimensional charge distribution of a GaN/AlGaN heterostructure.

4.1 Material models and parameters

The accuracy of a physical device simulator greatly depends on the accuracy of the physical material parameters employed. The data in the literature on some of the physical properties of III-nitrides still shows some inconsistencies and is evolving, especially for the Indium alloys e.g. the InN bandgap controversy [13] and the thermal conductivity of quaternary alloys [13]. Apart from the experimental variations and the different quality of test samples (the strong effect of strain-induced polarization in III-nitrides makes sample preparation especially important), the differences are mostly due to improvement and further development of measurement techniques. This especially holds true for physical material parameters derived by computational techniques which have advanced greatly in recent times[1]. When utilizing computed parameters care should be taken that all the parameters have been calculated by the same method to retain consistency [1]. Likewise, comparison between parameter values that were measured by different techniques requires great diligence. Parameters that show good agreement between the computed (theoretical) and experimental values, naturally, inspire the greatest confidence. Blindly calculating mean values between all published results might prove counterproductive.

Most of the published material parameters of GaN have converged to values that have found wide acceptance. Good reference literature for nitride materials has been established [1, 13, 25] which provides a very good overview and interpretation of the published material properties and parameters. These references, amongst others, have been consulted to select material parameters and their associated physical models.

The material parameters of ternary alloys, like $\text{Al}_x\text{Ga}_{1-x}\text{N}$, are usually calculated by a linear interpolation between the values of the binary compounds (AlN and GaN), known as Vegard's law. Vegard's law holds well for macroscopic quantities like lattice parameters but not always for microscopic properties where a nonlinear dependency on the mole-fraction x exists e.g. energy bandgap. This is remedied by introducing a so-called bowing parameter which leads to a 2nd order polynomial [13]. The nonlinearity found in some material and electronic properties is due to the internal stress caused by the size mismatch between the composing group III elements in the alloy. This is the reason why $\text{Al}_x\text{Ga}_{1-x}\text{N}$ shows a more linear behaviour than InGaN, for example, since the atomic size of Al is much better matched to Ga than is the case for In.

Some material parameters, especially those affected by the fabrication process, are used as empirical 'fit parameters' based on measured device characteristics [13] to 'calibrate' the simulator to match measured results.

The various material models and parameters are now discussed.

4.1.1 Bandgap

The ternary alloy of GaN and AlN, $\text{Al}_x\text{Ga}_{1-x}\text{N}$, provides a wide continuous range of bandgap values with a small associated change in the lattice constant. The energy bandgap of $\text{Al}_x\text{Ga}_{1-x}\text{N}$ is modelled with both a temperature and a compositional dependence. The compositional dependence of the principal bandgap of $\text{Al}_x\text{Ga}_{1-x}\text{N}$ shows a nonlinear increase of the bandgap with increasing Al-content. It is modelled by the empirical equation

$$E_g(x) = xE_g(\text{AlN}) + (1-x)E_g(\text{GaN}) - bx(1-x) \text{ eV} \quad (4.1)$$

with x and b representing the mole-fraction of Al and the bowing parameter, respectively. The values of the bowing parameter still show big inconsistencies in the literature.

The temperature dependence of the bandgap is modelled by the empirical Varshni equation

$$E_g(T) = E_g(0) - \frac{\alpha T^2}{(\beta + T)} \text{ eV} \quad (4.2)$$

with α and β being empirical fit parameters which are given along with other parameters in table 4.1.

Table 4.1: Energy bandgaps of AlN and GaN and empirical fitting parameters

Parameter	Value	Reference
Bandgap $E_g(\text{AlN})$ [eV]	6.1	[1]
Bandgap $E_g(\text{GaN})$ [eV]	3.42	[1]
Bowing parameter b	-1	[1]
Fitting parameter α	2.15×10^{-3}	[13]
Fitting parameter β	1561	[13]
E_V discontinuity [eV]	$0.85x$	[13]

The difference of the bandgaps between the materials present in the heterostructure leads to a discontinuity in the conduction and valence bands at the heterointerface. The value of this 'jump' is not split evenly between the conduction and the valence bands. The maximum value of the discontinuity in the valence band is 0.85 eV and is assumed to increase linearly with increasing Al-content. The associated of the conduction band discontinuity follows directly from this. This value is needed when solving Schrodinger's equation.

4.1.2 Effective carrier masses

An anisotropy in the effective carrier masses is evident for the hole carriers in GaN and AlN in the wurtzite phase; the electrons have an isotropic effective mass. The values used are given in table 4.2. The effective mass for $\text{Al}_x\text{Ga}_{1-x}\text{N}$ is once again a linear interpolation between the values of AlN and GaN (of either transverse or longitudinal values)

$$m^*(\text{Al}_x\text{Ga}_{1-x}\text{N}) = (1-x)m^*(\text{GaN}) + xm^*(\text{AlN}) \quad (4.3)$$

A distinction should be made between the effective carrier mass used for density of states (DOS) calculations and conductivity (mobility) calculations. The geometric mean of the effective masses is used for the DOS calculations, given by

$$m_{\text{DOS}}^* = M_C^{3/2} \sqrt[3]{m_{e-}^{\parallel} (m_{e-}^{\perp})^2} \quad (4.4)$$

with M_C the valley multiplicity (one for GaN). Whereas, the effective mass used for conductivity (mobility) calculations is

$$m_{cond}^* = \frac{3}{\frac{1}{m_{e-}^{\parallel}} + \frac{2}{m_{e-}^{\perp}}} \quad (4.5)$$

Table 4.2: Effective masses of carriers in AlN and GaN [26] (expressed in units of free electron mass m_0).

Mass	AlN	GaN
$m_{e-}^{\perp} (=m_{e-}^{\parallel})$	0.27	0.18
m_{hh}^{\perp}	0.25	0.15
m_{hh}^{\parallel}	3.68	1.10
m_{hh}^{\perp}	6.33	1.65
m_{hh}^{\parallel}	3.68	1.10

4.1.3 Dielectric constant

The relative dielectric constant of the $\text{Al}_x\text{Ga}_{1-x}\text{N}$ alloy is a linear interpolation between the dielectric constants of the binary alloys GaN and AlN and is given by [1]

$$\epsilon_{\text{Al}_x\text{Ga}_{1-x}\text{N}} = 10.28 + 0.03x \quad (4.6)$$

for an Al mole-fraction, x .

4.1.4 Lattice constant

The hexagonal unit cell of the wurtzite structure has two lattice parameters to which Vegard's law can be applied:

$$a_{\text{Al}_x\text{Ga}_{1-x}\text{N}} = 3.1986 - 0.0891x \quad (4.7)$$

$$c_{\text{Al}_x\text{Ga}_{1-x}\text{N}} = 5.2262 - 0.2323x \quad (4.8)$$

These values have been well established and their linear interpolation matches experimental values to within 2% [1].

4.1.5 Crystal strain

A crystal lattice is strained (deformed) when under stress. The stress may be due to several factors which are discussed here and in the following subsection.

Stress-strain relation The mechanical deformation of a crystal can be described using the linear stress-strain relation, known as Hooke's law. In its most general form

$$\sigma_{ij} = \sum_{kl} C_{ijkl} \varepsilon_{kl} \text{ for } i, j, k, l = 1, 2, 3 \quad (4.9)$$

where σ_{ij} , C_{ijkl} and ε_{kl} represent the stress, elastic stiffness and deformation (strain), respectively, along the various directions of the crystal (using tensors of appropriate rank). Due to symmetry considerations of the wurtzite crystal structure $\sigma_{xy} = \sigma_{yx}$, $\sigma_{yz} = \sigma_{zy}$ and $\sigma_{xz} = \sigma_{zx}$ which allows (4.9) to be written in a more compact form

$$\sigma_i = \sum_j C_{ij} \varepsilon_j \quad (4.10)$$

where the Voigt notation is employed: $xx \rightarrow 1$, $yy \rightarrow 2$, $zz \rightarrow 3$, $yz, zy \rightarrow 4$, $xz, zx \rightarrow 5$ and $xy, yx \rightarrow 6$.

The wurtzite crystal structure is assumed unconstrained in the growth direction [0001], therefore only biaxial strain in the (0001) plane remains to be considered ($\sigma_{4,5,6} = 0$).

$$\begin{bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{zz} \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{bmatrix} \begin{bmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \varepsilon_{zz} \end{bmatrix} \quad (4.11)$$

4.1.6 Polarization

There exist both piezoelectric and spontaneous polarization effects that can be treated independently.

4.1.6.1 Piezoelectric polarization

The piezoelectric polarization vector (\vec{P}_{pz}) is determined by the strain (deformation) that a crystal is subjected to when a stress (compressive or tensile) is applied. This is expressed by the relation [1]

$$\vec{P}_{pz} = \overleftarrow{e} \vec{\varepsilon} \quad (4.12)$$

where \overleftarrow{e} represents the piezoelectric tensor (the derivative of polarization w.r.t. strain) and the strain tensor $\vec{\varepsilon}$. Through symmetry consideration of the hexagonal wurtzite crystal structure 4.12 simplifies to [1]

$$\begin{bmatrix} P_x \\ P_y \\ P_z \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & e_{15} & 0 \\ 0 & 0 & 0 & e_{24} & 0 & 0 \\ e_{31} & e_{32} & e_{33} & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \varepsilon_{zz} \\ \varepsilon_{yz} \\ \varepsilon_{xz} \\ \varepsilon_{xy} \end{bmatrix} \quad (4.13)$$

where one remains with three independent components $e_{15} = e_{24}$, $e_{31} = e_{32}$ and e_{33} . The piezoelectric coefficient e_{15} relates to shear strain which is not present in epitaxially grown layers, thus 4.13 reduces to simply

$$P_{pz, z} = e_{31}\varepsilon_{xx} + e_{32}\varepsilon_{yy} + e_{33}\varepsilon_{zz} \quad (4.14)$$

The strain components in the epitaxial layer grown in the z-direction are approximated as

$$\varepsilon_{xx} = \varepsilon_{yy} = \frac{a - a_0}{a_0} \quad (4.15)$$

$$\varepsilon_{zz} = \frac{c - c_0}{c_0} \quad (4.16)$$

where a_0 , c_0 refer to the relaxed lattice constants of the substrate, or material, on which the epitaxial layer is pseudomorphically grown and a , c represent the lattice constant assumed by the epitaxial material. Under these assumptions of a completely relaxed substrate and a coherently strained epitaxial layer 4.14 simplifies to

$$P_{pz, z} = 2e_{31}\varepsilon_{xx} + e_{33}\varepsilon_{zz} \quad (4.17)$$

The above argument can also be developed in terms of stress tensors and piezoelectric moduli, as shown in [1]. By substituting 4.15 in 4.17 and expressing ε_{zz} i.t.o. ε_{xx} using Poisson's ratio one obtains [12]

$$P_{pz, z} = 2\frac{a - a_0}{a_0} \left(e_{31} - e_{33} \frac{C_{13}}{C_{33}} \right) \quad (4.18)$$

where C_{13} and C_{33} are elastic stiffness coefficients. Since the second term (in brackets) is constantly negative it becomes clear that $P_{pz} < 0$ under tensile strain ($a > a_0$) and $P_{pz} > 0$ under compressive strain ($a < a_0$).

The piezoelectric polarization can be calculated using 4.17 since all involved quantities are well established in the literature for GaN and AlN. The value of \vec{P}_{pz} for $\text{Al}_x\text{Ga}_{1-x}\text{N}$ is calculated as an linear interpolation between the values of AlN and GaN

$$\vec{P}_{\text{Al}_x\text{Ga}_{1-x}\text{N}}^{\text{pz}} = [x\overleftarrow{e}_{\text{AlN}} + (1-x)\overleftarrow{e}_{\text{GaN}}] \overrightarrow{\varepsilon}(x) \quad (4.19)$$

The nonlinear dependence of $\vec{P}_{\text{Al}_x\text{Ga}_{1-x}\text{N}}^{\text{pz}}$ on x enters through the nonlinear strain relations for the composing binaries AlN and GaN which given in [1] as

$$P_{\text{AlN}}^{\text{pz}} = \begin{cases} -1.808\varepsilon + 5.624\varepsilon^2 & \text{if } \varepsilon < 0 \\ -1.808\varepsilon - 7.888\varepsilon^2 & \text{if } \varepsilon > 0 \end{cases} \quad (4.20)$$

$$P_{\text{GaN}}^{\text{pz}} = -0.918\varepsilon + 9.541\varepsilon^2 \quad (4.21)$$

Using the material parameters provided in [1] the value for $\vec{P}_{\text{Al}_x\text{Ga}_{1-x}\text{N}}^{\text{pz}}$ on both a AlN and GaN substrate are calculated as

$$P_{\text{Al}_x\text{Ga}_{1-x}\text{N}/\text{GaN}}^{\text{pz}} = -0.0525x + 0.0282x(1-x) \text{ C m}^{-2} \quad (4.22)$$

$$P_{\text{Al}_x\text{Ga}_{1-x}\text{N}/\text{AlN}}^{\text{pz}} = -0.026x + 0.0282x(1-x) \text{ C m}^{-2} \quad (4.23)$$

4.1.6.2 Spontaneous polarization

The advent of the modern theory of polarization (MTP), allowed the accurate, yet simple, computation of P_{sp} by 1st principles [13]. P_{sp} is negative for all Ga-polarity heterostructure-s [12].

The value of $P_{\text{Al}_x\text{Ga}_{1-x}\text{N}}^{\text{sp}}$ exhibits a nonlinear dependence on x which is described by the introduction of a bowing parameter to the usual linear interpolation between values of AlN and GaN

$$P_{\text{Al}_x\text{Ga}_{1-x}\text{N}}^{\text{sp}} = xP_{\text{AlN}}^{\text{sp}} + (1-x)P_{\text{GaN}}^{\text{sp}} + b_{\text{Al}_x\text{Ga}_{1-x}\text{N}}x(1-x) \quad (4.24)$$

with the bowing parameter calculated as

$$b_{\text{Al}_x\text{Ga}_{1-x}\text{N}} = 2P_{\text{AlN}}^{\text{sp}} + 2P_{\text{GaN}}^{\text{sp}} - 4P_{\text{Al}_{0.5}\text{Ga}_{0.5}\text{N}}^{\text{sp}} \quad (4.25)$$

this yields, for the values provided in [1]

$$P_{\text{Al}_x\text{Ga}_{1-x}\text{N}}^{\text{sp}} = -0.09x - 0.034(1-x) + 0.0191x(1-x) \text{ C m}^{-2} \quad (4.26)$$

Table 4.3: Polarization related parameters [13]

	GaN	AlN
e_{13}	-0.338	-0.533
e_{33}	0.667	1.505
C_{13}	103	108
C_{33}	405	373
$P_{\text{sp}} [\text{C.m}^{-2}]$	-0.034	-0.090

4.1.6.3 Multi-layer heterostructure

When an extra material layer is added to the AlGaN/GaN heterostructure, an additional material interface is introduced. This additional layer may be a AlN barrier layer between the AlGaN and GaN layers or a GaN cap layer on top of the AlGaN producing AlGaN/AlN/GaN and GaN/AlGaN/GaN multilayer heterostructures, respectively. The AlGaN/AlN/GaN is considered here but the GaN/AlGaN/GaN structure follows similar arguments.

The GaN layer is assumed to be thick enough to be completely relaxed and the AlN is then pseudomorphically grown on this GaN, adopting the lattice constant of the underlying GaN close to this interface. The strain, and thus piezoelectric polarization, can only be roughly approximated at the AlN/AlGaN interface since the effective lattice constant (and consequently the strain) of the AlN at the second interface is not known. Some relaxation of the lattice occurs as the AlN epitaxial is grown. At the second heterointerface between the AlGaN and AlN it is unclear what the lattice constant of the AlN is that should be used to calculate the strain. If one assumes the strained AlN barrier layer is so thin that it maintains the assumed lattice constant of the underlying GaN throughout the thickness of the layer, the strain at the AlGaN/AlN interface can be calculated using this value (this is the currently employed approach).

It is more likely that there occurs some relaxation of the AlN lattice moving away from the AlN/GaN heterointerface. The typical thickness of the barrier layer (up to 5 nm) is not enough for it to attain its relaxed lattice constant and it is not known at what rate this relaxation occurs (without reverting to more complex computational methods). If one can establish what thickness is needed before the AlN returns to its relaxed lattice constant, when grown on GaN, one could make an interpolation to approximate the strained lattice constant in the thin AlN barrier layer. The use of such a crude approximation may not have a significant effect on the overall results since it would only affect the piezoelectric component of the polarization but the dominating spontaneous polarization will remain unaffected.

To accurately calculate the strain in the entire device structure one would need to minimize the total elastic energy of the structure using, for instance, a conjugate gradient method. This would require introducing further equations that will not be discussed here.

4.1.6.4 Thermally-induced strain

High-power devices and lasers can experience significant self-heating (Joule-heating) under operation. The resulting temperature variations cause a thermally-induced strain in the lattice due to the mismatched thermal expansion coefficients of the various materials. To take this effect into consideration, a phonon-transport model should be coupled to the other semiconductor equations but is not considered here.

4.1.7 Schottky barrier

A Schottky barrier is a metal-semiconductor contact with a high barrier¹ and low doping. For power HEMTs it is desirable to have a Schottky contact at the gate with a high barrier to achieve low gate-leakage currents and high breakdown voltages. Current transport is due to majority carriers generated through thermionic emission (described by Richardson's law). An ideal Schottky barrier between a metal and an n-type semiconductor is characterised by the difference between the metal work function metal (ϕ_m) and the electron affinity (χ) of the semiconductor material i.e. the difference across interface between the energy level of the majority carrier band edge of the semiconductor and the metal Fermi level. This barrier height is calculated as

$$\phi_{Bn} = \phi_m - \chi \text{ eV} \quad (4.27)$$

¹relative to the thermal energy kT ; $\phi_B \gg kT$

or in another form

$$\phi_{Bn} = \phi_m - \phi_{sc} + (E_C - E_F) \text{ eV} \quad (4.28)$$

where ϕ_{sc} , E_C and E_F denote the semiconductor work function and the conduction band and Fermi level energies, respectively.

The work functions for commonly used gate metals and their corresponding Schottky barrier heights with respect to bulk GaN [27] are given in table 4.4. The electron affinity of GaN and AlN are approximately 4.2 eV and 2.05 eV, respectively [28]; the electron affinity of $\text{Al}_x\text{Ga}_{1-x}\text{N}$ decreases with increasing Al mole-fractions. By using a linear interpolation between the electron affinities of GaN and AlN the barrier height between an arbitrary metal and the alloy $\text{Al}_x\text{Ga}_{1-x}\text{N}$ can be calculated as

$$\phi_B(\text{metal} - \text{Al}_x\text{Ga}_{1-x}\text{N}) = \phi_m - \chi_{\text{Al}_x\text{Ga}_{1-x}\text{N}} \quad (4.29)$$

$$\chi_{\text{Al}_x\text{Ga}_{1-x}\text{N}} = \chi_{\text{GaN}} + (1-x)\chi_{\text{AlN}} \quad (4.30)$$

This interpolation, applied to a nickel contact, is found to hold [1]:

$$\phi_B(\text{Ni} - \text{Al}_x\text{Ga}_{1-x}\text{N}) = 0.95 + 2x \text{ eV} \quad (4.31)$$

The linear increase in barrier height with increasing Al mole-fractions holds only up to approximately $x = 0.2$ after which it deviates, presumably due to an increased density of surface defects caused by the increased Al mole-fraction [28]. The barrier height increases monotonically with increasing metal work functions but does not scale linearly with the associated barrier height [27]. This suggests that other factors are at play, like surface states or process defects, that influence the barrier height. This is substantiated by the experimental values for barrier heights that depend on sample preparation and the measurement technique used.

The data reported hitherto are for Schottky contacts between a metal and a bulk semiconductor, as is common in literature. However, the barrier heights for GaN/AlGaN heterostructures, with a strained AlGaN layer, differ from the values measured for their corresponding bulk (i.e. relaxed) counterparts. A lowering of the Schottky barrier height due to polarization effects in GaN/AlGaN heterostructures has been proposed [29]. Furthermore, the normal thermionic emission theory cannot be applied due to the strong piezoelectric polarization in the strained AlGaN layer [30].

The Schottky barrier height for a Ni- $\text{Al}_x\text{Ga}_{1-x}\text{N}$ contact of equation (4.31) is modified for a GaN/AlGaN heterostructure [29] as

$$\phi_B(\text{Ni} - \text{Al}_x\text{Ga}_{1-x}\text{N}) = 0.917 + 1.784x \text{ eV} \quad (4.32)$$

Schottky barriers can also be used to simulate the effect know as Fermi level pinning. This effect is ascribed to surface states and requires the Fermi level to be a certain 'distance' from the conduction band implying, for example, a non-zero built-in electric field for a homogeneous semiconductor. This effect can be reproduced by a Schottky 'contact' with the appropriate barrier height.

