Performance Predictions for AlGaN/GaN Heterojunction Bipolar Transistors

by

Sasan Fathpour

B.Sc., Isfahan University of Technology, Isfahan, Iran, 1995

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

Master of Applied Science

in

THE FACULTY OF GRADUATE STUDIES

(Department of Electrical and Computer Engineering)

We accept this thesis as conforming
to the required standard

The University of British Columbia

August 2000

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Electrical and Computer Engineering

The University of British Columbia
Vancouver, Canada

Date August 8th, 2000

DE-6 (2/88)

Abstract

Heterostructure transistors made from wide band-gap, wurtzite, nitride semiconductors are candidates for high-power, and high-temperature electronics. AlGaN/GaN heterojunction bipolar transistors (HBTs) have been studied since 1998, and the performance of fabricated devices has been improving steadily. However, the limits to the performance of these devices are not known. The main objective of this work is to achieve theoretical performance predictions for the characteristics of abrupt-emitter, and graded-emitter, *n-p-n* $Al_xGa_{1-x}N$/GaN HBTs.

As a direct band-gap material, GaN has a high radiative recombination rate, which tends to dominate the base current. Therefore, the direct recombination probability is readily recognized as the limiting factor of the current gain. According to the employed semi-empirical model for the recombination lifetime, gains as high as 300 are predicted at room-temperature, while the figure decreases to around 100 at 600 K. There is a high spontaneous polarization in $Al_xGa_{1-x}N$ nitrides. It is shown that the related intrinsic charges can degrade the gain considerably for layers which are N-faced, but there is no impact on the gain for Ga-faced layers. The incomplete ionization of magnesium, the presently employed acceptor in these materials, is another limiting factor of the devices; it leads to high-level injection in the base, which limits the ideal operating of the collector current density to below a few kA cm^{-2}.

Regarding the high-frequency performance of HBTs, a comparison between the

traditional analytic expressions and numerical regional signal-delay times has been accomplished. Partitioning the device has been performed by means of a simple phenomenological scheme. The study has been mostly performed on $Al_{0.3}Ga_{0.7}As/GaAs$ HBTs, where the modeling parameters are mature, high-level injection is not a limitation, spontaneous polarization is not an issue, and complete ionization is a fair approximation. It is shown that the available expressions are inadequate to account for the mobile charge in the base-emitter space-charge region, as well as in the quasi-neutral emitter. It is suggested to simply multiply the base-emitter junction capacitance, calculated by the depletion approximation, by a correction factor of about two as a practical approach for accounting for these storage phenomena in graded-emitter HBTs. For the base-delay time, it is demonstrated that choosing the appropriate collection velocity is crucial in estimating the delay time analytically, if band-gap narrowing, quasi-ballistic transport, or hot-electron injection are important. For the collector-delay time, it is shown that the effect of the mobile charge in the collector-base space-charge region cannot be neglected, even at biases below the onset of the Kirk effect. The effects of incomplete ionization and high-level injection in AlGaN/GaN HBTs are discussed separately, and a practical cut-off frequency of 10 GHz is predicted for abrupt-emitter transistors.

Contents

List of Tables

List of Figures

Acknowledgements

This work would not have been successful without the exceptional supervision of Professor David L. Pulfrey. He has been much more than a coach, and has always been cheerfully and patiently available for any question and discussion, for which I am deeply grateful. His friendly smile after any tiny progress might have been the ultimate motivation for the fulfillment of my M.A.Sc. degree. The highlight of his paternal conduct was to invite me to his family party on the Christmas eve of the new millennium. Although in order to venture a new environment for my Ph.D. program I have decided to leave Vancouver, this gentleman will surely be missed.

I must thank my former senior colleagues, Drs. Mani Vaidyanathan and Anthony St. Denis. Their involvement in this work was not only to provide guidance to a first year graduate student on the course of research, but they also have direct contribution to chapter 5 of the thesis. In particular, Mani has generously devoted a lot of time on our discussions, even after leaving our group, which enhanced my understanding of bipolar transistors.

I should also express my gratitude for the nice friends I have found here. Landing in Canada as an international student, settling and adapting to the new environment would not be as smoothly as it had been without the help of Misters Farinam Farahmand, Hossein Saboksayr, Sassan Tabatabaei, Ali Moshtaghi and Mohammad Fallahi. The great time I have shared with them has been the best relaxation for me.

Finally, I would like to thank Professors Tom Tiedje and Mike K. Jackson for their valuable time spent as members of my examination committee.

SASAN FATHPOUR

The University of British Columbia

August 2000

To my parents,

who put me on the track!

Chapter 1

Introduction

1.1 Overview of Group-III Nitride Semiconductors

Most introductory texts on semiconductors state at some stage that a semiconductor is a solid material with a conductivity neither as high as conductors (like silver) nor as low as insulators (like mica). Obviously this definition, in terms of a property which covers a wide range without any exact limit, is vague. For instance, the conductivity of silver is about 10^{22} times greater than that of mica at room temperature [1]. According to Ref. [2] (Chap. 4, p. 118), a semiconductor is a material which has a resistivity from 10^{-2} to 10^{6} Ω-cm at a certain temperature. In principle, at sufficiently high temperature the resistivity of any insulator can decrease to this range. In terms of band-gap, the cited reference, published in the late 1980's, classifies semiconductors as materials with a band-gap less than 3.5 eV, with the justification that "all practical semiconductors" lie in this regime. On the other hand, it is possible to categorize semiconductors as those solids where the mean of the periodic table group

the participant elements belong to is group IV [1]. It should be emphasized here that this is another vague definition, and has many violations, *e.g.*, diamond, which is certainly not a semiconductor. This definition covers the well-known categories: elemental (Si, Ge), IV compounds (SiGe, SiC), binary III-V compounds (GaAs, InP), II-VI compounds (ZnS, CdTe), and, similarly, many ternary and quaternary combinations. These are the most common types, but one can generalize to other classes, like IIA-VI (MgO). From a combinational analysis point of view, one can count up to hundreds of alloys, made out of group III, IV, and V elements, that might have semiconducting properties in terms of conductivity or band-gap. However, it does not necessarily mean that the combination is chemically stable.

It was through a systematic study of the crystal structure and band-gap of various compound materials, that Juza and Hahn studied a needle-size sample of a nitride semiconductor (GaN) in 1940 [3]. This is not the first report on nitrides though, and actually AlN powder was made in the mid-eighteen hundreds. However, due to certain technological problems, which are discussed later, these interesting materials would not emerge as possibly useful materials until the 1990's. What we are interested in are the compounds (binary and ternary) of the group III elements (B^5, Al^{13}, Ga^{31}, and In^{49}) with nitrogen (N^7). Due to the large difference in electro-negativity, a stable covalent bond is created between the N anions and the corresponding cation(s). BN does not appear to have electronic applications, and is not discussed here. The other three binary alloys (AlN, GaN, and InN) and corresponding ternary alloys (AlGaN, AlInN, *etc.*) can attain one of three polytypes of crystal structure, namely wurtzite, zincblende, and rock-salt [4], [5]. The transition to the rock-salt phase occurs at high pressure. Thin films of the meta-stable zincblende phase can be grown on cubic substrates such as Si, MgO, and GaAs, but have not been extensively studied yet. The

thermodynamically stable phase is the wurtzite structure. It is the easiest to grow, and is presently the most promising one for electronic and opto-electronic applications. The employed substrates are usually sapphire or SiC. The band-gaps of the wurtzite nitrides are: GaN (~ 3.4), AlN (~ 6.2), InN (~ 1.9) eV at room temperature. These three direct band-gap wurtzite materials and the corresponding ternary alloys are what is meant by nitrides throughout this work, unless otherwise stated. As can be seen, the nitrides are wide band-gap materials, and can be roughly assumed to be semiconductors. Especially with respect to the aforementioned definition based on band-gap, AlN is virtually an insulator, recalling that the band-gap of diamond as a non-controversial insulator is about 0.2 eV smaller. However, it is usually employed as $Al_xGa_{1-x}N$, for which the band-gap of the alloy with an Al mole-fraction $x < 0.25$ is typically less than 4.0 eV . Recalling that nitrides have grown into practical semiconductor compounds in the 1990's, mostly for opto-electronic applications, it seems that 3.5 eV as an accepted demarcation between insulators and semiconductors in the 1980's needs some modification, and 4.0 eV might be a more up-to-date value.