Table 4.4: Work functions of metals used for n-GaN Schottky contacts and barrier heights to n-GaN determined by I-V measurements [27]

Metal	Work function ϕ_m [eV]	Barrier height ϕ_B [eV]
Au	5.1	0.87
Pd	5.12	0.94
Ni	5.16	0.95
Pt	5.65	1.01

4.1.8 Surface states

The effect of surface states on the charge distribution in the device structure can be included by using appropriate distribution statistics when calculating the charge density in Poisson's equation (by including the surface states through a δ -doping layer). The surface states are often modelled in a phenomenological manner by assuming the existence of certain acceptor and donor surface states, with a specific ionization energy, at the surface of the device but with no further knowledge on their physical behaviour or origin [31]. It may be the case that their ionization does not actually follow the assumed Fermi-Dirac statistics. A discussion of the incorporation of donor surface states in a simulator is discussed in [32].

4.2 Semiconductor equations

To study the channel formation (electron distribution) in HEMTs, taking quantum effects into account, a self-consistent solution of the Poisson (4.33) and Schrodinger (4.34) equations should be used. Additionally, the continuity equation (4.35) must be solved for the nonequilibrium case.

$$\nabla \cdot (\vec{D}) = \rho \quad (4.33)$$

$$-\frac{\hbar^2}{2m^*} \Delta \psi + V = \mathcal{E} \psi \quad (4.34)$$

$$\nabla \cdot (q\mu_n n \nabla E_{Fn}) = 0 \quad (4.35)$$

The symbols represent: the reduced Planck constant (\hbar), effective mass (m^*), elementary charge quantity (q), the electric displacement field (\vec{D}), the potential function (V), the wave function (ψ) with its associated energy (\mathcal{E}), the electron mobility (μ_n), the electron density (n) and the quasi-Fermi level of the electrons (E_{Fn}).

4.2.1 Poisson's equation

The Poisson equation (4.33) relates the electrostatic potential with the spatial charge distribution in the material. The significant polarization effects in III-nitrides requires the Poisson equation (4.33) to be solved for the electric displacement field

$$\vec{D} = \epsilon \vec{E} + \vec{P} \quad (4.36)$$

where ϵ , \vec{E} and \vec{P} represent the permittivity, electric field and polarization density, respectively. In the absence of a time-varying magnetic field, the electric field may be expressed as

$$\vec{E} = -\nabla \phi \quad (4.37)$$

Substituting (4.36) and (4.37) into (4.33) yields

$$\epsilon \nabla^2 \phi = -\rho + \nabla \cdot \vec{P} \quad (4.38)$$

The charge distribution in the semiconductor is a nonlinear function of the potential

$$\rho(\phi) = [p(\phi) - n(\phi) + N_D^+(\phi) - N_A^-(\phi)] \quad (4.39)$$

through the mobile charge carrier densities, p and n , and the ionized donor N_D^+ and acceptor N_A^- dopants' densities. The nonlinearities are introduced by the expressions used to calculate the carrier densities and the ionized dopant concentration (detailed in 4.2.3).

Electro-mechanical coupling In sections 4.1.5 and 4.1.6, linear elastic theory (Hooke's law) was used assuming that no electro-mechanical coupling exists i.e. no electric field induced strain exists in the crystal structure due to the piezoelectricity of the AlGa_N. Thermodynamics, however, dictates that such a coupling between electrical and mechanical effects exists in piezoelectric materials [10]. In this case the stress tensor and the electric displacement field are coupled and should be considered concurrently; they are expressed as

$$\sigma_{ij} = C_{ijkl}\varepsilon_{kl} - e_{kij}E_k \quad (4.40)$$

$$D_i = e_{ijk}\varepsilon_{jk} - \epsilon_{ij}E_j + P_i^s \quad (4.41)$$

where the meaning of all previously introduced quantities is retained; the second rank permittivity tensor ϵ_{ij} is introduced along with the the spontaneous polarization P_i^s .

The coupling between (4.40) and (4.41) is especially pronounced in materials with a strong piezoelectric effect, like AlGa_N, which calls for a simultaneous treatment of the stress tensor and the electric displacement vector. However, it has been found that the electro-mechanical coupling only has a significant influence on the results in heterostructures in the absence of free carriers - this is not the case in HEMTs and as a result, this coupling can be safely neglected [31]. There have, however, been recent attempts to incorporate the gate-dependent polarization in full-fledged simulators [33].

4.2.2 Schrodinger's equation

The Schrodinger equation is needed to accurately describe the quantum effects that occur at the Ga_N/AlGa_N interface. The discontinuity in the conduction band forms a potential well that leads to a quantization of the energy levels. The electrons in the potential well assume a 2-dimensional behaviour. The solution of the Schrodinger equation gives a wave function associated with an energy - an eigenpair - for the lowest k subbands i.e. the eigenpairs associated with the k smallest eigenvalues. This information allows the specification of a spatial probability distribution of finding an electron at a given energy. The state of the system, the wave function as expressed in (4.34), is a superposition of the wave functions of all the subbands. The equation

$$-\frac{\hbar^2}{2} \frac{d}{dx} \left(\frac{1}{m^*(x)} \frac{d}{dx} \right) \psi_m(x) + V(x)\psi_m(x) = \varepsilon_m \psi_m(x) \quad (4.42)$$

must be solved for $m = 1, 2, \dots, k$ i.e. for each subband. The potential function

$$V(x) = -q\phi(x) + \Delta E_C(x) \quad (4.43)$$

depends on the electrostatic potential ϕ (obtained from Poisson's equation) and the conduction band discontinuity ΔE_C .

At the heterointerface there occurs a discontinuity in the conduction (and valence) bands. When one assumes the level of E_C in the Ga_N bulk to be the reference i.e. $\Delta E_C = 0$, the value of ΔE_C in the AlGa_N layer will be positive since it lies above the reference value. The value of ΔE_C is thus space dependent - this is not always clear from equations quoted in literature. The potential function used here is very simple. Additional terms accounting for electron-electron interactions (the cross-correlation term) can also be added (usually denoted by V_{XC}).

Different Hamiltonians can be used to solve Schrodinger's equation, like effective masses, $k \cdot p$ expansion or tight-binding expansion (one band $k \cdot p$). The latter two approximations are still very sophisticated models that may be needed to accurately describe valence band states and interband transitions of carriers. The effective mass approximation, however, is very accurate if only conduction band processes are of interest as is the case for III-nitride HEMTs [10].

4.2.3 Calculation of charge carrier densities

The charge distribution in the semiconductor, introduced by (4.39), requires the calculation of the charge carrier densities contributed by ionized dopants and mobile charge carriers.

4.2.3.1 Ionization of dopants

Not all dopants are necessarily ionized when they are introduced into a semiconductor crystal, depending on their energy level and the lattice temperature. The ionized donor density is given by

$$N_D^+ = \frac{N_D}{1 + g_D e^{\left(\frac{E_F - E_D}{kT}\right)}} \quad (4.44)$$

where N_D is the donor concentration and $E_D = E_C - E_{ion}$ is the donor energy level, where E_{ion} is the energy required to ionise the dopant. Symbol g_D is the degeneracy in the ground state of the donor dopant energy level (E_D) and is taken to be 2 due to opposite electron spin [34]. A similar equation describes the ionized acceptors

$$N_A^- = \frac{N_A}{1 + g_A e^{\left(\frac{E_A - E_F}{kT}\right)}} \quad (4.45)$$

where E_A is the acceptor dopant energy level and g_A is the ground state degeneracy, taken to be 4 because of the valence band degeneracy of the heavy hole and light hole energy bands at $k = 0$ [34].

This dopant ionization model may be further extended to account for deep and shallow dopant levels by simply introducing additional terms of (4.44) and (4.45) with appropriate E_A and E_D .

4.2.3.2 Mobile charge carriers

The carrier density (electrons or holes) of a semiconductor in equilibrium, within the parabolic band approximation, can be calculated (for electrons) by

$$n = \int_{E_C}^{\infty} g(E) f(E) dE \quad (4.46)$$

where $g(E)$ describes the density of states (DOS) and $f(E)$ is the Fermi-Dirac distribution function. The Fermi-Dirac function (4.47) is a probability density function describing the probability of finding a fermion (e.g. electron) at a specific energy level.

$$f(E) = \frac{1}{1 + e^{\frac{E - E_F}{k_B T}}} \quad (4.47)$$

The 3-dimensional density of states (4.48) are the number of quantum states available per unit volume (in the k -space).

$$g_{3D}(E) = \frac{(2m^*)^{3/2}}{2\pi^2 \hbar^3} \sqrt{E - E_C} \quad (4.48)$$

In structures where one component of the wave vector $\vec{k} = (k_x, k_y, k_z)$ is restricted e.g. quantum wells, a 2-dimensional DOS, describing the number of quantum states per unit area, is needed.

$$g_{2D}(E) = \frac{m^*}{\pi \hbar^2} \quad (4.49)$$

Using (4.48) and (4.47) in (4.46) yields an expression for the 3-dimensional electron density

$$n_{3D} = \frac{(2m^*)^{3/2}}{2\pi^2\hbar^3} \int_{E_C}^{\infty} \frac{\sqrt{E - E_C}}{1 + e^{\frac{E - E_F}{k_B T}}} dE \quad (4.50)$$

and by making the substitution

$$\varepsilon = \frac{E - E_C}{k_B T} \quad (4.51)$$

and defining

$$\eta_F \equiv \frac{E_F - E_C}{k_B T} \quad (4.52)$$

which expresses the distance of the Fermi energy from the conduction band, the integral (4.50) becomes

$$n_{3D} = \frac{(2m^*k_B T)^{3/2}}{2\pi^2\hbar^3} \int_0^{\infty} \frac{\sqrt{\varepsilon}}{1 + e^{\varepsilon - \eta_F}} d\varepsilon \quad (4.53)$$

The Fermi integral of order $1/2$ is defined as

$$F_{1/2}(\eta_F) \equiv \int_0^{\infty} \frac{\sqrt{\varepsilon}}{1 + e^{\varepsilon - \eta_F}} d\varepsilon \quad (4.54)$$

Following the same procedure for the 2-dimensional electron density using (4.49) in (4.46) one obtains

$$n_{2D} = \frac{m^*k_B T}{\pi\hbar^2} \int_0^{\infty} \frac{\varepsilon^0}{1 + e^{\varepsilon - \eta_F}} d\varepsilon \quad (4.55)$$

where the Fermi integral of order 0 is defined as

$$F_0(\eta_F) \equiv \int_0^{\infty} \frac{\varepsilon^0}{1 + e^{\varepsilon - \eta_F}} d\varepsilon \quad (4.56)$$

The general expression for the Fermi integral is given by

$$\mathcal{F}_k(x) \equiv \frac{1}{\Gamma(k+1)} \int_0^{\infty} \frac{t^k}{1 + e^{t-x}} dt, \quad k > -1 \quad (4.57)$$

where Γ is the gamma function defined as

$$\Gamma(k) = (k-1)! \quad (4.58)$$

and has the property

$$\Gamma(k+1) = k\Gamma(k) \quad (4.59)$$

For $k = 1/2$ it is defined as

$$\Gamma(1/2) = \sqrt{\pi} \quad (4.60)$$

Adopting the general Fermi expression (4.57) in (4.53) one obtains

$$n_{3D} = N_{3D} \mathcal{F}_{1/2}(\eta_F) \quad (4.61)$$

where N_{3D} is the effective density of states, given by

$$N_{3D} = 2 \left(\frac{m^*k_B T}{2\pi\hbar^2} \right)^{3/2} \quad (4.62)$$

Similarly, for (4.55)

$$n_{2D} = N_{2D} \mathcal{F}_0(\eta_F) \quad (4.63)$$

$$N_{2D} = \frac{m^* k_B T}{\pi \hbar^2} \quad (4.64)$$

The Fermi integrals cannot be solved analytically and have to be evaluated numerically, in the general case. An analytical solution exists for the 0th order Fermi integral (4.56)

$$\mathcal{F}_0(\eta_F) = \int_0^\infty \frac{\varepsilon^0}{1 + e^{\varepsilon - \eta_F}} d\varepsilon = \ln(1 + e^{\eta_F}) \quad (4.65)$$

Using the following property of Fermi integrals

$$\frac{d}{dx} \mathcal{F}_k = \mathcal{F}_{k-1} \quad (4.66)$$

its solution can be extended to any $k \leq 1$ with an analytical expression of the Fermi integral.

4.2.3.3 Quantum carrier density

The quantum electron density is a function of the electrostatic potential, ϕ , through the dependence of the eigenenergy $E_n(\phi)$ and wave function $\psi_n(\phi)$. The quantum electron density (similarly for holes) is obtained by summing over all energy bands (i.e. eigenpairs) multiplying the occupancy N_n^q of the n^{th} energy band by the squared modulus of the wave function²:

$$n_q(\phi) = \sum_n N_n^q |\psi_n(\phi)|^2 \quad (4.67)$$

$$N_n^q \equiv \frac{1}{\pi} \left(\frac{2m^* k_B T}{\hbar^2} \right)^{1/2} \mathcal{F}_{-1/2} \left(\frac{E_F - E_n(\phi)}{k_B T} \right) \quad (4.68)$$

$$\mathcal{F}_{-1/2}(\eta_F) = \frac{2}{\sqrt{\pi}} F_{-1/2}(\eta_F) \quad (4.69)$$

The squared modulus of the wave function expresses a spatial probability distribution function i.e. the probability of finding an electron, with an energy E_n , at a specific point in space. This probability should also be multiplied by the occupancy: the probability of the electron occupying a state at energy E_n , described by the Fermi-Dirac distribution introduced before, and the appropriate density of states (which would be 2-dimensional for the quantum well we wish to describe). Taking this into consideration, the expression follows

$$n_q(\phi) = \frac{1}{\pi} \left(\frac{2m^* k_B T}{\hbar^2} \right)^{1/2} \sum_n \psi_n^2(\phi) \mathcal{F}_{-1/2} \left(\frac{E_F - E_n(\phi)}{k_B T} \right) \quad (4.70)$$

4.2.4 Continuity equation

When a gate potential is applied, the device is no longer in thermal equilibrium and the Fermi level is not constant throughout the structure in this case. The Fermi level must be solved using the continuity equation which is expressed in the following form

$$\frac{\partial n}{\partial t} = \frac{1}{q} \nabla \cdot \vec{J}_n + G \quad (4.71)$$

where J_n is the electron current density and G the net generation-recombination term. Since we are only considering steady-state solutions the temporal variation is ignored i.e.

²in the time-independent Schrodinger equation the wave functions are real and the modulus (i.e. complex conjugate) is not needed

$\frac{\partial n}{\partial t} = 0$. Furthermore, the gate current is assumed to be very small implying that the electron density does not deviate far from its equilibrium value. This makes the neglect of the generation-recombination term, $G = 0$, a reasonable assumption.

The best-known expression for current density is the drift-diffusion model:

$$\vec{J}_n = q\mu_n\vec{E} + qD_n\nabla n \quad (4.72)$$

where μ_n and D_n denote the electron mobility and diffusivity coefficients, respectively; the electric field is $\vec{E} = -\nabla\phi$. By using a linearization of the distribution function (of carriers) around the equilibrium point the current density can also be expressed as

$$\vec{J}_n = -q\mu_n n \nabla E_{Fn} \quad (4.73)$$

where E_{Fn} is the quasi-Fermi level for electrons. By substituting (4.73) into (4.71) and considering the mentioned assumptions, one obtains

$$\nabla \cdot (\mu_n n \nabla E_{Fn}) = 0 \quad (4.74)$$

The calculation of the electron-current across the heterointerface may require the use of thermionic emission theory, as proposed in [35]. The 2D electrons that are confined within the potential well have their motion constricted in the direction of the gate current flow; only unconstrained electrons in the upper subbands should contribute to the gate current. The consideration of this in the code implementation requires special attention. One solution could be to neglect the contribution of the bottom three or four subbands, that represent the quantized electrons in the quantum well, when calculating the electron density for continuity equation calculations.

4.3 Quasi-2D charge transport model

The current-voltage (I-V) characteristics of the HEMT can be calculated using a quasi-2D charge transport model, as detailed in [36, 37, 38]. Firstly the sheet charge density is calculated at several 1D sections perpendicular to the gate. This is done by solving the Schrodinger-Poisson system, using the appropriate gate voltage as a boundary condition. Using the calculated sheet charge density, the current flow in the parallel direction is calculated using a transport model e.g. drift-diffusion (DD) or hydrodynamic (HD) models. The The basic assumptions common for a quasi-2D simulator are

- negligible current flow in the perpendicular direction (to the layers);
- one-dimensional current density, restricted to the current channel (2DEG);
- linear variation of the voltage between the drain and the source of a biased transistor;
- only DC and steady-state conditions are calculated.

This quasi-2D model neglects the interaction between electrons in the parallel and perpendicular directions but, nonetheless, allows many physical effects to be incorporated in the model in a computationally efficient manner [37].

Chapter 5

Numerical simulation

This chapter discusses the numerical simulation of the system of coupled equations discussed in Chapter 4. An iterative solution approach to the equation system is discussed along with the discretization and solution of the individual equations using numerical schemes.

5.1 Solving the system of equations

The semiconductor equations presented in section 4.2 (repeated below for convenience) have to be solved self-consistently since they are coupled amongst each other. To solve such a system one can either solve all three equations simultaneously or use an iterative method where the separate equations are solved successively until convergence is achieved. The latter approach is preferred since it is easier to implement (each equation can be tested independently) and it also allows for an easier implementation of 'numerical tricks' to aid in convergence.

$$\epsilon \nabla^2 \phi = -\rho(\phi, \psi, E_F) + \nabla \vec{P} \quad (5.1)$$

$$-\frac{\hbar^2}{2m^*} \Delta \psi + V(\phi) = \epsilon \psi \quad (5.2)$$

$$\nabla \cdot (\mu_n n(\phi, \psi, E_F) \nabla E_F) = 0 \quad (5.3)$$

For the equilibrium case (when no gate voltage is applied) only the Schrodinger and the Poisson equations - the so-called SP system - needs to be solved. When a gate voltage is applied the continuity equation must be solved in addition to the SP system. First, the self-consistent solution of the SP system will be discussed and then the solution for the nonequilibrium case where all three equations must be solved self-consistently.

5.1.1 Self-consistent solution of Schrodinger-Poisson system

The equations (5.1)-(5.2) form a system of equations that are coupled through the quantum electron density. The solution of Poisson's equation yields a potential ϕ which depends on the value of the quantum electron density which, in turn, is determined by solving Schrodinger's equation which again depends on the potential.

To obtain a self-consistent solution of equations (5.1)-(5.2) an iterative technique is used where the two equations are solved sequentially [39, 40]. The flowchart in figure 5.1 outlines the basic procedure: If no initial value for the potential is available (a guess or an analytic approximation), Poisson's equation is first solved assuming a purely classical electron density to obtain ϕ_1 . Then the iterative procedure is started. For the k-th iteration the Schrodinger equation is solved to obtain a quantum electron density n_q^k

(using the resulting band energies and wave functions). Poisson’s equation is then solved again, using n_q^k , to obtain a new potential, ϕ_k . The process is repeated until the values of the potential converges to within a set tolerance, otherwise we replace $\phi_{k+1} \rightarrow \phi_k$ and repeat the process.

This system is intrinsically oscillatory [40]; a simple iteration between the Poisson and Schrodinger equations fails to converge because the quantum charge density changes the potential profile too much between iterations. The physical justification for this numerical oscillation is given by the high sensitivity that the energy values have on the confining potential and, in turn, the strong dependence of the quantum charge density on the energy values [40]. This makes the initial guess of the potential and the updating algorithm critical to ensure (fast) convergence. The updating algorithm governs how the variables are updated between iterations. There are several possible approaches to damp this numerical oscillation, two of which will now be discussed.

5.1.1.1 Under-relaxation

The under-relaxation method is the most obvious and very easy to implement. The updated carrier densities (or alternatively the potential) is damped sufficiently between iterations to obtain convergence

$$n_{new, damped} = \alpha n_{new} + (1 - \alpha) n_{old} \quad (5.4)$$

The convergence behaviour depends on the selection of the damping coefficient, $\alpha \in [0, 1]$, and is typically selected heuristically based on numerical experiments. If the damping coefficient is too large, the system oscillates and fails to converge; if the damping coefficient is too small the system converges slowly and requires (too) many iterations to converge. This method generally exhibits a poor convergence rate and is not feasible for computationally intensive iterations unless it is more refined e.g. one could dynamically adjust the value of the damping coefficient based on the norm of the change of a variable between iterations.