A brief summary of the technological aspects of nitrides, including the fabrication techniques and the historical challenges, is presented in Sec. 1.1.1. Various applications of nitrides in device engineering are the subject of Sec. 1.1.2, introducing nitride-based transistors among the candidates for high-temperature and high-power electronics. Sec. 1.2 presents the outline of the following chapters of the present work, where performance predictions for AlGaN/GaN HBTs are stated as the main objective of the thesis.

1.1.1 Technological Aspects of Nitrides

The history of nitrides is rather ambiguous. As mentioned before, Pankove and Moustakas [3] refer to the study by Juza and Hahn as the first systematic study of nitrides. They produced a needle-size sample of nitride GaN in 1940, by passing ammonia over hot gallium. Ref. [14] addresses an earlier publication by J. V. Lirman and H. S. Zhdanov in 1937 as the first study on the crystalline structure of nitrides. According to the same citation, the nitrides were among the first semiconductors ever discovered in 1907. Ambacher [12] refers to the oldest reference as being in 1862, when AlN powder was synthesized from liquid Al and N_2 gas.

There are basically three modern techniques utilized to grow nitrides, namely hydride vapor phase epitaxy (HVPE), metal-organic chemical vapor deposition (MOCVD), and molecular beam epitaxy (MBE). The details of the techniques are beyond the scope of this work and can be found elsewhere [8], [9], [10], [11], [5], [12]. Briefly, HVPE is an enhanced chemical vapor deposition process, using ammonia as the nitrogen source and GaCl as the Ga source (in the case of GaN), produced by flowing HCl over Ga melt at $800 - 900°C$ via the reaction [12],

$$2Ga(l) + 2HCl(g) \leftrightarrow 2GaCl(g) + H_2(g). \tag{1.1}$$

GaN is formed on the substrate at 1100°C through [8], [12],

$$GaCl(g) + NH_3(g) \leftrightarrow GaN(s) + HCl(g) + H_2(g). \tag{1.2}$$

This method has high growth rates, and thick films that can be removed from the substrate can be grown. Thus, potentially, it is useful to produce nitride substrates for homo-epitaxy [12]. Such HVPE-grown substrates have been used to grow homo-epitaxial films, but the method is still in an immature stage of development [10].

Maruska and Tietjen [13] used a chemical vapor deposition technique to produce the first large-area layer of GaN on a sapphire substrate. This was historically an important step in the technology of nitrides [3]. Modern MOCVD is the commonly used technique in the industry for fabricating nitride opto-electronic devices, partially due to the potential of large-scale manufacturing, and partially due to the high film quality. Basically, it employs the same nitrogen source as the HVPE technique, but uses the tri-methyls of the group-III element(s). The chemistry of the technique is rather complicated, but the basic reaction is [9],

$$Ga(CH_3)_3(g) + NH_3(g) \rightarrow GaN(s) + 3CH_4(g). \tag{1.3}$$

In the MBE technique, the cation source is liquid Ga. Regarding the atomic nitrogen source required for this method, three approaches have been utilized, namely: heat-assisted decomposition of ammonia on the substrate surface with a temperature of $700-900^{\circ}$C, ionization by plasma or by low-pressure discharge [11]. The advantages of the MBE technique are excellent process control and *in situ* monitoring of the growth for research purposes, as well as a lower growth temperature compared to MOCVD, which is crucial for high indium-containing alloys [12].

Sapphire (α-Al_2O_3) is the most common substrate used in manufacturing nitride films. It has the complex rhombohedral structure [12], and the *c*-plane, (0001), of that structure is usually employed to grow nitride films. Surprisingly, this plane has a large lattice mismatch of about 13.9% with the *c*-axis of GaN, and the critical thickness of the grown layer is less than a mono-layer, consequently the lattice strain is relieved immediately [14]. Also, the thermal expansion coefficients are substantially different between nitrides and sapphire. However, it remains the most promising substrate due to the low cost, lack of better alternatives, as well as transparency, and stability at high temperature. The other common substrate material is silicon carbide,

where the 6H structure is commercially available, and the (0001) plane is appropriate for growing the wurtzite structure. Specially important, it has a less than 1% lattice mismatch with AlN, and perfect junctions can be achieved. The important challenge for 6H-SiC at the present time, apart from its cost, appears to be the lack of appropriate chemical etch [5]. Available wurtzite ZnO could be another alternative as a substrate material, but it is unstable at high growth temperatures, and at the high operating temperatures of interest in some applications. Finally, as mentioned before, zincblende nitrides can be grown on cubic substrates such as Si, MgO, GaAs, and 3C-SiC. It is worthwhile noting herein that cubic nitrides are still in their infancy, and beyond the lattice constants, very little experimental data are available for them. Theoretical calculations show that the band-gaps appear to be typically smaller than that of their wurtzite older sisters [15]. The electrical and opto-electronic properties are not known yet. Therefore, the simple fact that wurtzite nitrides are more convenient, does not necessarily mean that cubic polytypes may not have interesting properties.

Virtually any semiconductor device of engineering interest requires at least one *p-n* junction. Therefore, technology of any semiconductor material cannot be prosperous if promising donor and acceptor dopants are not developed. This has actually been an important challenge in the history of nitride technology. GaN films are known as unintentionally *n*-type materials, as Hall measurements show a high background electron concentration. In early material, this was partially due to the defects in the lattice structure. However, this cannot be the case in the rather defect-free materials obtained by state-of-the-art growth techniques. So, the reason for the unintentional background electron concentration is still controversial, but has been long attributed to the nitrogen vacancies in the lattice, which act as defect centers [16].

For many applications, like heterojunction field effect transistors (HFETs), where a high-mobility channel is required on top of a low-conducting layer, the background doping can degrade the performance of the device. A commonly used technique to attain low background doping is growing a AlN or GaN buffer layer on the substrate at low-temperature, prior to the growth of functional layers of the device. This technique appears to be useful in reducing the background concentration from $10^{18} - 10^{20}$ cm^{-3} to mid-10^{16} cm^{-3} [10], and is practically employed in the fabrication of HFETs [17]. To grow controllable n-type nitrides, oxygen has been used, and germanium is another alternative. However, silicon has become the most commonly used donor impurity; it substitutes for Ga. The ionization (activation) energy is about 16 meV in GaN [18]. Si can also be used for all compositions of AlGaN, but the ionization energy increases with Al mole fraction.

Seeking the appropriate acceptor has been a major issue among researchers for two decades (1970-90). A lot of effort was concentrated on zinc in the 1970's. Nevertheless, the p-type materials were poorly conducting, due to the deep ionization level of Zn in the band-structure of GaN. Discovering the appropriate technique to employ magnesium as an alternative by Amano *et al.* [19], in 1989, is perhaps the most noteworthy breakthrough in the history of nitride technology. The ionization level is still deep, though, (about 175 meV in GaN), and increases with the mole fraction in AlGaN. This deep level causes low ionization of about 4% for a rather high doping concentration of 10^{19} cm^{-3} at room temperature. This hugely affects the electrical performance of the devices made out of nitrides. In addition, using Mg for AlGaN alloys is difficult, and synthesizing a p-type AlGaN with a mole fraction of over 30% has not proved feasible yet [16]. Therefore, finding a more promising acceptor impurity seems to await another breakthrough.

We conclude this section with a short discussion on n- and p-type contacts on nitrides. The requisites are thermal stability for high-power and high-temperature applications, as well as low resistance contacts. To achieve low resistivity, thin tunneling contacts are a solution, provided that high doping density contact layers can be grown. Al is the most common n-type contact metal [1], but employing Al/Cr is common too [20]. For thermal stability concerns, W and Ti/WSiN are appropriate choices, with low ohmic resistance and stability up to 800 °C for 1 hour [12]. Tungsten was actually the material used for the emitter contact of one of the first fabricated AlGaN/GaN HBTs [21]. For contacting p-type materials, usually multilayer metals like Cr/Al or Au/Mg are utilized [1]. These are basically borrowed from the GaAs technology, and a variety of multilayer combinations like Ti/Al/Pr/Au, Au/Ti/Al (for the n-contact), and Ni/Pt/Au and Au/Ti/Ni (for the p-contact) are reported in the fabricated devices as well [22], [23].

1.1.2 Why Nitrides ?

Nitrides have been primarily of interest for opto-electronic applications. Wurtzite nitrides are direct band-gap materials with high radiative recombination (this is not the case in all the zincblende phases, where AlN is indirect [24]). The wavelengths corresponding to an arbitrary wurtzite AlInGaN alloy cover, in principle, from 650 nm (orange) in InN, to yellow, green, blue and violet, and finally to 200 nm (ultraviolet) in AlN. Thus, a wide range of the visible and UV spectrum is covered. Manufacturing high-brightness, light-emitting diodes (LEDs) and lasers in the mentioned visible colors (especially blue) has been a dream for a long time. The LEDs can have application in full color LED displays, indicators, and lamps with high reliability, efficiency, and speed. Prior to the emergence of the nitrides in the early 1990's, SiC and GaP

had been commercialized for blue and green LEDs, respectively. Nevertheless, since they are indirect materials, the brightness is very low, and the efficiencies very small. II-VI compounds (such as ZnSe) are another alternative, but they appear to have many weaknesses, *e.g.*, short lifetime, and poor thermal stability [12]. Therefore, it is not surprising that the nitride technology started growing rapidly when the aforementioned appropriate acceptor impurity issue was solved in 1989. The first blue LED was reported by Nakamura *et al.* in 1991 [25]. In 1994 the first full-color LED display using an InGaN/AlGaN heterostructure for the blue color was demonstrated in Tokyo, Japan. Also, the first LED traffic light using green InGaN single quantum well, yellow AlInGaP, and red AlInGaP LEDs was demonstrated in Sweden in 1996 [26]. Regarding the nitride-based lasers, they are basically of interest for the purpose of digital information storage. The shorter wavelength coherent light of GaN-based lasers implies that a higher density of data (over 1 Gb/cm^2 [12]) can be stored compared to bulky expensive red lasers. Another optical application of nitrides is in UV photo-detection, where the wavelength is less than 400 nm. Solar-blind detectors are of especial commercial interest for detection within the Hartley Ozone band (200-300 nm). A mole fraction of about 0.33 in AlGaN can give a sharp cutoff at 290 nm, required for this application [27].

Nitride electronic devices are of interest primarily for high power, high temperature applications. Higher power is always limited by the inevitable higher temperature caused by internal losses. Silicon transistors need appropriate cooling means to keep the temperature below 200°C, so that the material continues to behave as an extrinsic semiconductor. The large band-gaps of nitrides keep them extrinsic at elevated temperatures. In addition, the chemical stability and inertness of nitrides marks them as promising candidates for high power applications, without the need

for expensive and bulky heat sinks. Meanwhile, the large band-gap leads to the high breakdown field desired for such applications. Also, the high electron saturation velocity and the somewhat high mobility implies reasonably high-frequency devices. Nitride transistors are predicted to find extensive applications in harsh environments, *e.g.*, as found in the automobile, and aerospace industries. The bulky and heavy hydraulic and mechanical control systems in aircraft will be replaced by nitride electronic circuitry, in a technology which is claimed will lead to the *all-electrical aircraft* [24]. The weighty heat radiators in satellite circuits can be eliminated as well, if the technology gets mature, and reliable enough. These are general statements valid for both bipolar and field effect transistors. Regarding the competition between these two types of transistors, the bipolar one is more advantageous, basically because of higher power that can be gained, as well as better linearity, which is required for amplification. The main disadvantage of nitride-based bipolar transistors seems to be the high radiative recombination in the base, which significantly affects the gain of the device. The material system that researchers have been more interested in is $Al_xGa_{1-x}N$. Using a wide band-gap emitter (AlGaN) improves the performance of the device. The technology of these devices is still in its infancy, and very few fabricated devices are reported. Prediction of the performance of AlGaN/GaN heterojunction bipolar transistors is the main objective of the present work as presented in the subsequent chapters. In particular, where unconventional physical phenomena arise, they are taken into account in modeling these emerging devices.

1.1.3 Summary

- Nitride semiconductors are the alloys of Al, Ga, and In from the group-III elements with nitrogen. These can be found in rock-salt, zincblende, and wurtzite

phases; the latter is the most thermodynamically stable and easiest to grow.

- Wurtzite nitrides are wide band-gap direct materials, widely suitable for short wavelength opto-electronic applications.

- Nitrides are also of interest for high-power, high-temperature, and high-frequency electronics in harsh environments; AlGaN/GaN HBTs are among the most promising devices.

- Modern HVPE, MBE, and MOCVD techniques can be used for growing nitride devices, MOCVD being the most common one. Si, and Mg are the best available donor and acceptor dopants, respectively; and sapphire and 6H-SiC the most common substrates, for the present time.

- The wide band-gap, the polarity of the *c*-axis in the wurtzite structure, the large lattice mismatch in heterostructures of different nitrides, the large piezoelectric coefficients, and the lack of a shallow acceptor impurity cause unconventional features in understanding the characteristics of nitride devices.

1.2 Thesis Outline

Chapter 2 presents the electrical properties of AlN, and GaN. As a new material system, not all the parameters required for modeling nitride transistors are widely known. It is attempted to collect the most recent data in the literature, favoring the reports of temperature-dependence due to the application of nitrides in high-temperature electronics. On the other hand, for a few cases, *e.g.*, majority-carrier mobilities and direct recombination probability, new fittings were achieved in this work.

Chapter 3 discusses the aspects of non-symmetry along the c-axis of wurtzite nitrides. It is thoroughly discussed how the consequent spontaneous and piezoelectric polarizations can affect Poission's equation and the corresponding boundary conditions needed to be solved in device modeling. It might seem a resolved classical problem, however, due to the fact that the polarity of the polarization charges is not employed consistently or unanimously in the nitride literature, it has been necessary to develop the theory from fundamental electrostatics to reconcile the issue.

Chapter 4 is the essence of this work in terms of the static performance of AlGaN/GaN HBTs. The short histories of experimental and theoretical publications on these device are reviewed. This work attempts to point out the bottlenecks of carrier transport, as well as the temperature-dependence of current gain. A comprehensive analytical model seems to be more suitable for this purpose, rather than using numerical, drift-diffusion simulators. Polarization surface charges at the hetero-interface of abrupt HBTs, and volumetric spontaneous charge in the graded layers are unconventional features of AlGaN/GaN HBTs. Also, high-level injection in the base occurs at relatively low current densities in these devices, due to the low ionization of dopants in the base. In order to study the effect of these two last features, we rely on the results of a commercial drift-diffusion simulator.

Chapter 5 is primarily devoted to the applicability of traditional signal-delay compact expressions in HBTs. Although compact expressions have been long used to estimate the cut-off frequency of these transistors, a regional comparison of analytical and numerical results has not been fully accomplished, even for conventional HBTs. Regionalizing the device into designated space-charge and quasi-neutral regions is a major issue in terms of interpreting the numerical results. Meanwhile, not all the subtle modeling features, like band-gap narrowing, are available for nitrides.