5.1.1.2 Predictor-corrector

If an expression could be derived describing the quantum electron density as only a function of the potential (opposed to a function of the wave functions and energies), the

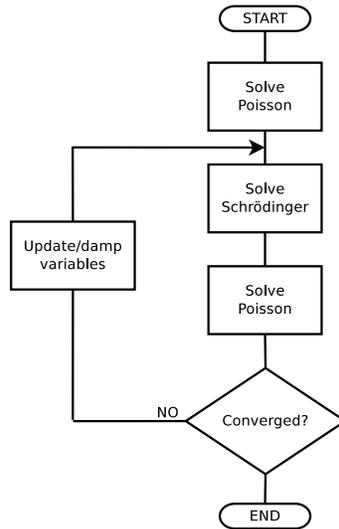


Figure 5.1: Flowchart outlining the approach to solving the Schrodinger-Poisson system self-consistently in an iterative manner.

quantum electron density could be calculated without solving the Schrodinger equation. By substituting this (hypothetical) function ($n_q(\phi)$) in 5.1, one could only solve a non-linear Poisson equation of the form

$$\epsilon \nabla^2 \phi = -\rho(n_q(\phi), \phi) - \nabla \vec{P} \quad (5.5)$$

The correct quantum electron density could then be simply calculated by substituting the solution ϕ into $n_q(\phi)$. The function $n_q(\phi)$ is not known but one may approximate such a function - the so-called predictor.

The Poisson equation is solved for ϕ , using a predictor of the quantum density $\tilde{n}_q(\phi)$ (the predictor step). This value of ϕ is then used when solving Schrodinger's equation to obtain the actual quantum electron density (the corrector step). These steps are repeated until convergence of the SP system is achieved. Trellakis *et al.* [39] first published the damping of the Schrodinger-Poisson system using this predictor-corrector algorithm, using a predictor of the quantum charge derived by first-order perturbation theory. The derivation of the quantum electron density predictor and the solution of the subsequent nonlinear Poisson equation are discussed below.

The convergence behaviour of the SP system is dramatically improved using this predictor-corrector approach compared to a simple under-relaxation method and is therefore preferred.

Quantum electron density predictor To derive a predictor of the change in the quantum carrier density, first-order perturbation theory is applied. A perturbation of the potential

$$\phi \rightarrow \phi + \delta\phi \quad (5.6)$$

changes the Hamiltonian of Schrodinger's equation

$$\hat{H} \rightarrow \hat{H} - q\delta\hat{\phi} \quad (5.7)$$

and the quantum carrier density

$$n_q(\phi) \rightarrow n_q(\phi + \delta\phi) = n_q(\phi) + \delta n_q(\phi, \delta\phi) \quad (5.8)$$

Exploiting the properties of the Fermi integral and some first-order perturbation theory, the perturbation $\delta n_q(\phi, \delta\phi)$ can be approximated as

$$\delta \tilde{n}_q(\phi, \delta\phi) = N_q \sum_n \psi_n^2(\phi) \mathcal{F}_{-3/2} \left(\frac{E_F - E_n(\phi)}{k_B T} \right) \frac{q\delta\phi}{k_B T} \quad (5.9)$$

This can then be applied to (5.8) and then using the Fermi-integral property (4.66) again, one obtains

$$\tilde{n}_q(\phi + \delta\phi) = N_q \sum_n \psi_n^2(\phi) \mathcal{F}_{-1/2} \left(\frac{E_F - E_n(\phi) + q\delta\phi}{k_B T} \right) \quad (5.10)$$

which is equivalent to the original expression (4.70) with $E(\phi) \rightarrow E_n(\phi) - q\delta\phi$. The prediction of the quantum density at the potential $\phi + \delta\phi$ is given by (5.10). Essentially one assumes that the wave functions remain unchanged at variations of the potential and that the eigenvalues are shifted according to the potential perturbation. It is clear that once the solution converges (i.e. $\delta\phi \rightarrow 0$) then the equation tends towards the unperturbed quantum density expression (4.70). For a more complete treatment of this approach refer to [39].

5.1.2 Nonequilibrium case

In the nonequilibrium case the continuity equation has to be solved for quasi-Fermi level E_{F_n} in a self-consistent manner with the Schrodinger and Poisson equations. To obtain such a self-consistent solution, the three equations are solved in succession until convergence is achieved. It is not clear what the best method is to achieve convergence - there are several possibilities:

1. Solve the SP system self-consistently, then the continuity equation (figure 5.2(a)).
2. Solve the Poisson and continuity equations self-consistently, then the Schrodinger equation (figure 5.2(b)).
3. Solve all three equations, one after the other (figure 5.2(c)).

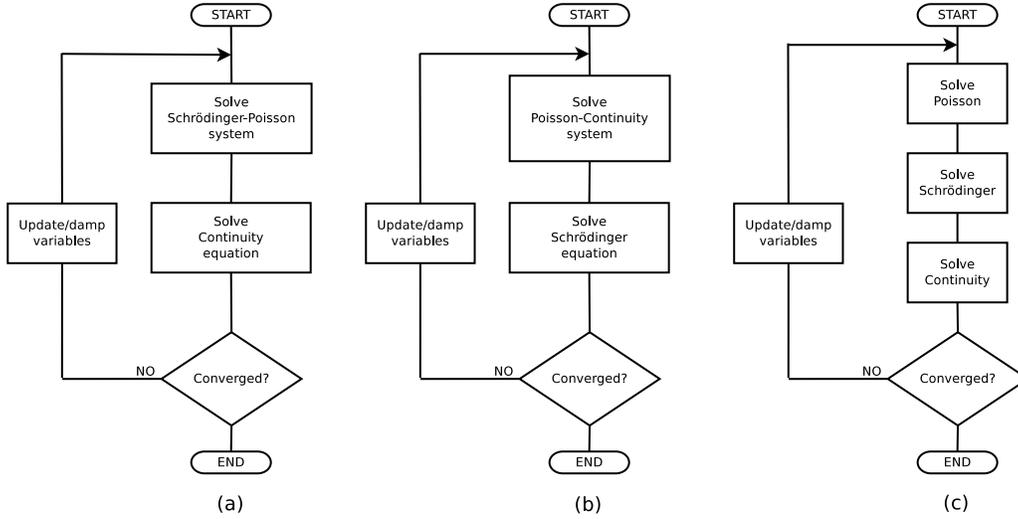


Figure 5.2: Flowchart outlining three possible approaches to solving the Poisson, Schrodinger and continuity equations self-consistently as detailed in the text.

It should be investigated which approach is feasible since the implementation of these schemes have not been tested thoroughly. It is not clear, for instance, how well the corrector-predictor stabilisation of the SP-system works if the continuity equation is solved additionally. Whether the quantum density predictor 5.10 should also be used in the continuity expression is also not known. The value of E_{F_n} may need to be damped to achieve convergence. Further comment on the attempted approaches is given in subsection 5.2.4.5.

5.2 Numerical schemes

This section discusses the discretization of the equations highlighted in section 4.4 and their subsequent solution using numerical methods.

5.2.1 Spatial discretization

A nonuniform spatial discretization is used (if specified by the user). A representation of the discretized space is given in figure 5.3. There are a total of N nodes, numbered from 1 to N ; $N-2$ 'internal' nodes and 2 boundary nodes. The length of the $N-1$ line segments are denoted by Δx_i for a line segment between nodes i and $i+1$.

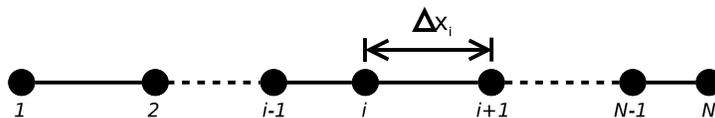


Figure 5.3: Representation of nonuniform spatial discretization showing node numbering.

The small variation in both carrier densities and potential within the GaN bulk region allows a coarse spatial resolution. Aiming for a node-to-node variation in potential of less than the thermal voltage (~ 26 mV) is a good reference value. A finer spatial resolution should be used in regions where large electron density gradients are present/expected e.g. around the potential well at the heterointerface (3 Å was used to generate presented results in Chapter 6).

By using a nonuniform mesh the symmetry of the matrices of the discretized equations is lost (as will become apparent in the following subsections) which negatively affects the performance of the algorithms used to solve the linear systems. Symmetry can be regained by applying transformations on the matrices but since all the linear systems are solved using MATLAB's robust procedures, this does not have to be done explicitly.

5.2.2 Poisson's equation

Rewriting (4.38) for the one-dimensional case, substituting (4.39), yields the equation

$$\epsilon\phi_{xx} = - [p(\phi) - n(\phi) + N_D^+(\phi) - N_A^-(\phi)] + P_x \quad (5.11)$$

which is a nonlinear elliptic equation in ϕ . It is discretized using the finite volumes method and then solved using the Newton-Raphson method (detailed in section 5.2.2.3).

Using the finite volumes discretization, the introduction of δ -doping layers (sheet charge) can be easily added.

5.2.2.1 Discretization by finite volumes

At the material interface(s) of the heterostructure a discontinuity of the material parameters occurs. The discontinuity may introduce numerical problems, especially for the spatial derivative of the polarization, P_x . This may be partly cured by introducing a gradual change in the Al-content (x) - as would be the case in an actual heterostructure - but the rate of change is (generally) not known.

Instead, the finite volume method is used to manage these material discontinuities in a more robust manner. By integration of (5.11) one obtains a weak formulation. Equation (5.11) is repeated here in more compact notation:

$$\epsilon\phi_{xx} = -\rho + P_x \quad (5.12)$$

Integrating over the element associated with node i (i.e. the line segment $x \in [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$) one obtains

$$\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \epsilon\phi_{xx} dx = - \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \rho dx + \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} P_x dx \quad (5.13)$$

By employing the second fundamental theorem of calculus¹ it is valid to write the definite integral of the polarization derivative, despite its discontinuity, as

$$\left[\epsilon \left(x_{i+\frac{1}{2}} \right) \phi_x \left(x_{i+\frac{1}{2}} \right) - \epsilon \left(x_{i-\frac{1}{2}} \right) \phi_x \left(x_{i-\frac{1}{2}} \right) \right] = -\rho(x_i) \left[x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}} \right] + \left[P \left(x_{i+\frac{1}{2}} \right) - P \left(x_{i-\frac{1}{2}} \right) \right] \quad (5.14)$$

The charge density (ρ) is assumed to be constant over the element (an exponential interpolation should be considered). The first-order derivatives are discretized using a

¹Let f be a real-valued function defined on a closed interval $[a, b]$ that admits an antiderivative g on $[a, b]$. That is, f and g are functions s.t. $\forall x \in [a, b], f(x) = g'(x)$. If f is integrable on $[a, b]$ then $\int_a^b f(x) dx = g(b) - g(a)$

first-order finite difference approximation, giving the left-hand side (LHS) and right-hand side (RHS) of (5.14) as

$$\text{LHS} = \left[\epsilon_{i+\frac{1}{2}} \frac{\phi_{i+1} - \phi_i}{\Delta x_{i-1}} - \epsilon_{i-\frac{1}{2}} \frac{\phi_i - \phi_{i-1}}{\Delta x_{i-1}} \right] \quad (5.15)$$

$$\text{RHS} = -\rho(x_i) \left[x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}} \right] + \left[P_{i+\frac{1}{2}} - P_{i-\frac{1}{2}} \right] \quad (5.16)$$

where $\Delta x_i = x_{i+1} - x_i$ and $P_{i+\frac{1}{2}}$ is the total polarization at the centre of the line segment between the nodes $i + 1$ and i .

The equations (5.15) and (5.16) can be rewritten in a form suitable for matrix implementation:

$$\frac{\epsilon_{i+\frac{1}{2}}}{\Delta x_i} \phi_{i+1} - \left(\frac{\epsilon_{i+\frac{1}{2}}}{\Delta x_i} + \frac{\epsilon_{i-\frac{1}{2}}}{\Delta x_{i-1}} \right) \phi_i + \frac{\epsilon_{i-\frac{1}{2}}}{\Delta x_{i-1}} \phi_{i-1} = -\rho_i \left(x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}} \right) + \left(P_{i+\frac{1}{2}} - P_{i-\frac{1}{2}} \right) \quad (5.17)$$

This allows the system to be written in the form $\mathbf{A}\phi = \mathbf{b}$ where

$$\mathbf{A} = \begin{bmatrix} \star & \star & 0 & \dots & 0 \\ \frac{\epsilon_{2-\frac{1}{2}}}{\Delta x_1} & -\left(\frac{\epsilon_{2+\frac{1}{2}}}{\Delta x_2} + \frac{\epsilon_{2-\frac{1}{2}}}{\Delta x_1} \right) & \frac{\epsilon_{2+\frac{1}{2}}}{\Delta x_2} & 0 & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & 0 & \frac{\epsilon_{i-\frac{1}{2}}}{\Delta x_{i-1}} & -\left(\frac{\epsilon_{i+\frac{1}{2}}}{\Delta x_i} + \frac{\epsilon_{i-\frac{1}{2}}}{\Delta x_{i-1}} \right) & \frac{\epsilon_{i+\frac{1}{2}}}{\Delta x_i} & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & 0 & \frac{\epsilon_{N-1-\frac{1}{2}}}{\Delta x_{N-2}} & -\left(\frac{\epsilon_{N-1+\frac{1}{2}}}{\Delta x_{N-1}} + \frac{\epsilon_{2-\frac{1}{2}}}{\Delta x_{N-2}} \right) & \frac{\epsilon_{N-1+\frac{1}{2}}}{\Delta x_{N-1}} \\ & & & & 0 & \star & \star \end{bmatrix}$$

$$\mathbf{b} = \begin{bmatrix} -\rho_i \left(x_{2+\frac{1}{2}} - x_{2-\frac{1}{2}} \right) + \left(P_{2+\frac{1}{2}} - P_{2-\frac{1}{2}} \right) \\ \vdots \\ -\rho_i \left(x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}} \right) + \left(P_{i+\frac{1}{2}} - P_{i-\frac{1}{2}} \right) \\ \vdots \\ \rho_i \left(x_{N-1+\frac{1}{2}} - x_{N-1-\frac{1}{2}} \right) + \left(P_{N-1+\frac{1}{2}} - P_{N-1-\frac{1}{2}} \right) \end{bmatrix}$$

The value of the elements denoted with the place holder \star depend on the type of boundary conditions that are enforced on the system and will be discussed hereafter.

The charge density ρ is currently assumed to be constant over each element, an exponential interpolation between nodal values may be sensible considering the high gradients ∇n present around the heterointerface.

5.2.2.2 Boundary conditions

Two boundary conditions must be specified corresponding to the gate electrode and the interface of the bulk GaN to the growth substrate. The gate contact is assigned a Dirichlet boundary condition specifying the electrostatic potential as the difference between the applied bias potential and the Schottky barrier height. The GaN/substrate interface is assigned a Neumann boundary (the derivative of the potential directed normal to the interface) condition of zero, implying that no electric field is present in a direction normal to the boundary surface. This implies that no current flows out of this contact. The validity of using the latter boundary condition under nonequilibrium conditions is not clear. A current does indeed flow when a gate voltage is applied. Perhaps a zero Dirichlet condition should be used.

The Dirichlet boundary condition at the gate contact is enforced by setting

$$\frac{1}{\Delta x_1} \phi_1 = \frac{qV_G - \phi_b}{\Delta x_1}$$

by replacing the appropriate \star matrix elements in \mathbf{A} and \mathbf{b} , where ϕ_b is the Schottky barrier height and qV_G the applied gate potential (both in eV). The Neumann boundary condition at the opposing barrier is enforced by approximating the derivative as

$$\frac{1}{\Delta x_{N-1}} \phi_{N-1} - \frac{1}{\Delta x_N} \phi_N = 0$$

Alternatively, one could also apply a zero Dirichlet boundary condition at the substrate boundary. At least one of the contacts should have a Dirichlet boundary condition for the PDE to be well-posed.

5.2.2.3 Newton-Raphson method

The nonlinear Poisson system is solved using the classic Newton-Raphson method along with an inexact line search to provide more robust convergence properties. A general definition of the method is given here. The implementation for the Poisson equation is rather straightforward (*cf.* `runstructure.m`).

Newton-Raphson method The Newton-Raphson method is surely the workhorse of numerical methods, finding application in countless applications. Its popularity stems from its fast (quadratic) local convergence properties and relative ease of implementation. A geometrical interpretation of the Newton-Raphson method is shown in figure 5.4. Starting at an arbitrary point arbitrary point x_i , a tangent is extended until it crosses zero at point x_{i+1} where a new value, $f(x_{i+1})$, is calculated. This process is repeated until the point x_n is found that satisfies $|f(x_n)| \leq \epsilon$. This point corresponds to the root of the function $f(x)$ (within the set tolerance ϵ).

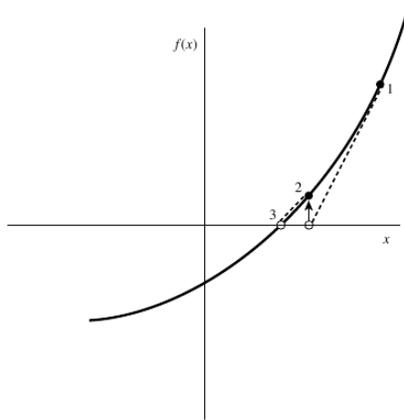


Figure 5.4: Geometrical representation of the Newton-Raphson method (taken from [41])

The Newton-Raphson method is derived algebraically from the Taylor expansion around x

$$f(x + \delta x) = f(x) + f'(x)\delta x + \frac{f''(x)}{2}\delta x^2 + O(\delta x^3) \quad (5.18)$$

By setting $f(x + \delta x) = 0$ and dropping higher order terms ($O(\delta x^2)$) one obtains an estimate for the correction

$$\delta x = -\frac{f(x)}{f'(x)} \quad (5.19)$$

which can be expressed as

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)} \quad (5.20)$$

Ignoring the higher order derivatives yields useless corrections if the initial guess of the root, x_0 , is too far from the actual root and includes local minima and maxima in the interval.

The Newton-Raphson formula requires the efficient evaluation of the function $f(x)$ and its derivative at each iteration step. It is highly desirable to have an analytic expression for the derivative $f'(x)$ since a numerical evaluation would destroy the fast convergence properties of the method, making the secant method a better alternative in the one-dimensional case [41].

$$F_i(\bar{x} + \delta\bar{x}) = F_i(\bar{x}) - \sum_{j=1}^N \frac{\partial F_{ij}}{\partial x_j} \delta x_j \quad (5.21)$$

$$J_{ij} \equiv \frac{\partial F_{ij}}{\partial x_j} \quad (5.22)$$

$$\bar{F}(\bar{x} + \delta\bar{x}) = \bar{F}(\bar{x}) - \bar{J}\delta\bar{x} \quad (5.23)$$

setting $\bar{F}(\bar{x} + \delta\bar{x}) = 0$ one obtains a correction vector

$$\delta\bar{x} = -\bar{J}^{-1}\bar{F}(\bar{x}) \quad (5.24)$$

such that

$$\bar{x}_{i+1} = \bar{x}_i + \delta\bar{x} \quad (5.25)$$

Building on the rapid local convergence behaviour of the Newton-Raphson method, the method is extended by the inexact line search technique to improve its global convergence properties.

Line search The goal of the Newton method is to find the value \bar{x} which yields $\bar{F}(\bar{x}) = 0$. The correction values of (5.24) might direct the algorithm away from the root when the current guess is far from the root and higher order terms become important. A criterium must be established on whether to accept the calculated correction step or not. A minimization of

$$f = \frac{1}{2}\bar{F} \cdot \bar{F} = \frac{1}{2}|\bar{F}|^2 \quad (5.26)$$

with each correction step provides a suitable criterium. The correction vector $\delta\bar{x}$ is always in the direction of descent for f . This is shown by the following argument

$$\nabla f \cdot \delta\bar{x} = (\bar{F} \cdot \bar{J}) \cdot (-\bar{J}^{-1} \cdot \bar{F}) = -\bar{F} \cdot \bar{F} \leq 0 \quad (5.27)$$

In a first attempt the full correction step $\delta\bar{x}$ is used, if it does not minimize f the step size is reduced by moving back until a point is reached where f is reduced. The best factor of correction is found using a line search algorithm where the goal is to find $\lambda \in (0, 1]$ such that $f(\bar{x} + \lambda\delta\bar{x})$ is sufficiently reduced. What constitutes a sufficient reduction is set by an adjustable parameter in the practical algorithm (`runstructure.m/linesearch`).