Therefore, we have performed the analytical versus numerical comparisons for the mature AlGaAs/GaAs HBTs with respect to a proposed partitioning scheme, called the integrated charge method. On the other hand, due to high-level injection and incomplete ionization in the base, the picture is more complicated in AlGaN/GaN HBTs, aspects of which are discussed in the last section of this thesis.

Finally, Appendix A reviews the wurtzite structure in detail, since understanding the crystallographic nomenclature is essential in developing Chapter 3.

Chapter 2

Electrical Properties of AlN and GaN

2.1 Overview

Optical properties of nitride materials have been largely studied in the last decade, primarily due to the growing interest in short-wavelength, light-emitting diode and laser diode applications. There has also been a substantial publication of the electrical properties of these wide band-gap materials, due to their emerging application in HFETs, as well as HBTs. However, as shown here, not all the essential parameters of AlN and GaN are fully studied yet. In addition, not all the available parameters in the literature, are in agreement. Herein, it is aimed to present a collection of the most recent electrical parameters of wurtzite AlN and GaN available in the literature, which are essential for modeling HBTs by means of analytic expressions, as well as by drift-diffusion simulators. It should be emphasized that, throughout this work, all the parameters of the ternary alloy $Al_xGa_{1-x}N$ have been assumed to be given by a linear interpolation between the corresponding values of AlN and GaN.

2.2 Energy Band-gap and Affinity

Band-gap is usually among the first parameters studied in a new material. In the case of nitrides, it does not have the same value for wurtzite and zincblende crystals, and, as discussed in Chap. 1, is typically smaller in the zincblende polytypes. In the wurtzite crystals of our interest, most of the reported values are usually at room temperature or lower. Since the nitride transistors are intended to operate at high temperatures, one needs a temperature-dependent analytic expression for simulation purposes. There is a universal model conventionally applied for semiconductor materials, which has been employed for the first time in nitride literature by Monemar [28],

$$E_g(T) = E_0 - \alpha_g T^2/(T + T_g), \tag{2.1}$$

where E_g is the band-gap of GaN in eV, as a function of temperature, T is the temperature in Kelvin, and the other symbols are fitting parameters. However, Monemar's original fitting parameters are invalid for $T > 300$ K, and have been carelessly employed by some workers [33]. Herein, following the discussion in the review of Mohammad *et al.* [24], we use in our modeling work their most recent parameters valid for T as high as 600 K, *i.e.*, $E_0 = 3.509$ eV, $\alpha_g = 7.32 \times 10^{-4}$ eV K^{-1}, and $T_g = 700$ K. For AlN, no study of the high-temperature dependence of band-gap exists, to the best of our knowledge. However, in Ref. [33], Monemar's expression is employed with $E_0 = 6.118$ eV, $\alpha_g = 1.799 \times 10^{-3}$ eV K^{-1}, and $T_g = 1462$ K. Although it is questionable whether these are valid at elevated temperatures, we have utilized them, due to the lack of more promising data.

Regarding the nonlinear mole-fraction dependence of the band-gap of ternary alloys, it is a common practice to take into account a quadratic term, with an appro-

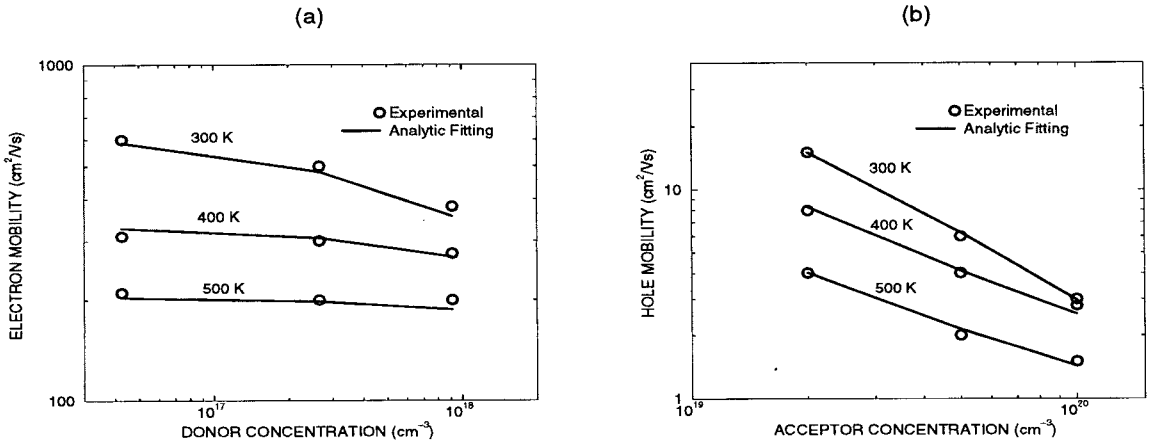

Figure 2.1: Majority carrier mobility versus doping concentration for 300, 400, and 500 K, with experimental data from Ref. [34] (circles), and corresponding analytic fittings in Eqns. (2.3), and (2.4) (solid line): (a) μ_e^n in n-type GaN. (b) μ_h^p in p-type GaN.

priate bowing factor. Nevertheless, in this work we have assumed that the bowing factor, b, is zero rather than $b \approx 1.0$ eV in $Al_xGa_{1-x}N$, following the discussion in Ref. [24], in which the controversy is reconciled.

There is no widely accepted model for the electron affinity of nitrides. The related ratio $|\Delta E_C/\Delta E_g|$ (ΔE_C is the conduction band offset, and ΔE_g is the difference between the band-gaps) has been reported in the $0.68 - 0.80$ range in the literature [24], [32], [33], and some more citations have been summarized in Ambacher's review article [12]. We have adopted the mean value of the mole-fraction-dependent affinity, fitted in Ref. [27]:

$$\chi = 4.1 - 1.87x(eV), \tag{2.2}$$

where χ stands for the affinity of AlGaN with mole fraction x.

2.3 Mobility

There have been considerable efforts to study the mobility of the nitrides in the past two decades, and there is a significant amount of scattered data on the subject. However, many of the old results are based on the immature crystal growth techniques, where there are lots of defects in the lattice structure, degrading the mobility of carriers. On the other hand, there are few publications where both the temperature- and the doping-dependence of the mobility has been measured or predicted. In addition, no minority carrier mobility study has been done to the best of our knowledge, which is actually crucial in estimating the diffusivity of electrons in Mg-doped, p-type bases of bipolar transistors. Also, most of the publications are on GaN and the reports on AlN are very limited. Chin *et al.* [30] have done a theoretical temperature- and carrier-concentration dependent study of electron majority-carrier drift and Hall mobilities in GaN, InN and AlN; Mohammad and Morkoç [31] have done a doping-dependent analytic fit to these results at room temperature. Götz and co-workers have done some of the most recent and complete measurements of the Hall effect, minority-carrier mobility for both electrons and holes, which are doping and temperature dependent [18], [34]. The results for electrons have been confirmed by one of the most recent theoretical calculations [35].

Therefore, we have picked the results in Ref. [34] as the basis of an analytic fit to obtain convenient expressions for device simulation purposes. The one corresponding to electrons in Si-doped n-type GaN is

$$\mu_e^n = 85(T/300)^{-2.2} + \frac{545(T/300)^{-2.2}}{1+(T/300)^{-4.3}(N_D/(9\times 10^{17}))^{0.8}}(cm^2/Vs), T > 300K, \quad (2.3)$$

where N_D is the donor concentration in cm^{-3}. The expression resembles the original well-known work on silicon [36]. A similar analytic relationship does not fit the hole

mobility in Mg-doped p-type GaN. However, one can fit the experimental curves with a slight modification to the form of the empirical expression used in Eqn. (2.3), *i.e.*,

$$\mu_h^p = (T/300)^{0.020T-4\times10^{-5}T^2} + \frac{28(T/300)^{-0.008T+4\times10^{-6}T^2}}{1+(N_A/(2\times10^{19}))^{1.6}}(cm^2/Vs), T > 300K, \quad (2.4)$$

where N_A is the acceptor concentration in cm^{-3} . The experimental values of Ref. [34], and the analytic fittings in Eqns. (2.3) and (2.4) are shown in Figs. 2.1(a) and (b).

The reasons for the non-applicability of the rather conventional form of Eqn. (2.3) for the holes could be two-fold. Firstly, since the ground-state energy level of magnesium is deep inside the band-gap, as discussed later, we do not have complete ionization of acceptors, and therefore the hole concentration is significantly smaller than the doping density. Secondly, due to the fact that GaN films grown in Ref. [34] were unintentionally n-type, *i.e.*, they had a non-negligible background donor doping, presumably one should also take this effect into account. It seems likely that one can fit the hole mobility with the addition of two expressions similar to Eqn. (2.3), one corresponding to the Mg concentration, and the other to the background doping. It has not been exercised here though, since the data in Ref. [34] was not extensive enough to evaluate so many fitting parameters in such a bulky expression. The hole mobility values and range are typically small ($\sim$1-$\sim$10 $cm^2(Vs)^{-1}$) [34], [38], so their precise characterization is not too important.

It must be noted that Eqns. (2.3) and (2.4) give the Hall-effect mobilities rather than the drift mobilities of interest in device engineering. It is a common practice to assume that they are equal, more specifically, the Hall factors are isotropic, temperature independent, and of unity value [18]. This has been assumed here, and is reasonably valid, according to the result of recent Monte Carlo simulations, which yield an approximately constant Hall factor of 1.2 for electrons in n-type GaN at

temperatures above 300 K [37].

Ref. [30] seems to be the only available work on the mobility of electrons in AlN. The authors have only evaluated the phonon-limited drift mobility in intrinsic AlN as a function of temperature. Therefore, we have not used their data, but basically employed Eqn. (2.3) for the $Al_xGa_{1-x}N$ emitter too, while employing the emitter donor concentration in Eqn. (2.3). Another assumption in this work is the use of Eqns. (2.3) and (2.4) for the minority electron and hole mobilities, μ_e^p and μ_h^n, with the corresponding majority carrier doping concentrations in different regions of the device.

It is worthwhile to conclude the section on mobility with a note on the saturation and peak velocities, as well as the critical field. GaN has a GaAs-like drift velocity versus electric field profile, rather than a Si-like one, *i.e.*, it has a peak velocity at a critical field. Ref. [24] gives a prediction of about 3×10^7 cm s^{-1} for the saturation velocity of GaN, v_{sat}. However it does not give an estimate of the peak velocity, v_{peak}. Khan *et al.* [39] estimate $v_{sat} = 1.5 \times 10^7$ cm s^{-1}, and $v_{peak} = 2.7 \times 10^7$ cm s^{-1}. In the present work, we rely on the more recent values of Ref. [40], *i.e.*, $v_{sat} = 2.5 \times 10^7$ cm s^{-1}, and $v_{peak} = 3.1 \times 10^7$ cm s^{-1}. Regarding the electric field dependence of electron mobility, μ_e^E, a GaAs-like dependence has been assumed [41],

$$\mu_e^E = \frac{\mu_e^n + \frac{v_{sat}}{E}(\frac{E}{E_c})^4}{1 + (\frac{E}{E_c})^4}, \tag{2.5}$$

where μ_e^n is the field-free mobility taken from Eqn. (2.3), and E_c is the critical electric field, *i.e.*, the electric field that gives the peak velocity ($v_{peak} = \mu_e^{E_c} E_c$). Thus, with respect to the above values for v_{peak} and v_{sat}, the critical field turns out to be $E_c = 3.7 \times 10^7 / \mu_e^n$ (V cm^{-1}). The electric field dependence of μ_e^n is important in the collection of electrons at the base-collector junction. A similar expression to

Eqn. (2.5) is used for holes, but in this case E_c is so high that a Si-like mobility-field relationship effectively exists. Meanwhile, the *breakdown* critical field is about 2 MV cm^{-1} [21]. This is required for studying the breakdown of the base-collector junction due to the avalanche process, but has not been implemented in this work. All the results in the following chapters are for such low collector-emitter voltages, that this nonlinear effect is not relevant.

2.4 Dopant Ionization Energies

Silicon has become the conventional donor dopant, and magnesium is the best candidate for the acceptor dopant of nitride materials at the present date, as discussed in Sec. 1.1.1. These impurities are employed in the recently fabricated nitride bipolar transistors [21], [22], [23], [87]. Different authors have reported different ionization or activation energies. Ref. [34] gives a range between 0.170-0.208 eV for the acceptor ionization level in GaN, $E_A - E_V$, and 12-16 meV for the Si donor ionization energy, $E_C - E_D$, where E_V and E_C are the valence and conduction band energy levels, respectively . Kozodoy *et al.* [32] report 0.230 eV for the former quantity in both GaN and AlGaN. Suzuki *et al.* [42] have done a mole-fraction-dependent study of $E_A - E_V$, and evaluate 0.250 eV for $x = 0.15$, and a linear variation to 0.220 eV at $x = 0.09$. In this work, we have picked 16 meV for $E_C - E_D$, and assumed that $E_A - E_V = 0.175 + 0.5x$ (eV), which is a linear interpolation of the above data from Ref. [42].

2.5 Density of States Effective Masses

There are scattered reports on the band-structure effective masses of electrons and holes in the nitrides [30], [33], [34]. Although the longitudinal masses are necessary for calculating the Richardson velocities and evaluating the tunneling analytic expressions of abrupt HBTs, the electron and hole density-of-state effective masses, m_e^* and m_h^*, are more important in device modeling, since the effective density of states is needed for estimating key parameters, such as carrier concentration and built-in potential. Thus, we have calculated m_e^* and m_h^* for GaN and AlN with respect to the longitudinal and transverse masses given in Refs. [44], [48], [45], and [46], through the standard equations

$$m_e^* = q(m_{el}m_{et}^2)^{1/3} \tag{2.6}$$

$$m_{hh} = (m_{hhl}m_{hht}^2)^{1/3}, m_{lh} = (m_{lhl}m_{lht}^2)^{1/3}, m_h^* = (m_{lh}^{3/2} + m_{hh}^{3/2})^{2/3}, \tag{2.7}$$

where m_{el} and m_{et} are the electron longitudinal and transverse masses, respectively; m_{hh} and m_{lh} are the heavy and light hole masses in terms of the corresponding longitudinal and transverse ones ($m_{hhl}, m_{hht}, m_{lhl}$, and m_{lht}); q is the number of equivalent valleys in the conduction band, where, following the discussion in Ref. [2] (Chap. 6, p. 243), a unity value for wurtzite structures has been assumed. The results are summarized in Tables 2.1 and 2.2. As can be seen, there is a large discrepancy between the estimates of Refs. [44] and [48], with the one predicted by Ref. [45] for m_h^* in the GaN case. Herein, we have chosen the average of the very close estimates of Refs. [44] and [48], *i.e.*, $m_h^* = 2.00m_0$ and $m_e^* = 0.19m_0$ (m_0 is the free electron mass) in GaN. $1.42m_0$ and $0.32m_0$ have been assumed for the corresponding masses in AlN, as shown in the tables.

Table 2.1: Hole density-of-state effective masses in units of free electron mass.

Material	m_{hhl}	m_{hht}	m_{hh}	m_{lhl}	m_{lht}	m_{lh}	m_h^*
GaN	2.04 [44]	1.81 [44]	1.88	2.04 [44]	0.19 [44]	0.42	2.01
GaN	1.96 [48]	1.87 [48]	1.90	1.96 [48]	0.14 [48]	0.34	1.99
GaN	2.09 [45]	0.37 [45]	0.66	0.74 [45]	0.39 [45]	0.48	0.91
AlN	3.52 [44]	0.73 [44]	1.23	—	—	0.471 [46]	1.42

Table 2.2: Electron density-of-state effective masses in units of free electron mass.