This approach minimizes f by minimizing \bar{F} using Newton steps. Since $\delta\bar{x}$ is in a descent direction one is guaranteed a reduction in f for an λ small enough.

5.2.2.4 Implementation

This section makes direct references to the code to show the implementation of the above to solve Poisson's equation. The following functions relate to the implementation/solution of Poisson's equation.

genrhs.m generates the right-hand side of Poisson's equation (or its derivative w.r.t potential) for a given charge.

genmatrixpoi.m generates the left-hand side (matrix **A**) of Poisson's equation. It remains constant between iterations (independent of potential or charge); changes with new boundary conditions (i.e. applied voltages).

gencharge.m generates the carrier densities (or the change between iterations) for a given potential and eigenpairs.

runstructure.m the equations system is solved here and also the Poisson equation through incorporated function **newton** and **linsearch**.

The polarization effect is added to Poisson's equation in **genrhs.m** using the total polarization values that are added in **buildstructure.m** using polarization values given in **ganmaterial.m**. Since CGS units are used special care should be taken on the difference in units in Poisson's equation since the units for the charge density and polarization derivative differ by a factor i.e. $\vec{D} = 4\pi\rho + \nabla\vec{P}$. This should be confirmed along with the units of the polarization values. The formulas calculate the polarization in units of C.m^{-2} which should be converted to \AA^{-2} to match the units of the carrier density. This is probably still not correct in the code.

The inverted behaviour of the simulator when doping values are changed (*cf.* Chapter 6) may be caused by an error of sign somewhere in **genmatrixpoi.m**.

The Newton method is implemented within sub-functions of **runstructure.m**. The Jacobian matrix is calculated easily from the matrix expression $\mathbf{A}\phi = \mathbf{b}$ s.t.

$$\mathbf{J} = \mathbf{A} - \frac{\partial}{\partial\phi}\mathbf{b} \quad (5.28)$$

where $\frac{\partial}{\partial\phi}\mathbf{b}$ expresses the derivative of the charge density ρ which simply is subtracted from the main diagonal of **A**.

5.2.3 Schrodinger's equation

5.2.3.1 Discretization by finite differences

The Schrodinger equation is discretized using a finite difference scheme. The Hamiltonian, in the effective mass approximation, yields a sparse tridiagonal matrix, when applying the finite difference discretization. The main diagonal and the first sub-diagonals are given as

$$d_i = \frac{2}{\Delta x_i + \Delta x_{i+1}} \left(\frac{1}{m_i \Delta x_i} + \frac{1}{m_{i+1} \Delta x_{i+1}} \right) \quad (5.29)$$

$$sd_i = -\frac{2}{m_i \Delta x_i (\Delta x_{i-1} + \Delta x_i) (\Delta x_i + \Delta x_{i+1})} \quad (5.30)$$

yielding the Hamiltonian matrix

$$\hat{\mathcal{H}} = -\frac{\hbar^2}{2m_0} \begin{bmatrix} d_1 & sd_2 & 0 & \cdots & 0 \\ sd_2 & d_2 & sd_3 & \ddots & \vdots \\ 0 & sd_3 & d_3 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & sd_{N-2} \\ 0 & \cdots & 0 & sd_{N-2} & d_{N-2} \end{bmatrix} +$$

$$-q \begin{bmatrix} \phi_1 - \Delta E_C^1 & 0 & 0 & \cdots & 0 \\ 0 & \phi_2 - \Delta E_C^2 & 0 & \ddots & \vdots \\ 0 & 0 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \phi_{N-2} - \Delta E_C^{N-2} & 0 \\ 0 & \cdots & 0 & 0 & \phi_{N-1} - \Delta E_C^{N-1} \end{bmatrix} \quad (5.31)$$

By using a space-dependent value of the effective mass (in addition to the non-uniform spatial grid) the Hamiltonian becomes non-symmetric and can thus no longer be Hermitian. This would imply that the energy spectrum is no longer discrete. This matter may need further investigation.

5.2.3.2 Boundary conditions

The Schrodinger equation requires boundary conditions to be specified for the wave functions. It is assumed that the wave functions vanish at the boundaries of the quantum regions (i.e. the region in which the Schrodinger equation is solved). This leads to a zero electron density at the boundaries. If the quantum region stretches over the entire structure this boundary condition, although not completely physically realistic, does not have any significant impact on the simulation results or their evaluation. On the other hand, if only a part of the device is specified as a quantum region then a discontinuity in the electron density will occur at the boundaries to the classical region. Possible treatments of this aspect are discussed in section (5.3) but a disappearing wave function at the boundary is used, nonetheless.

The vanishing electron density at the boundaries caused by the boundary condition introduces a large gradient of the electron density ∇n , which may be problematic (in a numerical sense) when this term occurs in the solution of the continuity equation e.g. using the finite difference scheme or the drift-diffusion formulation of current density.

A possible 'work-around' for the vanishing electron density at the edge of the computation domain could be to extend the domain with extra 'artificial' nodes outside the domain and to solve the SP system over this larger region but then neglecting these nodes again for further computation. This is conceptually feasible but may prove challenging to implement.

5.2.3.3 Implicitly restarted Arnoldi method

The Schrodinger equation presents an eigenvalue problem. The eigenpairs (eigenvector, eigenvalue) are computed using the Implicitly Restarted Arnoldi Method (IRAM) of the ARPACK library (facilitated through the MATLAB function `eigs`). The IRAM is a member of the Krylov subspace projection methods, that are used for large scale eigenvalue problems, and is a generalization of the Lanczos method for the non-symmetric case [42].

5.2.3.4 Inverse iteration

If the step change in ϕ is sufficiently small between iterations, the eigenpairs are not recalculated anew but only slightly adapted using inverse iteration. The inverse iteration method is a numerical algorithm used to solve eigenvalue problems for which good approximations of the eigenpairs exist, that will ensure (fast) convergence. In our case these approximations would be the eigenpairs obtained in the previous iteration (for ϕ^k). After solving the Poisson equation anew to obtain ϕ^{k+1} , if $|\phi^{k+1} - \phi^k| < \rho$ (where ρ is some tolerance), the change in the eigenvalues and eigenvectors is assumed to be sufficiently small such that the new eigenpairs of the Schrodinger equation can be calculated by inverse iteration. This is substantially faster to compute than recomputing the eigenvalues

from scratch, especially since good approximations of the eigenpairs are given by the existing eigenpairs of the previous iteration. Adopting the same notation as the Schrodinger equation, the improved approximation of the eigenvector (wave function) is

$$\psi^{k+1} = (\mathcal{H} - \mathcal{E}^k \mathbf{I})^{-1} \psi^k \quad (5.32)$$

The corresponding eigenvalue is updated as

$$\mathcal{E}^{k+1} = \mathcal{E}^k + \rho \quad (5.33)$$

where the correction is given by

$$\rho = \frac{\psi^{k+1} \psi^k}{|\psi^{k+1} \psi^k|} \quad (5.34)$$

5.2.3.5 Implementation

This section makes direct references to the code to show the implementation of the above to solve Schrodinger's equation. Also refer to subsection (5.3.5) for related issues. The following functions relate to the solution of the Schrodinger equation

genmatrix1D generates the Hamiltonian, for a given potential, for the Schrodinger equation.

schr.solve organises the solution of Schrodinger equation passing variables and selecting the appropriate solver.

schr.track1D sets up the numerical problem for solution by inverse iteration.

inviter1 implements the inverse iteration that approximate the eigenpairs for the eigenvalue problem (Schrodinger equation).

schr.solve1D sets up the numerical problem by generating the Hamiltonian (**genmatrix1D.m**) and solves it using IRAM.

Additional comments can be found in the respective files.

The ΔE_C value that appears in the Hamiltonian (5.31) does not appear as a material parameter in the code. The energy reference for the system is the conduction band energy in GaN i.e. $E_C(\text{GaN}) = 0$, therefore $\Delta E_C = E_C(\text{Al}_x\text{Ga}_{1-x}\text{N})$.

The value of change in potential for which inverse iteration is used to solved the Schrodinger, instead of IRAM, must be set in **initaquila.m**. The best value for this has not been investigated.

5.2.4 Continuity equation

The continuity equation expressed in the form of (4.35) is simplified for the one-dimensional case, dropping the mobility term μ_n , as

$$\frac{d}{dx} \left(n \frac{d}{dx} E_F \right) = 0 \quad (5.35)$$

$$n(E_F)_{xx} + n_x(E_F)_x = 0 \quad (5.36)$$

where E_F is understood to be the quasi-Fermi level for electrons E_{Fn} . The value of the electron density n is also dependent on E_F in a nonlinear fashion (*cf.* (4.70), (4.50)). Two possible approaches exist to solve (5.35) to obtain a self-consistent solution in the system of equations:

1. Solve the nonlinear equation with $n(E_F)$ using (say) the Newton-Raphson method to iteratively obtain a solution.
2. Keep the electron density n fixed and solve the linear equation.

The latter approach, while simpler, probably will require more iterations to reach a self-consistent solution of the system. The first approach is expected to result in faster convergence of the system but this has not been verified.

The solution of the SP system along with the continuity equation has not been successful, hence all the attempted solution approaches to solving the continuity equation are listed here. A discretization using finite differences (both forward and central difference) and finite volumes has been attempted, as shown below. It is still unclear whether the reason for the numerical problems (?) are in the discretization of the continuity equation or rather its self-consistent solution in the coupled equation system.

5.2.4.1 Discretization by finite differences

The discretizations using both a forward difference and a central difference scheme for the first-order derivatives are given below, as implemented in `dajmat.m`.

Forward difference The derivative of the electron density is approximated by a forward difference method.

$$\frac{2n_i}{\Delta x_i + \Delta x_{i-1}} \left(\frac{\partial E_F}{\partial x} \Big|_{i+\frac{1}{2}} - \frac{\partial E_F}{\partial x} \Big|_{i-\frac{1}{2}} \right) + \frac{n_{i+1} - n_i}{\Delta x_i} \frac{E_F(x_{i+1}) - E_F(x_i)}{\Delta x_i} = 0 \quad (5.37)$$

$$\begin{aligned} \frac{2n_i}{\Delta x_i + \Delta x_{i-1}} \left(\frac{E_F(x_{i+1}) - E_F(x_i)}{\Delta x_i} - \frac{E_F(x_i) - E_F(x_{i-1})}{\Delta x_{i-1}} \right) + \\ + \frac{n_{i+1} - n_i}{\Delta x_i} \frac{E_F(x_{i+1}) - E_F(x_i)}{\Delta x_i} = 0 \end{aligned} \quad (5.38)$$

Rewritten in a form suitable for matrix implementation

$$\begin{aligned} E_F(x_{i-1}) \left(\frac{2n_i}{\Delta x_{i-1} (\Delta x_i - \Delta x_{i-1})} \right) - E_F(x_i) \left(\frac{2n_i}{\Delta x_i \Delta x_{i-1}} + \frac{n_{i+1} - n_i}{(\Delta x_i)^2} \right) + \\ + E_F(x_{i+1}) \left(\frac{2n_i}{\Delta x_i (\Delta x_i + \Delta x_{i-1})} + \frac{n_{i+1} - n_i}{(\Delta x_i)^2} \right) = 0 \end{aligned} \quad (5.39)$$

Central difference The approximation of the first term of (5.36) remains the same as in (5.38). The derivative of n is also approximated using a central difference scheme, using $n_{i+\frac{1}{2}} = n_{i+1} - n_i$.

$$\begin{aligned} \frac{2n_i}{\Delta x_i + \Delta x_{i-1}} \left(\frac{E_F(x_{i+1}) - E_F(x_i)}{\Delta x_i} - \frac{E_F(x_i) - E_F(x_{i-1})}{\Delta x_{i-1}} \right) + \\ + \frac{(E_F(x_{i+1}) - E_F(x_i))(n_{i+1} - n_{i-1})}{(\Delta x_i + \Delta x_{i-1})^2} = 0 \end{aligned} \quad (5.40)$$

Rewritten in a form suitable for matrix implementation

$$\begin{aligned} \frac{E_F(x_{i-1})}{\Delta x_i + \Delta x_{i-1}} \left(\frac{2n_i}{\Delta x_{i-1}} - \frac{n_{i+1} - n_{i-1}}{\Delta x_i + \Delta x_{i-1}} \right) - E_F(x_i) \left(\frac{2n_i}{\Delta x_i \Delta x_{i-1}} \right) + \\ + \frac{E_F(x_{i+1})}{\Delta x_i + \Delta x_{i-1}} \left(\frac{2n_i}{\Delta x_{i-1}} + \frac{n_{i+1} - n_{i-1}}{\Delta x_i + \Delta x_{i-1}} \right) = 0 \end{aligned} \quad (5.41)$$

5.2.4.2 Discretization by finite volumes

The finite volumes discretization of (5.35) avoids the evaluation of second order derivatives. This discretization scheme is implemented in `dajmat2.m`.

$$\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \nabla \cdot (n \nabla E_F) dx = 0 \quad (5.42)$$

$$n \frac{\partial E_F}{\partial x} \Big|_{x_{i+\frac{1}{2}}} - n \frac{\partial E_F}{\partial x} \Big|_{x_{i-\frac{1}{2}}} = 0 \quad (5.43)$$

$$n_{i+\frac{1}{2}} \frac{E_F(x_{i+1}) - E_F(x_i)}{\Delta x_i} - n_{i-\frac{1}{2}} \frac{E_F(x_i) - E_F(x_{i-1})}{\Delta x_{i-1}} = 0 \quad (5.44)$$

the midpoint electron density is taken as a nodal average $n_{i+\frac{1}{2}} = \frac{n_{i+1} + n_i}{2}$ such that

$$\frac{n_{i+1} - n_i}{2} \frac{E_F(x_{i+1}) - E_F(x_i)}{\Delta x_i} - \frac{n_i - n_{i-1}}{2} \frac{E_F(x_i) - E_F(x_{i-1})}{\Delta x_{i-1}} = 0 \quad (5.45)$$

This may be rewritten in a form suitable for matrix representation

$$\frac{n_i - n_{i-1}}{2\Delta x_{i-1}} E_F(x_{i-1}) - \left(\frac{n_{i+1} - n_i}{2\Delta x_i} + \frac{n_i - n_{i-1}}{2\Delta x_{i-1}} \right) E_F(x_i) + \frac{n_{i+1} - n_i}{2\Delta x_i} E_F(x_{i+1}) = 0 \quad (5.46)$$

Using a fixed value of n , the value of E_F can be calculated by solving the linear system. If the nonlinear dependence $n(E_F)$ is considered the use of the Newton-Raphson was attempted to solve the equation. All efforts were unsuccessful and both the approach and implementation (described in the following section) should be checked.

5.2.4.3 Newton-Raphson method

The residual vector of the continuity equation takes the same form as the (numerical implementation) of equation (5.35) since there are no recombination terms on the right-hand side.

The calculation of the Jacobian depends on the chosen numerical scheme. The derivation using the finite volumes discretization (5.46) will be shown here. The discretization yields a system of equations in the form

$$\mathbf{A} \mathbf{e} = \mathbf{b} \quad (5.47)$$

where \mathbf{e} here represents the vector of quasi-Fermi values E_{Fn} . The vector $\mathbf{b} = \mathbf{0}$ except for the boundary node values of the Fermi level.

Using the chain rule, the Jacobian will take the form

$$\mathbf{A} \frac{\partial \mathbf{A}}{\partial \mathbf{e}} = \mathbf{0} \quad (5.48)$$

It is not clear whether this is indeed valid since the matrix \mathbf{A} was derived assuming n to be fixed and the equation to be linear. Now the dependence of \mathbf{A} on E_F is introduced by $n(E_F)$. We continue rewriting (5.46), with $\theta_i \equiv \frac{n_i - n_{i-1}}{2\Delta x_{i-1}}$, as

$$\theta_i E_F(x_{i-1}) - (\theta_{i+1} + \theta_i) E_F(x_i) + \theta_{i+1} E_F(x_{i+1}) = 0 \quad (5.49)$$

This allows \mathbf{A} to be written i.t.o. θ_i like

$$\mathbf{A} = \begin{bmatrix} \star & 0 & & & 0 \\ \theta_2 & \theta_2 + \theta_3 & \theta_3 & 0 & \\ 0 & \theta_3 & \theta_3 + \theta_4 & \theta_4 & 0 \\ & & \ddots & \ddots & \ddots \\ & & & 0 & \star \end{bmatrix} \quad (5.50)$$

The required derivatives for the matrix $\frac{\partial \mathbf{A}}{\partial \mathbf{e}}$ are

$$\frac{\partial \mathbf{A}_{i,i}}{\partial \mathbf{e}_i} = \frac{\partial}{\partial \mathbf{e}_i} (\theta_i + \theta_{i+1}) \quad (5.51)$$

$$\frac{\partial \mathbf{A}_{i,i+1}}{\partial \mathbf{e}_i} = \frac{\partial}{\partial \mathbf{e}_i} (\theta_{i+1}) \quad (5.52)$$

$$\frac{\partial \mathbf{A}_{i-1,i}}{\partial \mathbf{e}_i} = \frac{\partial}{\partial \mathbf{e}_i} (\theta_i) \quad (5.53)$$

Calculating the composing derivatives as

$$\frac{\partial}{\partial \mathbf{e}_i} (\theta_{i+1}) = \left(\frac{\partial n_{i+1}}{\partial \mathbf{e}_i} - \frac{\partial n_i}{\partial \mathbf{e}_i} \right) \frac{1}{\Delta x_i} \quad (5.54)$$

$$= -\frac{\partial n_i}{\partial \mathbf{e}_i} \frac{1}{\Delta x_i} \quad (5.55)$$

$$\frac{\partial}{\partial \mathbf{e}_i} (\theta_i) = \frac{\partial n_i}{\partial \mathbf{e}_i} - \frac{\partial n_{i-1}}{\partial \mathbf{e}_i} \quad (5.56)$$

$$= \frac{\partial n_i}{\partial \mathbf{e}_i} \frac{1}{\Delta x_{i-1}} \quad (5.57)$$

The value of $\frac{\partial n_i}{\partial \mathbf{e}_i}$ denotes the derivative at node i , with a summation over all subbands, as will become apparent from (4.70) rewritten here with the adopted notation

$$n_q(\phi) = \frac{1}{\pi} \left(\frac{2m^* k_B T}{\hbar^2} \right)^{1/2} \sum_n \psi_n^2 \mathcal{F}_{-1/2} \left(\frac{e_i - E_n}{k_B T} \right) \quad (5.58)$$

$$= N^{2D} \sum_n \psi_n^2 \mathcal{F}_{-1/2} \left(\frac{e_i - E_n}{k_B T} \right) \quad (5.59)$$

The derivative is calculated as

$$\frac{\partial n_i}{\partial \mathbf{e}_i} = \frac{N^{2D}}{k_B T} \sum_n \psi_n^2 \mathcal{F}_{-3/2} \left(\frac{e_i - E_n}{k_B T} \right) \quad (5.60)$$

When utilizing the function `genqcharge.m` the vector \mathbf{e} is passed returning a vector of derivatives $\frac{\partial n_i}{\partial \mathbf{e}_i}$. The constrained electron in the potential well do not contribute to the gate current. Thus, when calculating the derivatives the bottom three or four subbands must be neglected in the summation. Since these subbands provide a significant contribution to the electron density it is important to take the fact into account also in the calculation of matrix \mathbf{A} .

To complete the derivation we write

$$\frac{\partial \mathbf{A}_{i,i}}{\partial \mathbf{e}_i} = \frac{\partial n_i}{\partial \mathbf{e}_i} \left(\frac{1}{\Delta x_{i-1}} - \frac{1}{\Delta x_i} \right) \quad (5.61)$$

$$\frac{\partial \mathbf{A}_{i,i+1}}{\partial e_i} = -\frac{\partial n_i}{\partial e_i} \frac{1}{\Delta x_i} \quad (5.62)$$

$$\frac{\partial \mathbf{A}_{i-1,i}}{\partial e_i} = \frac{\partial n_i}{\partial e_i} \frac{1}{\Delta x_{i-1}} \quad (5.63)$$

If a uniform grid is used where $\Delta x_{i-1} = \Delta x_i$ the terms on the main diagonal are zero. This matrix was singular when implemented in MATLAB. Consequently, the Newton-Raphson method could not be implemented successfully.

The Jacobian matrix requires many numerical evaluations; the secant method may be better suited (refer to [41] for the implementation).

5.2.4.4 Boundary conditions

Dirichlet boundary conditions are enforced on both contact for the quasi-Fermi levels such that the difference equals the applied potential

$$E_{Fn}(0) - E_{Fn}(L) = qV_G \quad (5.64)$$

5.2.4.5 Implementation

The following MATLAB functions relate to the (attempted) implementation of the numerical solution of the continuity equation.

runstructure the implementation of the Newton iteration loop happens here (along with the self-consistent solution of the other equations).

dajmat discretization of the continuity equation, for a fixed n , using the finite difference scheme.

dajmat2 discretization of the continuity equation, for a fixed n , using the finite volumes scheme.

genrhsJ generates the right-hand side of the continuity equation. If the recombination term is neglected it is simply a zero vector with the boundary values replacing the first and last elements in the vector.

cont_Jacobian calculates the discussed Jacobian matrix using derivatives that are calculated by `cont_derivative.m`.

cont_current attempts to calculate the Fermi level by first calculating the current density (with drift-diffusion equation) and then setting it equal to $J = \mu_n n \nabla E_F$.

The fact that a zero electron density results on the boundary nodes, because of the boundary conditions used in Schrodinger's equation, causes problems when calculating the electron density gradient ∇n near the boundaries. Large gradients result that may hinder a numerical convergence when solving the continuity equation.

5.3 Code implementation

The code for this simulator was adapted from the AQUILA² code - an open source Schrodinger-Poisson solver for GaAs/AlGaAs structures - written in MATLAB. Various additions, corrections and modifications have been made of which many are not explicitly documented here. It is therefore recommended to use the latest version of the new simulator - GaNQUILA - as a starting point/reference for future development work.

²available from <http://www.mathworks.com/matlabcentral/fileexchange/3344>

5.3.1 Additions and modifications to AQUILA code

The most significant modifications that were made to the AQUILA code are discussed below. The affected/new MATLAB functions are indicated and should be referenced for further comments.

5.3.1.1 Material parameters

The GaAs/AlGaAs material database was changed to GaN/AlGaN values (where available). The values for polarization-related constants were added; all material parameters have been retained (only values were changed). This allows the GaAs and GaN databases to be interchanged quite easily. The GaAs database would require the addition of dummy polarization values and the call to the function `ganmaterial.m` should be changed to `gaasmaterial` in all the affected functions. A consolidation of both material databases into a single function with selection by a flag as input argument should be considered if a frequent change between materials is anticipated.

Some parameters/effects present for the GaAs material in the database have not been adopted for the GaN material database:

- The direct-indirect band gap crossover point depending Al-content could not be found for GaN. It is not evident if this effect exists for hexagonal (wurtzite) GaN.
- The non-parabolicity correction for Gamma electrons when calculating the Fermi integrals in `gencharge.m` was also not found for GaN - it has been removed.
- The acceptor and donor ionization energies are kept at the same levels as for GaAs. Currently all dopants are assumed to be fully ionized (which is reasonable at 300 K) so it does not matter. The consideration of surface states may require closer attention to proper ionization energies.

5.3.1.2 Polarization effects

The polarization effects evident in wurtzite GaN have been added to the Poisson's equation. This was done using the finite volumes scheme to avoid a discontinuity at the heterointerface. The additional terms (spatial derivative of polarization $\nabla \vec{P}$) were added in `genrhspoi.m`. Since the polarization value is assumed to be independent of the potential (electric field), the derivative w.r.t. potential (i.e. the Jacobian), used when solving Poisson's equation using Newton's methods, remains unchanged. The derivative of the RHS of Poisson's equation is called using an appropriate flag in `genrhspoi.m`.

5.3.1.3 Continuity equation

The framework for a self-consistent solution of the SP system with the continuity equation has been added to simulate heterostructure in non-equilibrium condition. This included the numerical scheme to discretize the equation (`dajmat.m`, `dajmat2.m`) and the integration into equation system to obtain a solution self-consistent with the Schrodinger and Poisson equations. As noted before, this implementation has not been successful and requires re-investigation.

5.3.1.4 Simplifications for 1D problem

The original AQUILA code was written to also work for 2-dimensional problems. The GaNAQUILA code was developed with only 1-dimensional problems in mind; the functionality in 2-D has not been tested in any way. Some segments of code have been reduced to

only represent the 1-D case to improve readability of the code but some artifacts of the 2-D implementation may still remain. e.g. having indices for both x and y directions. Should a 2-D version be desired (again) the original AQUILA code should be reviewed to identify the necessary additions.

5.3.1.5 Plotting functions

Plotting functions have been extended with `plot_this.m`, `someplot.m` and `bandeng_compare.m`. These functions generate various plots but must be modified to meet needs e.g. axis limits.

5.3.2 Assumptions and simplifications

There are several simplifying assumptions that have been made in the simulator, as detailed below.

5.3.2.1 Polarization-induced sheet charge

Apart from at the AlGa_N/Ga_N heterointerface, there are also induced sheet carrier charges at the top (gate contact) and bottom (Ga_N/nucleation layer interface) interfaces due to the discontinuity in the polarization values that occur there. These sheet charges are neglected.

If the Ga_N bulk is thick enough (> 100 nm - as is usually the case) the sheet charge that forms at the interface between the Ga_N layer and the nucleation layer gets shielded by free carriers in the Ga_N bulk. Similarly, the induced charge at the top interface is shielded by charge in the metal gate contact.

These polarization-induced sheet charges are expected to be positive and thus would also require the consideration of holes in the simulator.

5.3.2.2 Surface states

Surface states are not considered but they may be of importance as they act as a 'source' of electrons (similar to dopants). Thus, the doping of the structure layers is not the only source of electrons contributing to the 2DEG. The respective importance of each is not clear (*cf.* [32]).

5.3.2.3 Carrier types considered

The HEMT is considered a unipolar device i.e. only electrons are considered. Furthermore only the Γ electrons are considered, as the nearest valley in Ga_N lies more than 2 eV higher which makes any contribution by electrons in these valleys to the total electron distribution insignificant [43].

5.3.3 Units and scaling of equations

The semiconductor equations, as employed in the code, have not been scaled. No numerical overflow issues have become apparent in MATLAB because of this but a scaling may be sensible to improve the algorithms efficiency.

The CGS unit system is used for the equations/constants in the code. Dimensions are expressed in Ångstroms and energies and potentials in electronvolts. One should remember to convert the values to the appropriate units for output e.g. the calculated densities returned from `gencharge.m` are expressed in Å^{-3} and should be multiplied with a factor of 10^{24} to be expressed in the commonly used density unit for semiconductors of cm^{-3} .

5.3.4 Initial values of variables

All variables are initialised to zero or their equilibrium values, as applicable.

An initial guess for the potential ϕ is needed. By default $\phi_i = 0$ is used but by providing a better guess the system reaches convergence in less iterations. For this purpose the previous simulation result is normally used as it should provide a good guess for the next time iteration with small changes in the applied gate voltage. An initial value for the potential can be specified using `startpotential.m` but it has not been found to show effect.

The Arnoldi (IRAM) method (through the `eigs.m` function) used to solve the Schrodinger equation allows a guess to be given for the wave functions when the iteration process is started; a sinusoid is used as an initial guess for the wave functions.

The Fermi level is assumed to be in the middle of the energy bandgap (as is the case in equilibrium).

5.3.5 Consideration of the quantum region

This subsection discusses various aspects that should be considered when/how a 'quantum region' is specified where the Schrodinger is solved.

5.3.5.1 Extent of the quantum region

Solving the SP system has a high computational cost (compared to only solving Poisson's equation). Therefore, the Schrodinger-Poisson system is only solved in regions where quantum behaviour is expected - within the region of the quantum well formed at the heterointerface - and only the Poisson equation is used in the remaining parts of the structure. This approach saves much computational time and has been often implemented [44, 45]. This approach, however, requires special consideration of the coupling between the quantum and classical regions (see subsection 5.3.5.2).

The best convergence behaviour, however, has been found when considering a quantum region covering the entire device structure with a sufficient number of subbands to account for all the significant electron contributions. This avoids the coupling problem which is thought to hinder numerical convergence. This topic is discussed, nonetheless, as it might be of interest.

5.3.5.2 Coupling of quantum and classical regions

The original AQUILA code makes the assumption that all charge within the quantum region is contributed only by electrons in the quantized energy bands and that all charge outside the quantum region is purely classical. However, not all electrons in the (user-defined) quantum region are necessarily sufficiently confined by the potential well and there may be electrons (above a certain energy level) that behave classically. The original AQUILA does not account for this. In its implementation the Poisson equation is solved over the entire domain to obtain ϕ and n . The value of n is then replaced in the quantum region by those calculated from solving the SP system. The continuity of the electron density is thereby lost at the boundaries between the classical and quantum regions since the Schrodinger equation is solved with the boundary condition of vanishing wave functions (implying zero electron density) at the edge of the quantum region. To achieve continuity the electron density obtained from the Poisson equation, should serve as a boundary condition when solving the Schrodinger equation. This boundary condition is not implementable when considering more than one subband (which is the always the case for a practical device structure). There are several approaches to solving this problem, as will be discussed below.

5.3.5.3 Classical correction to quantum regions

Another approach has been adopted by [35, 40] where the charge in the quantum region is not considered to be solely of a quantum nature. Only the electrons within the potential well show two-dimensional (quantized) behaviour, while the electrons with an energy above E_{lim} (refer to figure 5.5) are considered in a classical sense and assume three-dimensional behaviour again since their energy levels are not quantized but form a continuum of values. It has been suggested that all subbands should be considered quantized up to the point where their eigenenergies differ by less than kT from the next value (just to set a reasonable transition point between quantized and unconstrained electron behaviour).

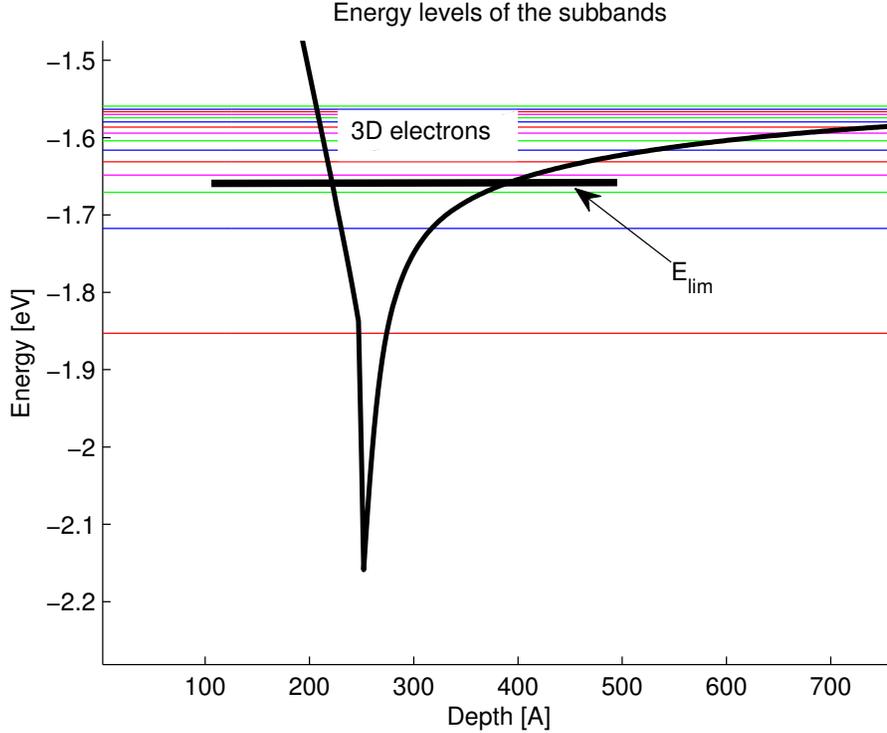


Figure 5.5: Separation of 2D and 3D behaviour of electrons in quantum well. Electrons above level E_{lim} are considered unconstrained (3D); below they are considered quantized (2D).

The classic 3D electron density, as calculated in (4.50), calculates the integrals from the conduction band energy E_C . For the case of the classical electron density contribution in the quantum region this integral should be calculated from $\max(E_{lim}, E_C)$ instead, making the subsequent use of the Fermi integrals no longer valid. There are methods suggested in [44] to make corrections to the integrals.

The limits of the integral in (4.55) must also be changed to avoid a contribution by both the classical and the quantum electron density integrals for energies above E_{lim} . The upper bound of the integral needs to be replaced $\infty \rightarrow E_{lim}$. This is the same as writing

$$\int_E^{E_{lim}} dx = \int_E^{\infty} dx - \int_{E_{lim}}^{\infty} dx \quad (5.65)$$

The Schrodinger equation is solved for (ψ_n, E_n) over all the subbands considered and the quantum charge density is then calculated for the quantum well region.

The use of a 'classical correction' in the quantum regions allows much less subbands to be used and also a smooth coupling between quantum and classical regions. This has, however, been found to introduce numerical instabilities in the solution of the SP system. This is thought to be because the intersection of the energy values, where $E_C = E_{lim}$, does not necessarily coincide with a node point causing a non-smooth jump in density to the next node.

5.3.5.4 The appropriate number of subbands

The question arises how many subbands should be considered to capture all the significant contributions to the quantum electron density in the quantum region. This depends on the specific device structure considered and cannot be predicted reliably *a-priori*. The best approach is to simulate the device and plot the wave functions along with their corresponding contributions to the total quantum electron density and then increase or decrease the number of subbands accordingly. In all cases enough subbands should be computed such that the highest eigenenergy (above the conduction band) is always above the Fermi level. By the nature of the Fermi-Dirac statistics, the probability of subbands above the Fermi level being occupied decreases exponentially and thus quickly becomes negligible. For the heterostructures considered here, three subbands have been found to capture most of the quantum effects around the heterointerface as can be seen from results (Chapter 6).

5.3.5.5 Boundary condition

The assumption of a vanishing wave function ($\psi = 0$) on the boundary of the quantum region is only physically valid if the energy of the highest subband lies below E_C at the quantum boundary (since no states exist in the energy bandgap). If the entire device is covered by a quantum region without any 'classical correction' this can never hold true for all subband energies. In the case where the quantum region only surrounds the heterointerface and a 'classical correction' is used, the value of E_{lim} should be chosen such that it lies below E_C at the boundary of the quantum region (NOTE: remember to calculate the 'classical correction' integrals from $\max(E_C, E_{lim})$).

5.3.6 Fixes for nonequilibrium calculation

A work-around for implementing applied gate voltages without being able to solve the continuity equation is to manually adjust the Fermi level in the structure in certain regions by using the `addbias.m` command in GaNQUILA. This is of course extremely rudimentary. One could increase the Fermi level, by an amount equal to the applied gate voltage, in the region from the gate contact up to some distance (say 150 Å) after the heterointerface. The actual Fermi level will decrease more gradually so this approximation will probably overestimate the effect of the applied voltage further from the surface of the device.

The reason for non-convergence when solving the continuity equation has not been pinpointed. Firstly the numerical solution of the continuity equation in a simple test structure should be established. After this has been established it can be integrated with the Poisson and Schrodinger equations. The iterative solution, along with the SP system, may require some numerical damping and can become computationally expensive.

It is commonly recommended to increase the gate voltage in steps smaller than 25 mV to retain numerical stability. This requires a great number of iterations when a large gate voltage is applied. Additionally the self-consistent solution of the continuity equation with a damped iterative technique quickly drives up the computational cost. Once the simulator shows satisfactory results one must save the output at various voltages to be used in a lookup table for use in the quasi-2D simulator.

5.4 Alternative simulators

An overview is given of freely available simulation software that may be used for comparison/validation of the results of the developed simulator.

5.4.1 BandEng

The BandEng software is a freely available³ software for simulating the carrier distributions in 1-dimensional heterostructures, taking polarization effects into account. The software was developed by M. Grundmann as part of his PhD thesis at the University of California Santa Barbara (UCSB) in the research group of Prof. Mishra investigating GaN HEMT structures.

The simulator is often used for validation of results but lacks complete documentation. It is not clear what methods/equations were used in the implementation of the simulator. The only possible source of this information may be the dissertation of Grundmann but access to it must be requested; the author could not be contacted directly.

5.4.1.1 Using BandEng for validation

The BandEng simulator has been used to generate validation results for the GaNQUILA simulator. There are several important things to note to make the comparison of results between the simulators possible

The material database of BandEng can be edited for most material parameters and has been adapted to reflect the material parameters used in GaNQUILA. The piezo-electric polarization is calculated in GaNQUILA using (4.17) but BandEng probably uses (4.18) since the parameters $C_{13,33}$ appear in the material database. In principle there should be no difference in the result between the two formulations.

The temperature dependence of the energy bandgap (subsection 4.1.1) is not taken into account in BandEng, therefore this model must be deactivated in `ganmaterial.m` to match the bandgap values of BandEng.

The energy reference level in GaNQUILA is the conduction band energy $E_C(\text{GaN})$ whereas in BandEng the Fermi level serves as reference (as is common). The values must be readjusted to match (for plotting; `bandeng_compare.m`).

The contact potentials must be specified in BandEng to match the ones in GaNQUILA i.e. a zero Neumann condition on the GaN bulk and a Schottky barrier (of appropriate height for the Al-content) at the AlGaN boundary.

The dopants in BandEng can be set to be fully ionized (under Setup menu); in GaNQUILA this is set in `initaquila.m`. If partially ionized dopants are considered in the future the ionization energies in material database must be homogenised.

The results of BandEng can be exported to a text file that is read and plotted in `bandeng_plot.m`. For a comparative plot with GaNQUILA results use `bandeng_compare.m`.

5.4.2 nextnano

The nextnano software is available as a free download after registration on the website⁴. The website provides extensive tutorials on the use of the software and general topics of computational microelectronics. The nextnano simulator is very extensive. It has not been tested for the simulation of the heterostructures considered.

³available from my.ece.ucsb.edu/mgrundmann/bandeng.htm

⁴www.nextnano.org

5.4.3 1D Poisson/Schrodinger

This simulator was developed by Greg Snider at the University of Notre Dame. It has a text interface and requires an external plotting program. Polarization in nitrides is taken into account but the details of the implementation are not clear. The user manual provides some reference papers detailing the equations of the simulator. This simulator has not been used for validation.

The simulator also assumes thermal equilibrium and applied gate voltages are implemented only by specifying a Schottky barrier of 'appropriate' height. This assumption, as is the case for our simulator, is only valid for negligible current.

Chapter 6

Results

This chapter discusses several representative examples of the simulation results obtained by the developed simulator, GaNQUILA. Firstly, the qualitative behaviour of the simulator is investigated with respect to the change of various device and numerical parameters. Using the BandEng simulator as a comparison, the GaNQUILA simulation results are validated for a selection of test structures.

Test structures

Except if mentioned otherwise, all the following simulation results are for a $\text{Al}_x\text{Ga}_{1-x}\text{N}/\text{GaN}$ heterostructure, where the Al-content is denoted by x . A Schottky barrier is assumed at the left gate contact; no applied voltage bias. A zero derivative of potential is imposed on the right 'contact' boundary.

A compact notation to denote the various test structures and their parameters is adopted here. The GaN bulk layer is assumed to have a constant background doping of $N_D = 10^{15} \text{ cm}^{-3}$ and a thickness of 1000 Å (thick enough to make the Neumann boundary condition plausible). The doping and thickness of the $\text{Al}_x\text{Ga}_{1-x}\text{N}$ layer is denoted by a letter (A-D) and a number (0-3), respectively. The nomenclature is defined in table 6.1; the Al-content is appended at the end e.g. $B2_x0.3$ would describe a 250 Å $\text{Al}_{0.3}\text{Ga}_{0.7}\text{N}$ layer doped with $N_D = 10^{17} \text{ cm}^{-3}$. This nomenclature is also used to name files (cf. `bandeng_compare.m`). The latter structure, $B2_x0.3$, is used as the 'standard' structure to investigate various effects since it represents dimensions and doping commonly found in GaN HEMT heterostructures.

Table 6.1: The nomenclature adopted to describe the doping and dimensions of test structures.

	Doping [cm^{-3}]	Layer thickness [Å]	
A	10^{15}	50	0
B	10^{17}	150	1
C	10^{18}	250	2
D	10^{19}	350	3

Plot classical vs quantum densities

A comparison between the classical solution (solving Poisson's equation) and the quantum solution (solving the SP system) is illustrated in figure 6.1. It can be seen that the consideration of quantum effects, by solving the Schrodinger equation, reduces the peak value of the electron density and 'spreads' the 2DEG. This is because of the spatial probability distribution that the wave functions introduce, as depicted in figure 6.2.