Material	m_{el}	m_{et}	m_e^*
GaN	0.23 [44]	0.19 [44]	0.20
GaN	0.19 [48]	0.17 [48]	0.18
AlN	0.33 [45]	0.32 [45]	0.32

2.6 Recombination Lifetime

Regarding the minority carrier lifetime, τ_n, we have followed the discussion in Ref. [47], and assumed that radiative recombination is the dominant process. However, the value of the direct recombination probability, B, is quite controversial, as shown in Fig. 2.2. Dmirtriev *et al.* [50] have done a quantum-mechanical theoretical calculation, which is very close to the one performed here through Hall's formula [51],

$$B = 0.58 \times 10^{-12} n \Big(\frac{m_0}{m_e^* + m_h^*}\Big)^{3/2} \Big(1 + \frac{m_0}{m_e^*} + \frac{m_0}{m_h^*}\Big) \Big(\frac{300}{T}\Big)^{3/2} E_g^2, (cm^3/s). \qquad (2.8)$$

E_g and T are in eV and K respectively, and n is the refractive index. The same figure also presents our semi-empirical calculation of B through the well-known detailed-balance analysis of van Roosbroeck and Shockley [52]. In both cases, the experimental values of n, and the absorption coefficient, α, were borrowed from Ref. [47]. Nonetheless, all the mentioned calculations are optimistic regarding the experimental value

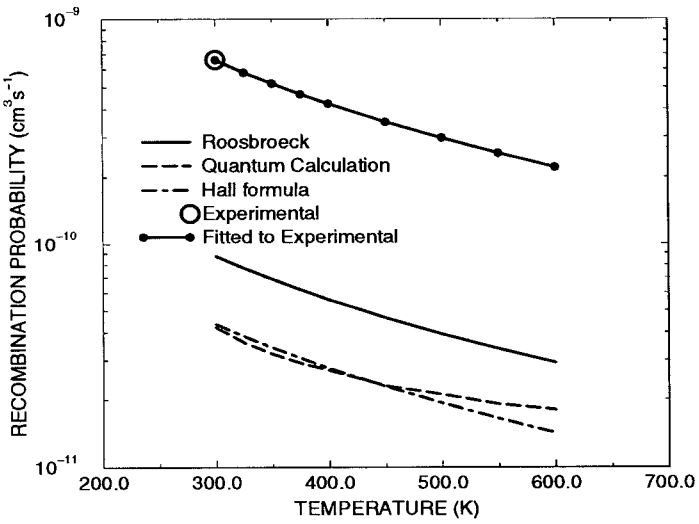

Figure 2.2: B versus T; (a) quantum-mechanical calculation from Ref. [50] (dashed line); (b) Hall formula (dot-dashed line); (c) van Roosbroeck and Shockley integral (solid line); (d) experimental value at 300 K [53] (circle); (e) the theoretical values in (c) normalized to the experimental value in (d) (solid-circled line).

of $B = 6.6 \times 10^{-10}$ cm^{3}s^{-1} measured at 300 K [53]. Therefore, in the following chapters, we have used the van Roosbroeck and Shockley temperature-dependent results, normalized to 6.6×10^{-10} cm^{3}s^{-1}. As Fig. 2.2 shows, B decreases with temperature, giving 3.5×10^{-10} cm^{3}s^{-1} at 450 K, and 2.2×10^{-10} cm^{3}s^{-1} at 600 K, for instance. We will see that B is a crucial parameter in predicting the gain of GaN bipolar transistors.

2.7 Other Parameters

The dielectric constant, ϵ_r, is essential for solving Poisson's equation. The reported values range from 8.5-9.0 for AlN, and from 9.0-9.5 for GaN [29], [32], [33], [12]. Here we use $\epsilon_r^{GaN} = 9.5$, and $\epsilon_r^{AlN} = 8.5$ [33] . Ref. [34] assumes a value of two for both the acceptor- and donor-site degeneracy factors g_A, and g_D. Scrutinizing the recent

band-diagram calculation of GaN [48], reveals that there is actually a degeneracy of two in the valence band of GaN, and $g_A^{GaN} = 4$ seems more appropriate. In spite of the fact that there is a small split in the light and heavy valence bands of AlN [43], since one can visualize that for the mole fractions of interest, $Al_xGa_{1-x}N$ is more inclined towards the characteristics of GaN, we neglect the slight band split, and have assumed the same values of GaN for AlN, *i.e.*, $g_A^{AlN} = 4$ and $g_D^{AlN} = 2$. For the case of spontaneous polarization, the calculated values of Ref. [49] have been implemented in the simulations, *i.e.*, $P_{SP}^{GaN} = -0.029$ C m^{-2}, and $P_{SP}^{AlN} = -0.081$ C m^{-2}.

2.8 Summary

A collection of the electrical properties of wurtzite GaN and AlN semiconductors required for modeling AlGaN/GaN HBTs has been presented. Many parameters are still controversial, or not available. The guideline through this work has been to be as up-to-date, and consistent as possible. For a few cases, some new calculations and/or interpolations have been done. For the case of majority-carrier mobility , a temperature- and doping-dependent analytic fit to experimental results was achieved, which should be helpful and convenient for device modeling purposes.

Chapter 3

Polarization in Nitrides

3.1 Overview

As discussed in Appendix A, wurtzite nitrides are polarized materials. The spontaneous polarization is large, and is only 3-5 times smaller than in pyroelectric materials like the perovskites. Due to the large lattice mismatch and piezoelectric constants, piezoelectric polarization is also significant and of the same order of magnitude as the spontaneous polarization in strained structures. Therefore, it is important to model these features, that are not apparent or very weak in conventional compound semiconductors like GaAs, in any electronic or opto-electronic device based on nitrides. The corresponding charges can enormously influence the band diagrams, and the carrier distributions in the devices. There are some experimental and theoretical publications regarding the effect of these charges on quantum wells and HFETs, however, the effect on the performance of HBTs has not been studied yet. The available theories and published parameters appear to be adequate to be transferred into an appropriate framework to model these phenomena in a drift-diffusion simulator. Nevertheless, there are some ambiguities in the literature, the resolving of which is

the main objective of this chapter.

Sec. 3.2 gives a synopsis of polarization in nitrides. Sec. 3.3 is a formal presentation of the theory of polarized materials with pyroelectric and piezoelectric effects. The sign conventions of polarization in nitrides is presented in Sec. 3.4, while the ambiguities in the literature are discussed in Sec. 3.5.

3.2 Synopsis of Polarization in Nitride HBTs

The lattice mismatch between AlN and GaN in the wurtzite structure is relatively large. Numerically the edge length a of the basal hexagon is 2.4% smaller in AlN. Therefore, unrelaxed heterostructures are strongly strained. On top of that, the various piezoelectric constants are at least 5 to 10 times larger in nitrides than in conventional III-V compounds, and resemble those found in typical piezoelectric materials, such as the II-VI oxides [49]. Consequently, the piezoelectric effect is important in strained layers of nitrides. The lattice misfit issue in the growth of heterostructural epitaxies is as old as the technology itself. It is sometimes possible to cancel it through certain mole fractions, *e.g.*, in $InP/In_{0.53}Ga_{0.47}P$ quantum wells. It is also possible to grow *pseudomorphic* films in which the perseverance of the lattice constants of the lower unstrained layer in the plane parallel to the hetero-interface is maintained by compressive or tensile strain in the upper layer, provided that the latter layer is thinner than a critical thickness. Otherwise, relaxation causes the formation of dislocations and defects in the lattice, which degrade the transport of carriers. In AlGaAs/GaAs HBTs, there is a small lattice mismatch between AlAs and GaAs (the lattice constant is 0.78% larger in the former). Thus, strain is not a concern in these devices, and quite thick pseudomorphic layers of $Al_xGa_{1-x}As$ with rather high mole

fractions can be grown without significant strain in the lattice. The lattice misfit is quite large in $Si_{1-x}Ge_x$ HBTs (the lattice constant of germanium is over 4% larger than that of silicon). However, due to the fact that the base thickness is typically smaller than the critical thickness, it is feasible to grow strained base layers. To give another example of a pseudomorphically grown, strained-base transistor, we can mention an InGaP/GaAs HBT which has been fabricated with a strained $In_xGa_{1-x}As$ base [54].