The electron density can be seen to drop to zero at the boundaries when the entire structure is considered a quantum region (with no classical correction), whereas the density is non-zero in the classical case.

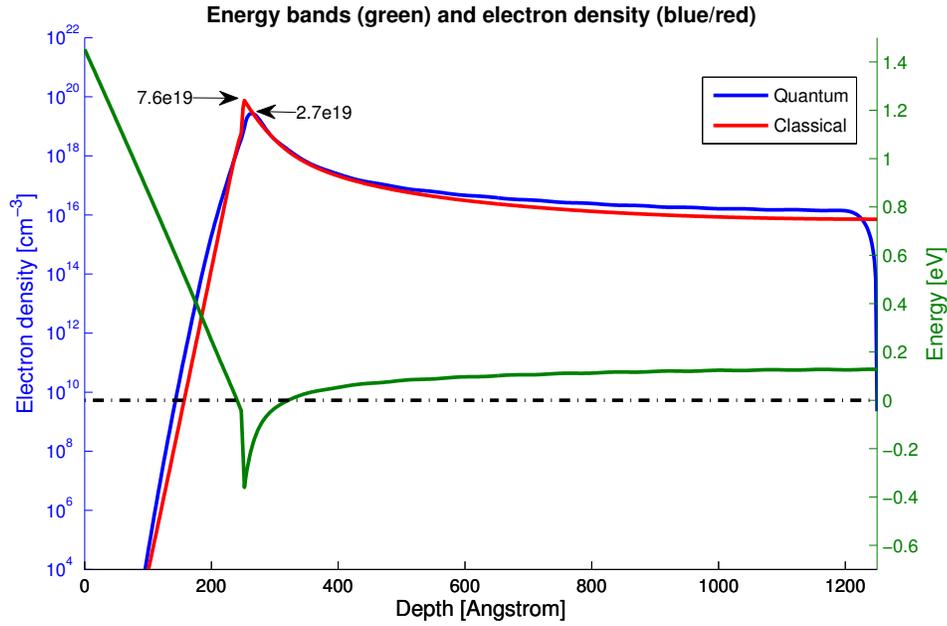


Figure 6.1: Comparison between the classical and quantum electron densities, shown in red and blue, respectively for test structure *B2_x0.3*.

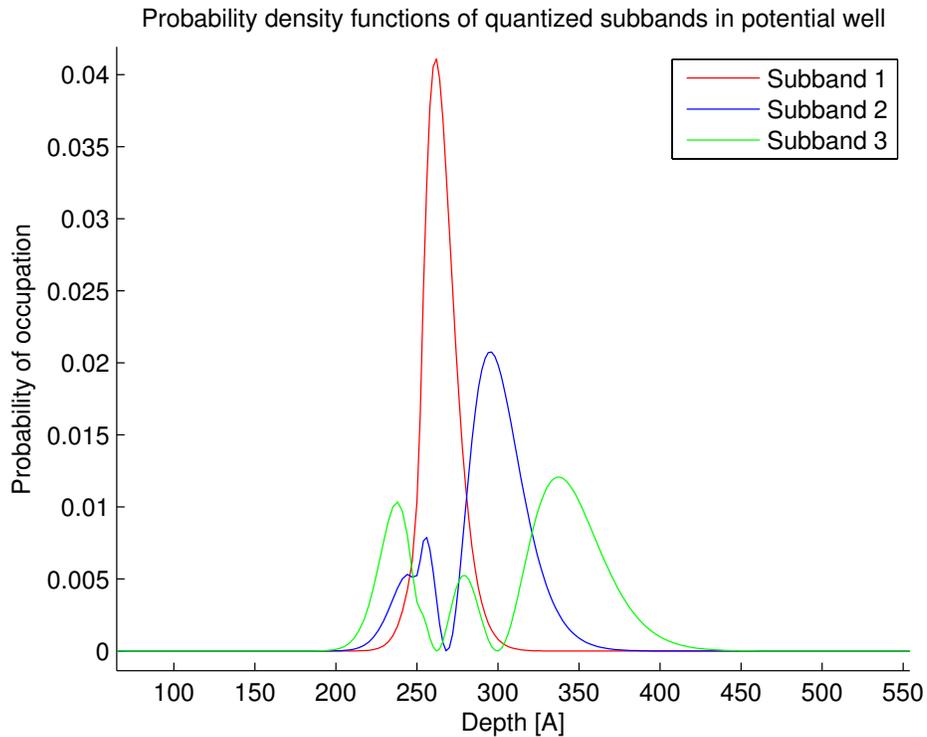


Figure 6.2: The spatial probability distribution $|\psi_i|^2$ of the three lowest (conduction) subbands in the potential well of *B2_x0.3*.

Contributions of subbands to the quantum density

Figure 6.3 shows the contribution of the lowest three subbands to the electron density in the quantum well region. It is evident that these subbands make the significant contribution to the electron density inside the quantum well region. The higher subbands are too far above the conduction band energy to have a significant probability of occupation, due to the Fermi-Dirac statistics.

Figure 6.4 shows the energy values of the subbands which can be seen to converge to a continuum of values i.e. an energy separation less than kT above the 5th subband. This would be the point from where the electrons could be again considered unconstrained (3D behaviour) and the 'classical correction' applied.

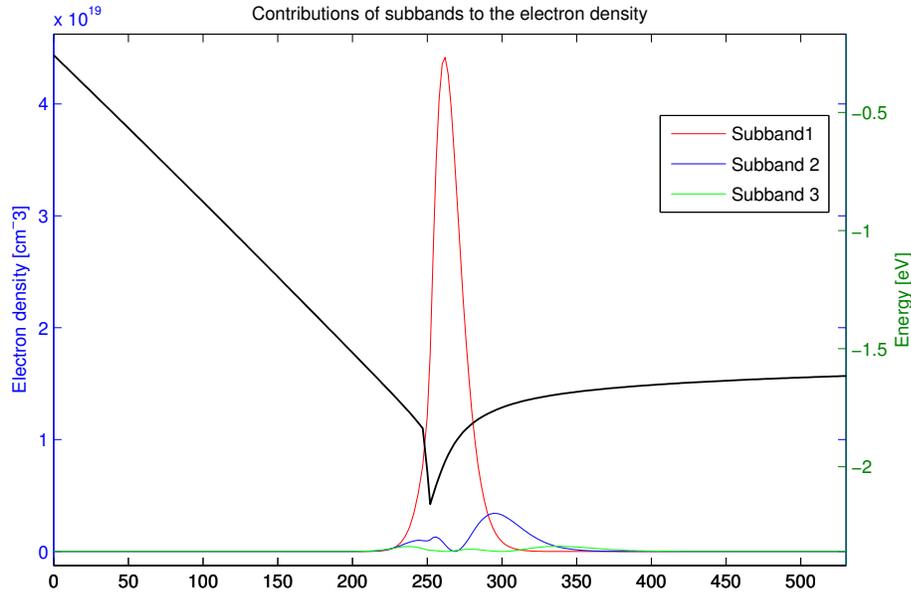


Figure 6.3: The electron density contributed by the three lowest subbands in the quantum well superimposed on a plot of the energy bands.

Effects on computation time of subbands and 'classical correction'

Table 6.2 allows an interpretation of the effects that the number of subbands and the 'classical correction' have on the computation time and charge densities. The results shown in table 6.2 are for the test structure *B2_x0.3* but the same trends should apply to other structures.

When only quantum charge is considered (without a 'classical correction') one has to use many subbands (at least 20) to cover a sufficient range of energy to cover the energy bands' bending to achieve convergence. There is not a significant increase in computation time when adding additional subbands (20 vs. 50) since the wave functions remain mostly unchanged sinusoidal waves that stretch over the length of the device. This is explained by the fact that the eigenvalue problem (Schrodinger's equation) is solved using an iterative technique (IRAM) using the sinusoidal solution as an initial guess. Thus, only few iterations are need from this initial guess to find the correct solution.

When 10 or less subbands are used the 'classical correction' must always be used since the 10 subbands do not cover a sufficient energy range to achieve a convergent result. Because of the numerical destabilisation that occurs if the top subband crosses the conduction band

energy (*cf.* section 5.3.5) the Schrodinger-Poisson system fails to converge to a plausible solution when four to nine subbands are used with a classical correction; if three subbands are used their effect is restricted to the quantum well and convergence is achieved.

The computation parameters do not affect the charge densities substantially since they remain within an order of magnitude of each other in all (converging) cases. There is no change in the results when increasing the number of subbands from 20 to 50. There is negligible change between the results of three and ten subbands (with correction).

The full quantum approach provides the most accurate results (when compared to BandEng - shown later) but the computational advantage of the quantum-classical model may be beneficial if this simulator is eventually coupled to a transport model (see section 4.3) but the underlying reason for the discrepancies must be investigated.

Table 6.2: CPU time, peak electron density and integrated sheet density for structure *B2_x0.3* using different number of subbands and 'classical correction' (where marked with *c*).

Number of subbands	CPU time [s]	n_{\max} [cm^{-3}]	n_s [cm^{-3}]
3 (<i>c</i>)	1.92	2.63×10^{19}	8.78×10^{12}
7 (<i>c</i>)	no	convergence	
10 (<i>c</i>)	6.47	2.76×10^{19}	9.07×10^{12}
10	no	convergence	
20	21.9	4.48×10^{19}	1.13×10^{13}
50	28.6	4.48×10^{19}	1.34×10^{13}

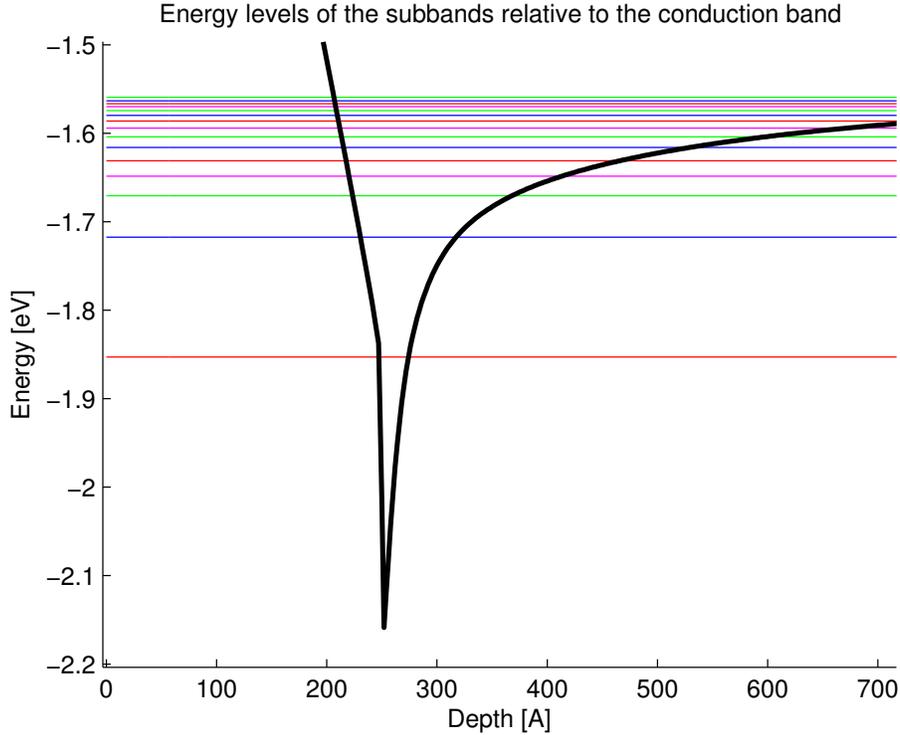


Figure 6.4: Subband energies superimposed on the conduction band. The energies of the subbands are no longer quantized above the 5th subband where they become 'continuous'.

The effect of the width of the quantum region (QBOX)

The placement of the QBOX, signifying the region where quantum effects are considered, requires some experimentation to establish the position that yields the best convergence. As a general guideline, the QBOX should extend from the gate (surface) to at least 150 Å behind the heterointerface. The appropriate placement is also dependent on the number of subbands which have been selected: when no 'classical correction' is used in the quantum region (QBOX) a sufficient number of subbands (~ 20) must be considered. However, the use of a mixed classical-quantum approach and the 'classical correction' in the QBOX has sometimes been found to be problematic for both convergence and results.

Table 6.3 shows a comparison between the computation time of three different heterostructures using either a quantum region that covers the entire device, or a combination between quantum and classical regions. The combination of classical and quantum regions presents roughly a 50% saving in computation time when compared to the full quantum approach; there is no difference in the obtained charge densities (a QBOX with 50 subbands, extending 150 Å behind the heterointerface was used in the 'mixed' case). A further advantage of the classical-quantum approach is that one avoids zero electron densities at right structure extremity, but a discontinuity in the electron density is introduced at the quantum/classical interface when no 'classical correction' is used.

Figures 6.7 and 6.6 show the result (for test structure *B2_x0.3*) of a QBOX extending from the surface to 150 Å behind the heterointerface with three subbands and a 'classical correction' and 50 subbands without a correction, respectively; figure 6.5 shows the result for a QBOX extending over the entire device with 50 subbands. It is evident that the best agreement to the full quantum treatment is achieved when no 'classical correction' is used in the QBOX, however a discontinuity in the electron density at the transition to the classical region results. Although, unphysical it does not affect the peak electron density and integrated sheet electron density appreciably.

To reduce the computation time one has the choice of limiting the quantum region either spatially, or 'energetically' i.e. restricting the number of subbands and using a 'classical correction'. Both approaches pose some problems. The classical correction impacts the accuracy and the use of a spatial classical-quantum separation introduces problems at the transition point between the regions.

Remark1: The lastly mentioned discontinuity does not affect the charge density results appreciably but the discontinuity at the quantum/classical interface will surely cause problems when solving the continuity equation for the quasi-Fermi level if gradients of the electron density (or their numerical equivalents) are considered.

Remark2: It has been found that the use of the 'classical correction' in the QBOX with 50 subbands yields lower electron densities, similar to when only three subbands are considered. This is a clear indication that there is a problem with calculating the charge integrals of the correction terms. The use of the 'classical correction' should therefore be re-investigated.

Table 6.3: Comparison of computation time in seconds between a full quantum region and quantum-classical approach.

Structure	Full quantum	Quantum-Classical
<i>C1_x0.1</i>	21.6	9.5
<i>B2_x0.3</i>	28.6	13.2
<i>A3_x0.2</i>	30.6	13.8

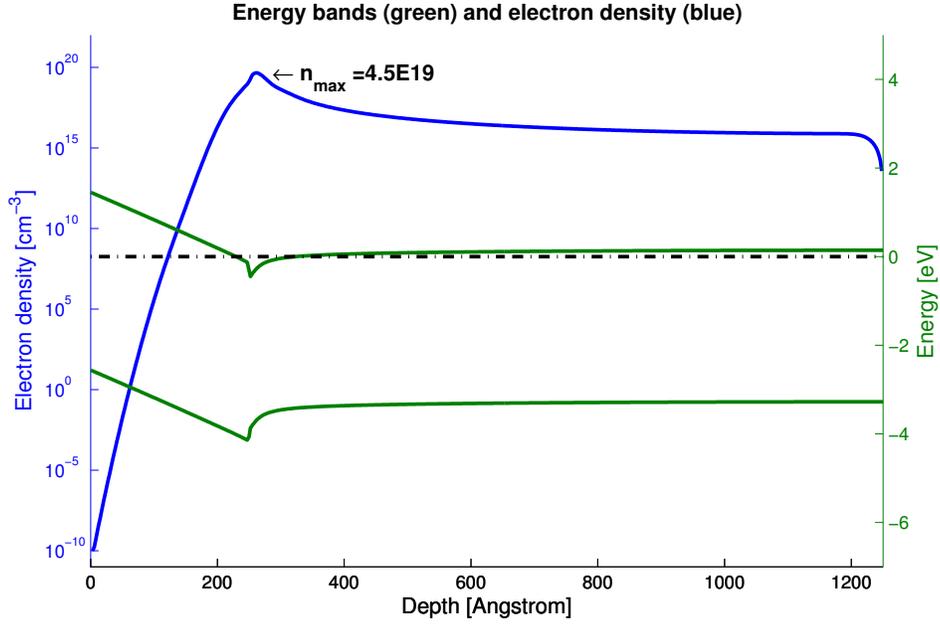


Figure 6.5: A QBOX covering the entire device domain with 50 subbands for structure $B2_x0.3$. A peak electron density of $n_{\max} = 4.7 \times 10^{19} \text{ cm}^{-3}$ and a sheet electron $n_s = 1.4 \times 10^{13} \text{ cm}^{-2}$ are found.

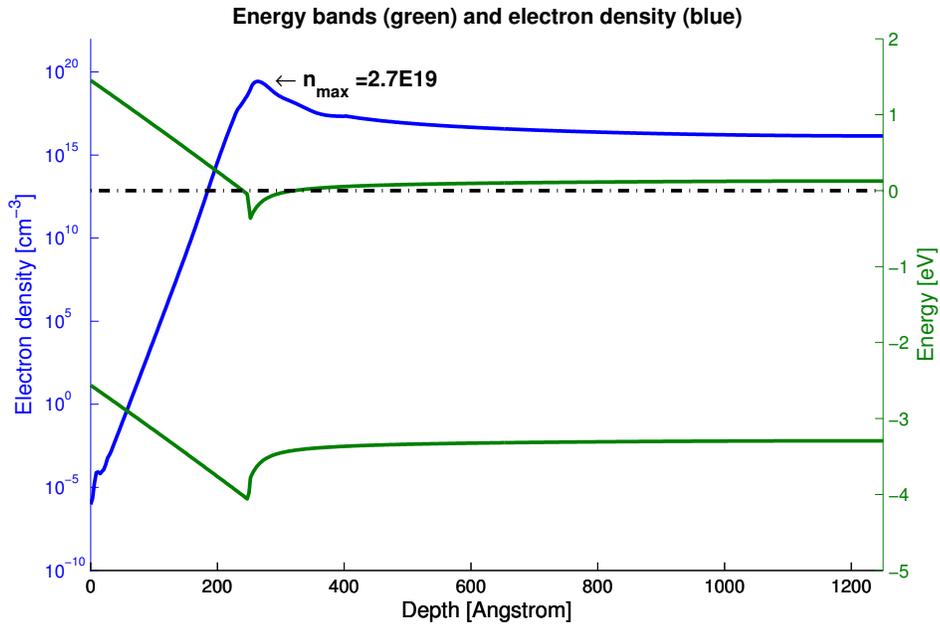


Figure 6.6: A QBOX extending from the surface to 400 \AA , 150 \AA behind the heterointerface, with 3 subbands and a 'classical correction'. A smooth transition to classical region is evident but the charge densities deviate from the full quantum treatment. A peak electron density of $n_{\max} = 2.7 \times 10^{19} \text{ cm}^{-3}$ and a sheet electron $n_s = 8.8 \times 10^{12} \text{ cm}^{-2}$ are found.

Effect with(out) a AlN barrier layer

The effect of the addition of a AlN barrier layer is investigated here by considering a AlGaN/AlN/GaN heterostructure. The polarization due to this interface(s) should also be taken into account by calculating the strain (*cf.* section 4.1.6). Some simulation results with an $\text{Al}_{0.8}\text{Ga}_{0.2}\text{N}$ barrier layer are shown in figure 6.8. The high Al-content causes very strong band bending which is probably not correct. An overlap of the energy levels of the valence and conduction bands occurs at the AlN/GaN interface which may make quantum tunnelling (and recombination) effects no longer negligible. This would require the consideration of holes when solving the equations.

Compared to the same structure without the AlN barrier layer (*cf.* figure 6.5), a stronger spatial confinement of the 2DEG is evident due to the barrier layer. The large energy bandgap of the AlN increases the potential barrier that electrons in the quantum well must transcend and thus decreases their probability of being able to leave the quantum well. Furthermore, the peak electron density is increased due to the larger difference in spontaneous polarization between AlN and GaN (vs. $\text{Al}_x\text{Ga}_{1-x}\text{N}$ and GaN where $x < 1$) and the increased piezoelectric polarization due to the increased strain caused by the increased mismatch in lattice constants.

Effect of the thickness of the AlGaN supply layer

If the AlGaN layer is too thin, for sufficient bending of the energy bands to occur, no 2DEG forms at the heterointerface. What constitutes 'too thin' also depends on the Al-content as can be seen when comparing figure 6.9 and figure 6.10. A thicker AlGaN layer allows more band-bending to occur - also for low Al-content layers - as is evident from figure 6.11. Notwithstanding, an increase of the AlGaN layer thickness reduces the control that the gate potential has on the 2DEG since there are unionized donor dopants

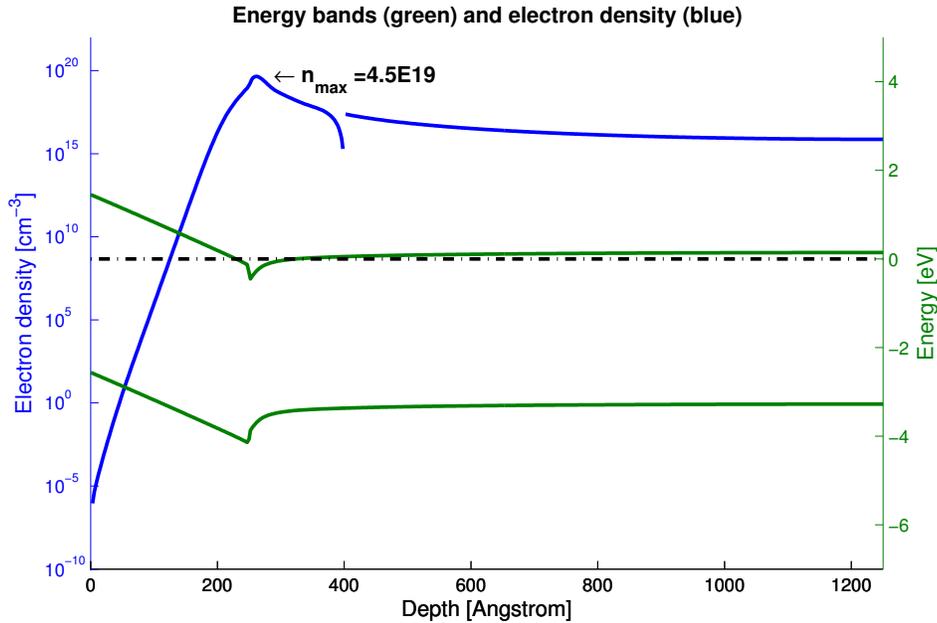


Figure 6.7: A QBOX extending from the surface to 400 Å, 150 Å behind the heterointerface, with 50 subbands and no 'classical correction'. A discontinuity in the transition to the classical region is evident but the charge densities match the ones obtained by full quantum treatment very well. A peak electron density of $n_{\text{max}} = 4.5 \times 10^{19} \text{ cm}^{-3}$ and a sheet electron $n_s = 1.3 \times 10^{13} \text{ cm}^{-2}$ are found.

in the middle of the AlGa_{0.7}N layer that provide shielding. The latter effect cannot be made clear by the simulator currently, since only the thermal equilibrium case can be simulated.

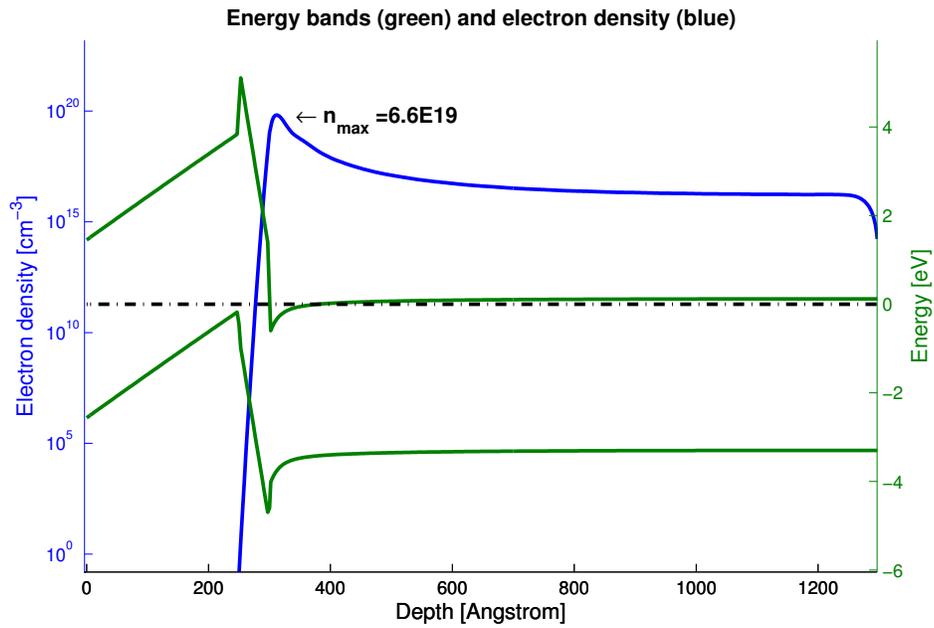


Figure 6.8: A Al_{0.3}Ga_{0.7}N/Al_{0.8}Ga_{0.2}N/GaN heterostructure showing increased peak electron density and spatial isolation of the 2DEG.

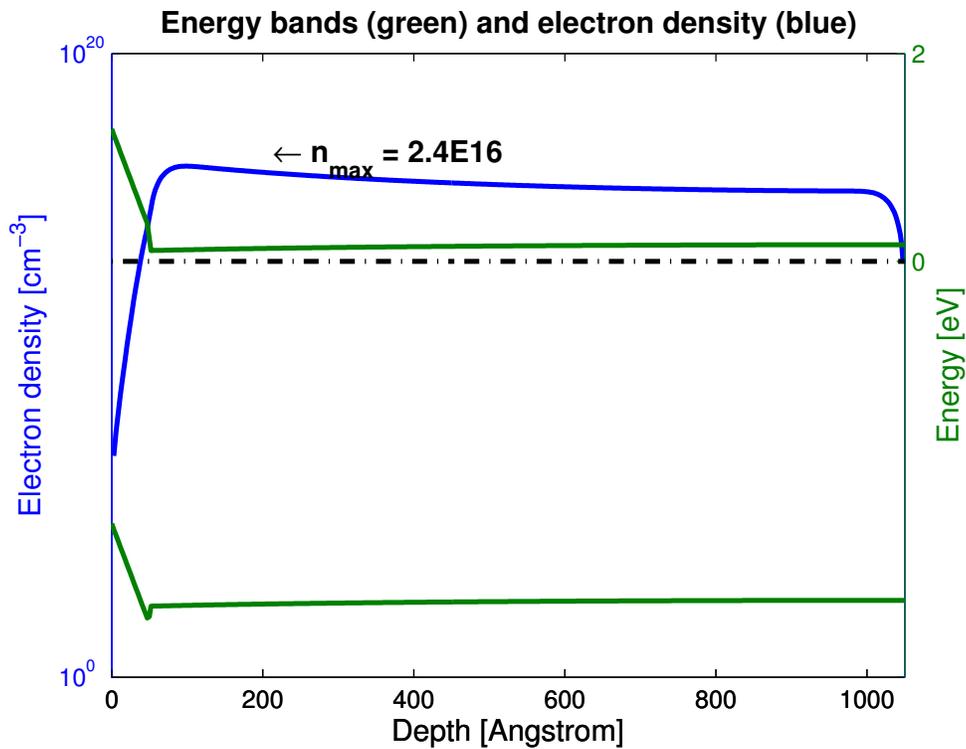


Figure 6.9: Test structure *A0_x0.2* (a 50 Å Al_{0.2}Ga_{0.8}N layer) showing insufficient bending of energy bands to form a 2DEG.

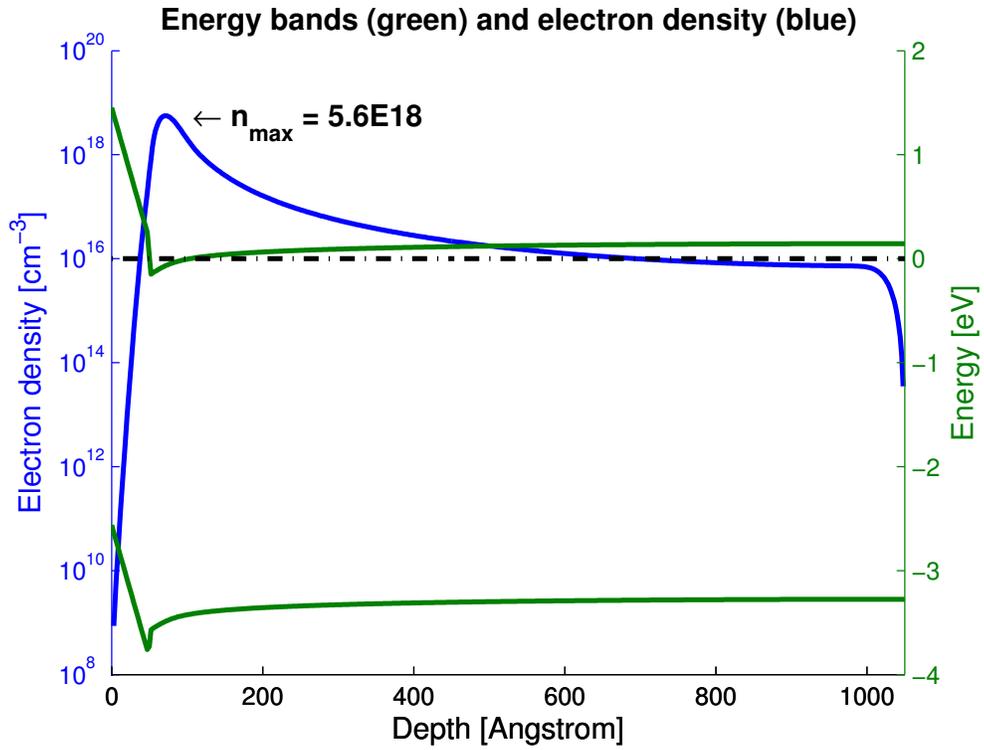


Figure 6.10: Test structure *A0_x0.3* (a 50 Å $Al_{0.3}Ga_{0.7}N$ layer) showing increased bending of the energy bands with increased Al-content forming a 2DEG.

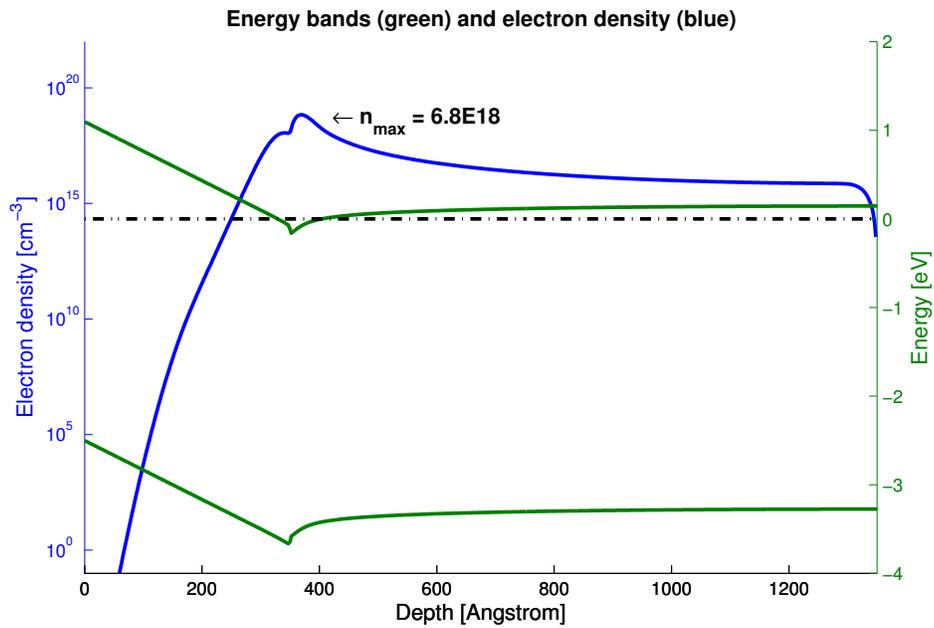


Figure 6.11: Test structure *A3_x0.1* (a 350 Å $Al_{0.1}Ga_{0.9}N$ layer) showing the formation of a 2DEG despite the low Al-content.

Effect of doping concentration

An increase in the peak density value of the 2DEG and the total integrated sheet electron density is expected if the donor concentration is increased, the opposite effect is evident in simulation results. The effect of the doping concentration is not reflected correctly by the simulator. The problem only becomes evident for high doping concentrations that approach the (equivalent) density of the polarization induced charge; the polarization effects 'disguise' this effect otherwise.

Inconsistencies are also evident from the bending of the energy bands: the conduction band should bend towards the conduction band (like a positive parabola) as the donor concentration in the AlGaIn layer is increased. The opposite is evident in the simulator when comparing the figures 6.12 and 6.13. This behaviour remains the same in both cases when solving the SP system and when solving only the Poisson equation. This indicates that the problem is caused by incorrect values of the electrostatic potential (as solved by Poisson's equation).

Remark: Some papers have considered the solution of the charge neutrality condition for GaN heterostructures to investigate the effects of doping in heterostructure [31]. The value of this is not well understood. The polarization effects present in Wurtzite crystals affect only the distribution of free charge carriers but does not actually act as a 'source' of electrons as doping does.

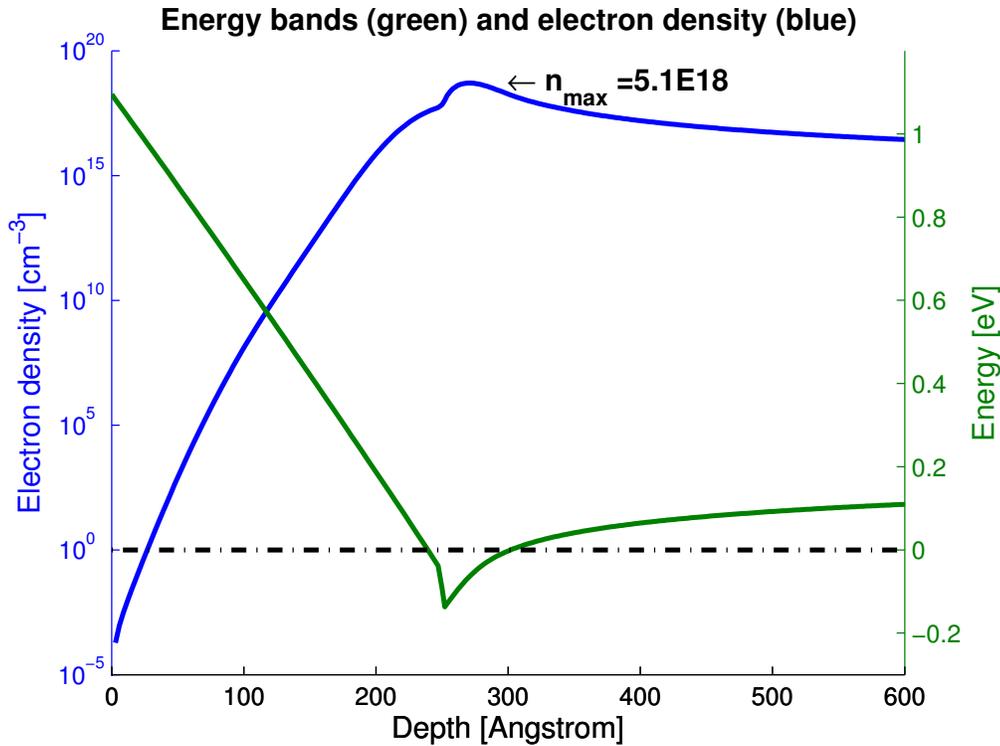


Figure 6.12: Test structure *B2_x0.1* with a doping of $N_D = 10^{17} \text{ cm}^{-3}$ showing a slight upward bending of the energy band. A peak electron density of $n_{\max} = 5.1 \times 10^{18} \text{ cm}^{-3}$ and the sheet electron density $n_s = 2.6 \times 10^{12} \text{ cm}^{-3}$.

Comparison to published results

A direct comparison to published experimental/simulation results for GaN HEMTs has been avoided but the results were studied to help interpret the behaviour of the GaN-

QUILA results.

A sensible comparison to published experimental results of GaN HEMT structures is very difficult since the significant polarization effects in GaN depend strongly on the preparation of the test structure(s). One can, however, gain insight on the what the qualitative behaviour of the device should be under various conditions. If own experimental measurements can be performed, a 'calibration' of the simulator is needed. This is done by adjusting the process-dependent material parameters (within a plausible physical range) to reproduce the experimental results with the simulator.

A comparison to published simulation results may be more sensible but the complete device dimension, doping, material parameters etc. are seldomly given in full detail which, again, makes only a qualitative comparison sensible.

Comparison to BandEng results

For comparison to the results of the BandEng simulator, the material parameters in the database were matched to those used in GaNQUILA (to the extent possible). The only notable absence is the lack of a temperature dependent bandgap in BandEng; this model was switched off in GaNQUILA for comparison purposes.

Table 6.4 (on page 60) shows a comparison of the charge densities for various test structures. A good agreement, although not exact, between the simulators is evident for most test structures. The best agreement is achieved for high Al-content where the polarization-induced charge is dominant. A notable exception is *A3_x0.3* (figure 6.14): it appears as if BandEng converges to an incorrect result as seems irregular compared to its other results - the band-bending is much too strong.

The problem of the incorrect band-bending in GaNQUILA with increased donor doping becomes evident again here when comparing structures with low Al-content ($x=0.1$) and high doping (10^{18} cm^{-3}) e.g. *C1_x0.1*. Figure 6.15 shows the incorrect band-bending very well. This is the underlying reason for the difference in electron densities.

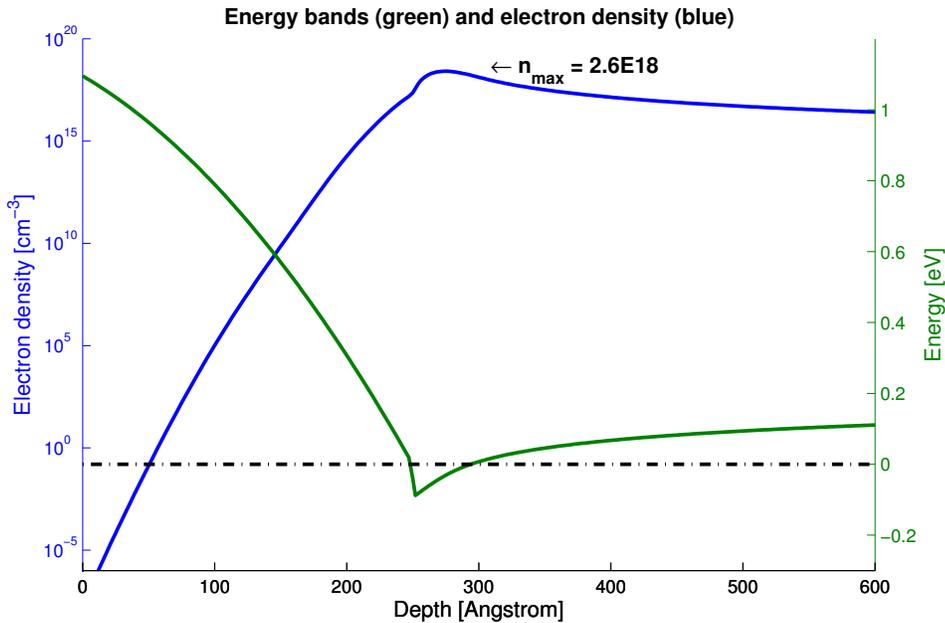


Figure 6.13: Test structure *C2_x0.1* with a doping of $N_D = 10^{18} \text{ cm}^{-3}$ showing a clear upward bending of the energy band. A peak electron density of $n_{\max} = 2.6 \times 10^{18} \text{ cm}^{-3}$ and the sheet electron density $n_s = 1.6 \times 10^{12} \text{ cm}^{-3}$.

In all cases both the peak electron density and the integrated electron sheet density are underestimated by GaNQUILA. The primary reason for this is believed to be the lastly discussed wrong bending of the energy bands, the calculation of the polarization charge may be another. The neglect of the cross-correlation term V_{xc} in the Hamiltonian of Schrodinger's equation could also explain the reduced peak density value but only to a negligible extent (as noted in [31] where a difference of 5×10^{18} occurs for a peak value 5.5×10^{19}).