Piezoelectric charges are important in AlGaN/GaN quantum wells, where they can cause shifts in the photoluminescence spectra[1] [55]. The piezoelectric charge is also apparent and influences the two dimensional electron gas concentration in HFETs [17]. In these cases, at least one of the layers is less than the critical thickness for strain relaxation. In the typical $Al_xGa_{1-x}N$/GaN HBTs studied here with mole-fraction $x \leq 0.3$, the thickness of the AlGaN emitter layer, grown on GaN layers of the base and the collector, is 0.5 μm (see Table 5.5). This value is much greater than the critical thicknesses determined from studies on GaN/AlN/GaN structures (0.003 μm [57], [58], [59][2]) and $Al_{0.1}Ga_{0.9}N$/GaN superlattices (0.004-0.06 μm [10]). Therefore, the piezoelectric effect does not seem to be relevant in HBTs. Nevertheless, the following discussion considers piezoelectric polarization charges, for the purpose of the comprehensiveness of the subject. It is worthy of note that the effect of dislocations due to relaxation in the emitter region *might* be important in interpreting the high diode-ideality factor in the Gummel plots of some fabricated devices [21]. Similar diode-ideality factors of greater than 2 for the base current in a reliability study of strained-base AlGaAs/GaAs HBTs have been related to defect formation

[1]This is on top of the known effect of the strain on the band diagrams of semiconductors [56], which is a bulk property.

[2]This reports the critical thickness of GaN layer grown pseudomorphically on AlN.

near the junction after a time-dependent relaxation at high currents [60]. The same phenomenon is likely to happen in the dislocated lattice on the emitter side of the base-emitter space-charge region in AlGaN/GaN HBTs.

Due to the lack of inversion symmetry between the growth directions of interest ([0001] and [000$\bar{1}$]), the nitrides have a nonzero polarization in those axes in a zero electric field state, known as spontaneous polarization. Efforts on the measurement of piezoelectric constants can be traced back to 1973 (References 27 and 30 in Ref. [12]). Nevertheless, it is not feasible to directly measure the spontaneous polarization in bulk materials, since the induced charges at the surfaces are compensated by ambient charges. There are, however, reports on the induced voltage with temperature variation [58], known as the pyroelectric effect. The only published values on spontaneous polarization are in the recent theoretical study of both spontaneous polarization and piezoelectric constants by Bernardini *et al.* [49], based on the elegant quantum-mechanical theory of polarization in solids by King-Smith and Vanderbilt [61]. The corresponding magnitudes show that they are quite large, and only 3-5 times less than in typical perovskite pyroelectrics. The spontaneous and piezoelectric polarizations appear to have the same orders of absolute value in the various nitrides [12], [62]. Thus, lots of previous calculations on the effect of polarization in nitride structures can be ignored, whenever the spontaneous polarization has not been taken into account, *e.g.*, due to the lack of any estimate for the quantity [63].

The following section attempts to present an exact formal treatment of intrinsic polarization. It is important to be careful in the nomenclature, as well as in the mathematical vector notations and signs employed, since the material in the literature is often ambiguous, confusing, and occasionally conflicting.

3.3 Electrostatics of Polarized Materials

In the classical theory of electromagnetism, the *electric* polarization, $\mathbf{P}_E$, in response to the macroscopic electric field, $\mathbf{E}$, present in the dielectric is expressed in terms of the *electric susceptibility*, $\chi(E)$,

$$\mathbf{P}_E = \chi(E)\mathbf{E}. \tag{3.1}$$

It is evident that the polarization is in the same direction of the field ($\chi(E) > 0$), and vanishes if $\mathbf{E} = \mathbf{0}$. The scalar χ is usually a constant in ordinary dielectrics at low fields and frequencies. These are called *linear dielectrics* and follow the same back and forth path in the P_E versus E plots, when the applied field is varied. However, there are certain crystals, such as $BaTiO_3$ (barium titanate), that are *intrinsically* polarized in a zero-field state. This is due to the fact that the center of the positive and negative charges in the primitive cell are not at the same point. In other words, they do not have inversion symmetry in a certain crystallographic direction, called the *polar axis.* These dielectrics are called *electrets.* It is not possible to measure this intrinsic polarization directly, since it is neutralized by the charges in the ambient in steady state. In spite of this, the value of the polarization changes with strain and temperature. The variation with strain causes the *piezoelectric effect* (typical materials are II-VI oxides, *e.g.*, ZnO) and the variation with temperature is known as the *pyroelectric effect* (typical ones are pervoksites, *e.g.*, lithium niobate). When the value of the intrinsic polarization is varied in investigating either of these phenomena, the stray charges cannot respond instantaneously, and the induced voltage in response to the corresponding variation can be measured. There is also another variational effect, which is the nonlinear response to an external electric field in certain types of electrets. These materials have hysteresis loops similar to ferromagnetic

materials, and are therefore called *ferroelectrics*. The zero-field value of polarization indicated by extrapolation from the high-field, saturation regime is called the *spontaneous polarization*, which is typically very close to the polarization at that point. Obviously the polarity of this spontaneous polarization depends on the direction of the previously applied electric field. Both pyroelectricity and ferroelectricity disappear at a certain temperature called the *transition temperature* or *Curie point*. The precise distinction between the pyroelectric and ferroelectric materials is beyond the scope of this work and can be found elsewhere [64], but it is worthy of note that these phenomena usually occur in insulators.

The wurtzite nitrides of interest here are semiconductors and have a non-zero polarization in a field-free state, but do not show ferroelectric characteristics. This polarization is really an intrinsic polarization, but is usually referred to as spontaneous. The magnitude of the spontaneous (intrinsic) polarization, $\mathbf{P}_{SP}$, is quite large, as mentioned previously. The polar axis is the c-axis; however, in which crystallographic direction ([0001] or [000$\bar{1}$]) it is oriented is a matter to be discussed later. Piezoelectric polarization is also apparent and strong in nitrides. The details of the relationship between the stress, strain, piezoelectric and polarization tensors are thoroughly dealt with by Nye [65], and are not elaborated upon here. Briefly, in Voigt notation, the *first-rank polarization tensor*, P_i^{PE}, can be related to either the *second-rank stress tensor*, σ_{ij}, through the *third rank piezoelectric moduli tensor*, d_{ijk}, or equivalently (by means of Hooke's law) to the *second-rank strain tensor*, ϵ_{ij}. The latter is more suitable for our case, and since, due to the symmetries of the wurtzite structure, only 3 components of the 27 elements of d_{ijk} are independent, a simpler *matrix notation* is more common in the nitride literature [12],

$$\mathbf{P}_{PE} = \Big(e_{31}(\epsilon_x + \epsilon_y) + e_{33}\epsilon_z\Big)\mathbf{p} = 2\frac{a - a_0}{a_0}\Big(e_{31} - e_{33}\frac{C_{13}}{C_{33}}\Big)\mathbf{p}, \tag{3.2}$$

where a_0 is the edge length of the top (secondly grown) layer in an unstrained condition, which is presently under strain with the edge length of a. For a pseudomorphically grown layer, a is presumably equal to the corresponding edge length of the bottom unstrained layer. $\epsilon_x = \epsilon_y = (a - a_0)/a_0$ are the strains in the plane of the interface, and $\epsilon_z = (c - c_0)/c_0$ is the strain in the direction of the polar axis with unit vector $\mathbf{p}$, parallel or anti-parallel (as discussed later) to $\mathbf{z}$ in the (x, y, z) fixed coordinates, where $\mathbf{z}$ is in the direction of the growth from "bottom" to "top". c_0 is the height of the hexagonal prism in the top layer in an unstrained condition, and c is the corresponding figure in the strained condition. In contrast to a, c is not necessarily equal to the corresponding height in the lower unstrained layer. However, ϵ_z is related to ϵ_x through the elastic constants C_{13} and C_{33}, as can be seen in Eqn. (3.2). e_{31} and e_{33} are the piezoelectric coefficients in the matrix notation related to the independent terms of d_{ijk} through Hooke's law [3].