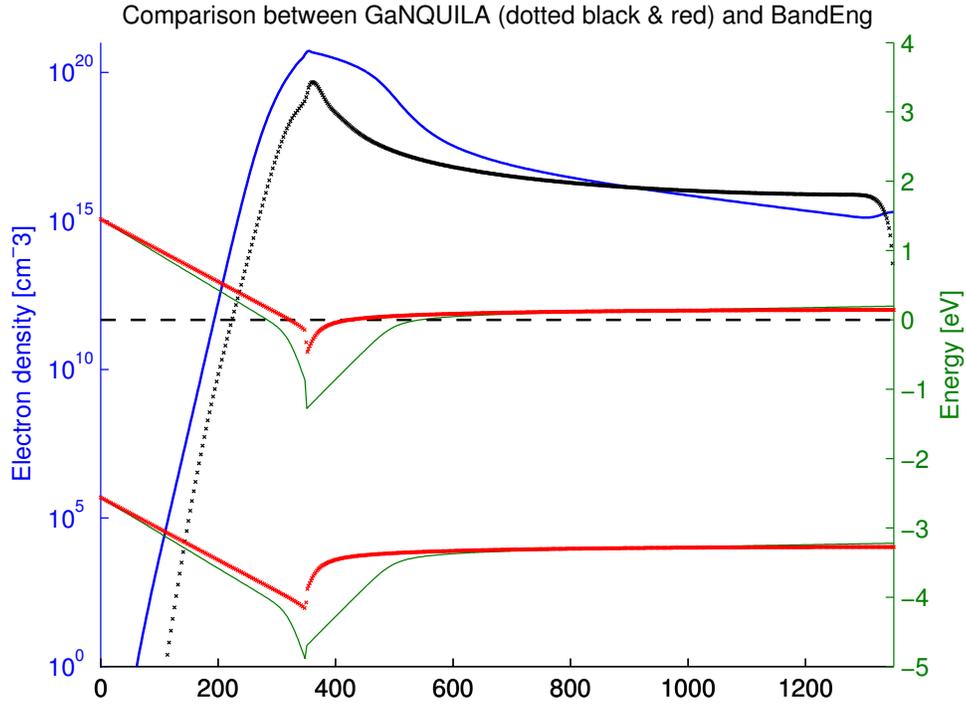


Figure 6.14: Test structure $A3_x0.3$ showing an irregular result from BandEng.

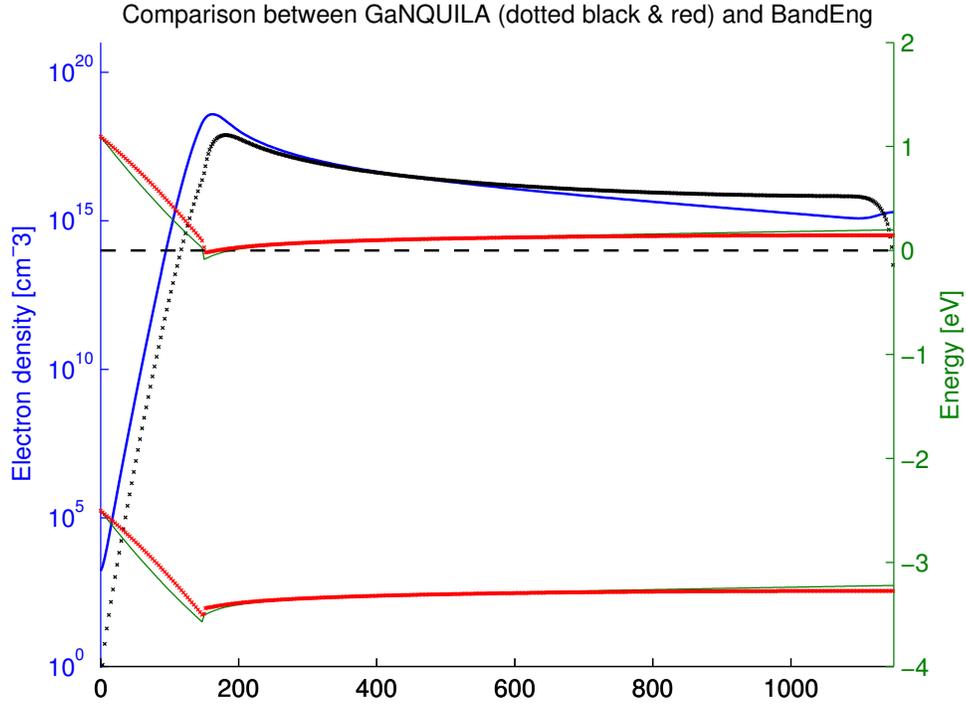


Figure 6.15: Comparison to BandEng for test structure $C1_x0.1$ showing a clear error of the band-bending of GaNQUILA (in red).

Table 6.4: Comparison of the peak electron densities and integrated electron densities of GaNQUILA and BandEng for various test structures.

Structure	$n_{\max} [\text{cm}^{-3}]$	$n_s [\text{cm}^{-2}]$	Simulator
$C1_x0.1$	3.87×10^{18}	2.07×10^{12}	BE
	7.66×10^{17}	6.91×10^{11}	GQ
$C1_x0.2$	2.12×10^{19}	7.28×10^{12}	BE
	1.40×10^{19}	5.25×10^{12}	GQ
$C1_x0.3$	4.95×10^{19}	1.28×10^{13}	BE
	3.51×10^{19}	1.05×10^{13}	GQ
$B2_x0.3$	5.71×10^{19}	1.59×10^{13}	BE
	4.48×10^{19}	1.34×10^{13}	GQ
$B2_x0.2$	2.58×10^{19}	8.61×10^{12}	BE
	2.18×10^{19}	7.75×10^{12}	GQ
$A3_x0.2$	2.81×10^{19}	9.36×10^{12}	BE
	2.36×10^{19}	8.79×10^{12}	GQ
$A3_x0.3$	5.26×10^{20}	4.03×10^{14}	BE
	4.68×10^{19}	1.45×10^{13}	GQ
$B3_x0.1$	8.22×10^{18}	3.82×10^{12}	BE
	6.53×10^{18}	3.26×10^{12}	GQ
$A2_x0.1$	6.04×10^{18}	2.93×10^{12}	BE
	5.39×10^{18}	2.71×10^{12}	GQ

Chapter 7

Conclusion

7.1 Summary

The simulation of the electron distribution in GaN/AlGa_N heterostructures was investigated for the use in a quasi-2D GaN HEMT simulator. The numerical solution and implementation of a one-dimensional Poisson-Schrodinger system was discussed, as well as the solution of the continuity equation in the nonequilibrium case. The simulator was implemented in MATLAB using the existing AQUILA code for GaAs/AlGaAs as a basis.

The developed simulator demonstrates the correct qualitative behaviour under the change of various, but not all, device parameters. The effect of (high) doping concentrations in the supply layer is not reflected correctly. Generally, a good agreement to the results generated by the BandEng software exists, for simple AlGa_N/GaN heterostructures in the thermal equilibrium case. The nonequilibrium case (with an applied gate voltage) has not been solved in a satisfactory manner. The numerical solution of the continuity equation together with the Schrodinger-Poisson system has failed. There is no convergence of the Fermi level and the obtained results are not plausible and cannot be used. This matter should be re-investigated.

7.2 Recommendations

This report has summarized the essential aspects that are needed to develop the intended simulator. It may prove fruitful to rewrite the entire simulation code from a clean slate to re-check the details of the code and to improve the general software architecture. Detailed attention needs to be given to numerical aspects e.g. equation scaling and matrix conditioning and the consequent efficiency of algorithms. These aspects have not been investigated properly up to now.

With a proper numerical treatment of the system of equations a convergence for the nonequilibrium case should be attainable using the approach sketched in this report. The solution of the nonequilibrium case is an essential feature for the eventual integration in the quasi-2D simulator. Furthermore, the presented alternative simulators do not account for the nonequilibrium case either, making its addition to GaNQUILA a valuable addition to the existing offering in free simulators. There are two essential problems that must be addressed before solution of the continuity equation is again attempted:

- Error in Poisson's equation causing incorrect band-bending with increased doping;
- The 'classical correction' term essential for a smooth coupling of classical and quantum regions and/or electrons. without causing discontinuities electron densities and/or zero electron densities at the boundaries of quantum region(s).

The primary goal should be to ensure the correct qualitative behaviour of the simulator to results reported in literature. Once this has been established the quantitative behaviour can be further improved using suggested softwares as comparisons. There are several further additions and improvements that can be made to the simulator once the basic equations can be solved in a numerically robust manner.

The rigorous verification of the numerical solutions of the individual equations is strongly recommended to better separate the convergence issues of the equations from those of the system. A selection of simple test structures should be designed for which the solution is well known or an exact analytic solutions exist. These structures can then be used to validate the numerical solution of equations and identify problems more effectively. Possibilities may be a n^+n junction to test potential or quasi-Fermi levels; a triangular potential well, for which an exact solution is known, to test the numerical code of Schrodinger's equation.

Currently, all electrons are calculated in a quantum manner without a 'classical correction' (section (5.3.5.3)) to ensure better convergence of the SP system. It would be beneficial to separate the calculation and treatment of 3D and 2D electrons more rigorously since only 3D electrons should be considered for the gate current and it also offers reduced computation time. The results obtained with the current 'classical correction' have shown some inconsistencies. This matter should be thoroughly investigated both in theory and code - a good starting point is given by Takano [35].

The accurate simulation of multilayer heterostructures, like GaN/AlGaIn/GaN, should also be pursued since they are becoming more common in GaN HEMTs. The method with which the strain in the layer(s) is currently calculated is not expected to accurately handle multilayer structures. The most accurate method would be to minimize the total strain in the structure using a conjugate gradient method, for instance, and then calculating the polarization effects from this using strain tensors.

Further additions the GaNQUILA simulator may include:

- The expansion of the Hamiltonian in Schrodinger's equation to incorporate cross-correlation effects which may be important to account for scattering effects outside the quantum well.
- The consideration of surface states - they play an important role in the ungated sections of the HEMT to supply the free electrons needed to form the 2DEG. Using the finite volumes scheme they can be incorporated using a δ -doping layer on the surface.
- The investigation of the electro-mechanical coupling that exists between the electric field and the stress tensor by introducing an additional equation to the system.
- The consideration of electrothermal effects that are important in high-power devices by introducing a heat (phonon) transport equation.
- The addition of C-V profiling to the simulator by calculating the charge at several voltages and calculating $dC = \frac{dQ}{dV}$ (once the non-equilibrium case can be solved).

For further queries or support, the author can be contacted <pellinhaus@ieee.org>.

Bibliography

- [1] H. Morkoç, *Handbook of Nitride Semiconductors and Devices: Electronic and Optical Processes in Nitrides Vol. 1: Materials Properties, Physics and Growth*. Weinheim: Wiley-VCH, 2008.
- [2] A. Ashok, D. Vasileska, S. Goodnick, and O. Hartin, “Importance of the gate-dependent polarization charge on the operation of GaN HEMTs,” *Electron Devices, IEEE Transactions on*, vol. 56, pp. 998–1006, May 2009.
- [3] J. Pankove, E. Miller, and J. Berkeyheiser, “GaN electroluminescent diodes,” in *Electron Devices Meeting, 1971 International*, vol. 17, p. 78, 1971.
- [4] S. Nakamura, “GaN growth using GaN buffer layer,” *Japanese Journal of Applied Physics*, vol. 30, no. Part 2, No. 10A, pp. L1705–L1707, 1991.
- [5] S. Nakamura, N. Iwasa, M. Senoh, and T. Mukai, “Hole compensation mechanism of p-type GaN films,” *Japanese Journal of Applied Physics*, vol. 31, no. Part 1, No. 5A, pp. 1258–1266, 1992.
- [6] S. Nakamura, T. Mukai, and M. Senoh, “High-power GaN p-n junction blue-light-emitting diodes,” *Japanese Journal of Applied Physics*, vol. 30, no. Part 2, No. 12A, pp. L1998–L2001, 1991.
- [7] M. A. Khan, A. Bhattarai, J. N. Kuznia, and D. T. Olson, “High electron mobility transistor based on a GaN-Al_xGa_{1-x}N heterojunction,” *Applied Physics Letters*, vol. 63, no. 9, pp. 1214–1215, 1993.
- [8] M. Asif Khan, J. N. Kuznia, A. R. Bhattarai, and D. T. Olson, “Metal semiconductor field effect transistor based on single crystal GaN,” *Applied Physics Letters*, vol. 62, pp. 1786–1787, Apr 1993.
- [9] J. Pankove, S. Chang, H. Lee, R. Molnar, T. Moustakas, and B. Van Zeghbroeck, “High-temperature GaN/SiC heterojunction bipolar transistor with high gain,” in *Electron Devices Meeting, 1994. IEDM '94. Technical Digest., International*, pp. 389–392, Dec 1994.
- [10] H. Morkoç, *Handbook of Nitride Semiconductors and Devices: Electronic and Optical Processes in Nitrides Vol. 3: GaN-based optical and electronic devices*. Weinheim: Wiley-VCH, 2009.
- [11] M. Micovic, N. Nguyen, P. Janke, W.-S. Wong, P. Hashimoto, L.-M. McCray, and C. Nguyen, “GaN/AlGaN high electron mobility transistors with f_{tau} of 110 GHz,” *Electronics Letters*, vol. 36, pp. 358–359, Feb 2000.
- [12] O. Ambacher, J. Smart, J. R. Shealy, N. G. Weimann, K. Chu, M. Murphy, W. J. Schaff, L. F. Eastman, R. Dimitrov, L. Wittmer, M. Stutzmann, W. Rieger, and J. Hilsenbeck, “Two-dimensional electron gases induced by spontaneous and piezoelectric polarization charges in n- and Ga-face AlGaN/GaN heterostructures,” *Journal of Applied Physics*, vol. 85, pp. 3222–3233, Mar 1999.

- [13] J. Piprek, *Nitride Semiconductor Devices : Principles and Simulation*. Weinheim: Wiley-VCH, 2007.
- [14] Y. Developpement, “GaN rf market analysis ’08,” 2008.
- [15] K. Gurnett and T. Adams, “Native substrates for GaN: the plot thickens,” *III-Vs Review*, vol. 19, no. 9, pp. 39 – 41, 2006.
- [16] J. Brown, R. Borges, E. Piner, A. Vescan, S. Singhal, and R. Therrien, “AlGaIn/GaN hfets fabricated on 100-mm GaN on silicon (1Å 1Å 1) substrates,” *Solid-State Electronics*, vol. 46, no. 10, pp. 1535 – 1539, 2002.
- [17] C. C. Kim, J. H. Je, M. S. Yi, D. Y. Noh, F. Degave, and P. Ruterana, “Microstructure of GaN nucleation layer during initial stage movcd growth,” *Materials Science and Engineering B*, vol. 82, no. 1-3, pp. 108 – 110, 2001.
- [18] O. Ambacher, “Ga(n)laktisch,” *Aktuelle Technik*, pp. 56 –58, Mar 2011.
- [19] Y.-S. Lee and Y.-H. Jeong, “A high-efficiency class-e GaN HEMT power amplifier for wcdma applications,” *Microwave and Wireless Components Letters, IEEE*, vol. 17, pp. 622 –624, Aug. 2007.
- [20] T. Mimura, “The early history of the high electron mobility transistor (hemt),” *Microwave Theory and Techniques, IEEE Transactions on*, vol. 50, pp. 780 –782, Mar 2002.
- [21] Y.-F. Wu, B. Keller, P. Fini, S. Keller, T. Jenkins, L. Kehias, S. Denbaars, and U. Mishra, “High al-content AlGaIn/GaN modfets for ultrahigh performance,” *Electron Device Letters, IEEE*, vol. 19, pp. 50 –53, Feb 1998.
- [22] H. Ilati, M. R. Ashrafi, and S. Khorasani, “Simulation and optimization of nanoscale HEMTs with finite elements method,” *ArXiv e-prints*, Feb 2011.
- [23] W. Lu, V. Kumar, R. Schwindt, E. Piner, and I. Adesida, “A comparative study of surface passivation on AlGaIn/GaN HEMTs,” *Solid-State Electronics*, vol. 46, no. 9, pp. 1441 – 1444, 2002.
- [24] J. Bernájt, P. Javorka, A. Fox, M. Marso, H. LÄEth, and P. Kordos, “Effect of surface passivation on performance of AlGaIn/GaN/si HEMTs,” *Solid-State Electronics*, vol. 47, no. 11, pp. 2097 – 2103, 2003.
- [25] R. Doe, “Electronic archive: New semiconductor materials. characteristics and properties type @ONLINE,” Sep 2011.
- [26] M. Suzuki, T. Uenoyama, and A. Yanase, “First-principles calculations of effective-mass parameters of aln and GaN,” *Phys. Rev. B*, vol. 52, pp. 8132–8139, Sep 1995.
- [27] A. C. Schmitz, A. T. Ping, M. A. Khan, Q. Chen, J. W. Yang, and I. Adesida, “Schottky barrier properties of various metals on n-type GaN,” *Semiconductor Science and Technology*, vol. 11, no. 10, p. 1464, 1996.
- [28] D. Qiao, L. S. Yu, S. S. Lau, J. M. Redwing, J. Y. Lin, and H. X. Jiang, “Dependence of ni/AlGaIn schottky barrier height on al mole fraction,” *Journal of Applied Physics*, vol. 87, no. 2, pp. 801–804, 2000.
- [29] A. Anwar, E. Faraclas, and K. Smith, “Schottky barrier height in GaN/AlGaIn heterostructures,” in *Semiconductor Device Research Symposium, 2005 International*, pp. 221 –222, Dec. 2005.
- [30] Z. Lin, W. Lu, J. Lee, D. Liu, J. S. Flynn, and G. R. Brandes, “Barrier heights of schottky contacts on strained AlGaIn/GaN heterostructures: Determination and effect of metal work functions,” *Applied Physics Letters*, vol. 82, no. 24, pp. 4364–4366, 2003.

- [31] B. Jogai, "Free electron distribution in AlGa_N/Ga_N heterojunction field-effect transistors," *Journal of Applied Physics*, vol. 91, no. 6, pp. 3721–3729, 2002.
- [32] J. A. C. Pã©rez, "On the significance of the surface states in isolated alxgal-xn/Ga_N heterostructures," *Solid-State Electronics*, vol. 49, no. 4, pp. 612 – 617, 2005.
- [33] A. Ashok, D. Vasileska, S. Goodnick, and O. Hartin, "Importance of the gate-dependent polarization charge on the operation of Ga_N HEMTs," *Electron Devices, IEEE Transactions on*, vol. 56, pp. 998 –1006, may 2009.
- [34] S. Sze and K. Ng, *Physics of semiconductor devices*. Wiley-Interscience publication, Wiley-Interscience, 2007.
- [35] C. Takano, Z. Yu, and R. Dutton, "A nonequilibrium one-dimensional quantum-mechanical simulation for algaas/gaas HEMT structures," *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, vol. 9, pp. 1217 –1224, Nov 1990.
- [36] F. Sacconi, A. Di Carlo, P. Lugli, and H. Morkoc, "Quasi two-dimensional modeling of Ga_N-based modfets," *physica status solidi (a)*, vol. 188, no. 1, pp. 251–254, 2001.
- [37] A. Asgari, M. Kalafi, and L. Faraone, "A quasi-two-dimensional charge transport model of AlGa_N/Ga_N high electron mobility transistors (HEMTs)," *Physica E: Low-dimensional Systems and Nanostructures*, vol. 28, no. 4, pp. 491 – 499, 2005.
- [38] R. Singh and C. Snowden, "A quasi-two-dimensional HEMT model for dc and microwave simulation," *Electron Devices, IEEE Transactions on*, vol. 45, pp. 1165 – 1169, Jun 1998.
- [39] A. Trellakis, A. T. Galick, A. Pacelli, and U. Ravaioli, "Iteration scheme for the solution of the two-dimensional schrodinger-poisson equations in quantum structures," *Journal of Applied Physics*, vol. 81, pp. 7880 –7884, Jun 1997.
- [40] A. Pacelli, "Self-consistent solution of the schrodinger equation in semiconductor devices by implicit iteration," *Electron Devices, IEEE Transactions on*, vol. 44, pp. 1169 –1171, Jul 1997.
- [41] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes in C (2nd ed.): the art of scientific computing*. New York, NY, USA: Cambridge University Press, 1992.
- [42] D. C. Sorensen, "Implicitly restarted arnoldi/lanczos methods for large scale eigenvalue calculations," in *IN*, pp. 23–25, Kluwer, 1995.
- [43] S. C. Jain, M. Willander, J. Narayan, and R. V. Overstraeten, "Iii?nitrides: Growth, characterization, and properties," vol. 87, no. 3, pp. 965–1006, 2000.
- [44] E. A. B. Cole, T. Boettcher, and C. M. Snowden, "Corrections to the calculation of bulk electron densities in quantum wells of HEMTs," *Semiconductor Science and Technology*, vol. 12, no. 1, p. 100, 1997.
- [45] A. Spinelli, A. Benvenuti, and A. Pacelli, "Self-consistent 2-d model for quantum effects in n-mos transistors," *Electron Devices, IEEE Transactions on*, vol. 45, pp. 1342 –1349, Jun 1998.