Consequently, in the electrostatics of nitrides, all the i components of the total polarization vector $\mathbf{P}$ must be added together:

$$\mathbf{P} = \sum_i \mathbf{P}_i = \mathbf{P}_E + \mathbf{P}_{SP} + \mathbf{P}_{PE}. \tag{3.3}$$

This alters the conventional relationship between the *electric displacement* and the field ($\mathbf{D} = \epsilon\mathbf{E}$) to,

$$\mathbf{D} = \epsilon_0\mathbf{E} + \mathbf{P} = (\epsilon_0 + \chi)\mathbf{E} + \mathbf{P}_{SP} + \mathbf{P}_{PE} = \epsilon\mathbf{E} + \mathbf{P}_{SP} + \mathbf{P}_{PE}, \tag{3.4}$$

where ϵ and ϵ_0 are the *permittivities* of the electrically polarized media and vacuum respectively. It is evident from the last equation that one can imagine a situation where there is a non-zero displacement, even though $\mathbf{E} = \mathbf{0}$. An example is a speci-

[3]The third independent coefficient, e_{15}, is due to the shear strain, and is not discussed in nitride literature.

men placed between two conducting plates at the same potential. The electric field vanishes in the material; however, $\mathbf{D} \neq \mathbf{0}$ in the media due to the *free* surface charge densities on the plates with the absolute values of $|P_{SP} + P_{PE}|$. On the other hand, we can have a situation where $\mathbf{E} \neq \mathbf{0}$, even though $\mathbf{D} = \mathbf{0}$. One should be aware that the macroscopic electric field, $\mathbf{E}$, is in response to all the real charges in the media (either free or polarized), and is not solely due to the external applied voltage. Thus a non-zero electric field is present in the material in an electrically isolated ambient due to the polarized charges in such a way as to satisfy $\mathbf{D} = \mathbf{0}$ (see Eqn. (3.11)). In this imaginary experiment, *polarized* surface charge densities of $\pm(P_{SP} + P_{PE})/\epsilon_r$ are induced at the interfaces, where $\epsilon_r = \epsilon/\epsilon_0$ is the relative permittivity of the dielectric.

Any of the $\mathbf{P}_i$ components in Eqn. (3.3) (as well as the total $\mathbf{P}$) can be mathematically attributed to a *polarization volume charge density*, $\rho_{p,i}$, in the bulk material and a *polarization surface charge density*, $\sigma_{p,i}$, at the surface of the material [66] (Chap. 4, p. 72),

$$\rho_{p,i} = -\nabla.\mathbf{P}_i, \tag{3.5}$$

$$\sigma_{p,i} = \mathbf{P}_i.\mathbf{n}, \tag{3.6}$$

where $\mathbf{n}$ is the *outward* normal vector at each point of the surface element of the dielectric (see Fig. 3.1). This last effective charge formulation is quite useful in determining the necessary modifications that need to be made to conventional numerical simulators used in semiconductor device work to take into account the presence of polarization charges. With regard to Gauss's law, we can always link $\mathbf{E}$ to all the contributing charges (either free or polarized) present in the media,

$$\epsilon_0 \nabla.\mathbf{E} = q(N_D^+ - N_A^- + p - n) + \rho_E + \rho_{SP} + \rho_{PE}, \tag{3.7}$$

where q is the electron charge, N_D^+ and N_A^- are the ionized donor and acceptor densi-

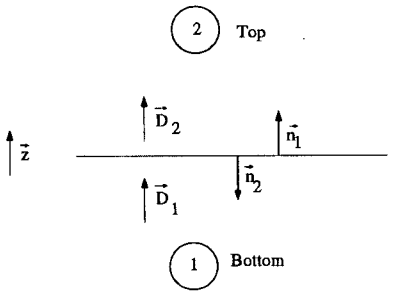

Figure 3.1: Unit vectors at the boundary of two media 1 (bottom) and 2 (top) with electric displacements.

ties respectively, and p and n are the hole and electron concentrations, respectively. We can define

$$\rho_{pol} \equiv \rho_{SP} + \rho_{PE} = -\nabla.(\mathbf{P}_{SP} + \mathbf{P}_{PE}). \tag{3.8}$$

ρ_{pol} is zero in homogeneously grown structures, *i.e.*, the polarizations are constant through the media. ρ_{SP} is non-zero in any graded layer, but ρ_{PE} is non-zero only if the graded layer is strained. On the other hand, $\rho_E = -\chi\nabla.\mathbf{E}$ can be absorbed in ϵ to obtain the *modified* Poisson's equation governing polarized semiconductors:

$$\nabla.(\epsilon\nabla\psi) = -\nabla.(\epsilon\mathbf{E}) = -q(N_D^+ - N_A^- + p - n) + \rho_{pol}, \tag{3.9}$$

where ψ is the vacuum potential. This equation can be solved with the ordinary Neumann boundary conditions:

$$\psi_1(0) = \psi_2(0), \tag{3.10}$$

$$\mathbf{D}_1(0) = \mathbf{D}_2(0), \tag{3.11}$$

where 1 and 2 refer to the two sides of the interface at $z = 0$ (see Fig. 3.1). The first boundary condition, Eqn. (3.10), is for the avoidance of an infinite electric field at the interface. The second boundary condition can be simply inferred from Eqns. (3.4), (3.5) and (3.7), if σ_{free} , the free surface charge density at the interface is zero.

This is the situation we are interested in. Recalling that $\mathbf{E}$ is always $-\nabla\psi$, according to Eqn. (3.4), Eqn. (3.11) yields:

$$\epsilon_1 d\psi_1(0)/dz - \epsilon_2 d\psi_2(0)/dz = \sigma_{pol} \equiv \sigma_{SP} + \sigma_{PE}, \tag{3.12}$$

where

$$\sigma_{SP} = (\mathbf{P}_{SP1} - \mathbf{P}_{SP2}).\mathbf{z} = \mathbf{P}_{SP1}.\mathbf{n}_1 + \mathbf{P}_{SP2}.\mathbf{n}_2, \tag{3.13}$$

$$\sigma_{PE} = (\mathbf{P}_{PE1} - \mathbf{P}_{PE2}).\mathbf{z} = \mathbf{P}_{PE1}.\mathbf{n}_1 + \mathbf{P}_{PE2}.\mathbf{n}_2, \tag{3.14}$$

The most right-hand sides of Eqns. (3.13) and (3.14) have been written in consistency with Eqn. (3.6). To have a rule of thumb it might be easier to express the above equations in terms of "bottom" and "top" layers, as defined previously in Fig. 3.1, *i.e.*,

$$\sigma_{pol} = P(bottom) - P(top), \tag{3.15}$$

where from P we mean the algebraic addition of piezoelectric and spontaneous polarizations projected in the growth direction, $\mathbf{z}$, *i.e.*, $P = (\mathbf{P}_{SP} + \mathbf{P}_{PE}).\mathbf{z}$. In a conventional emitter-up HBT configuration, $\mathbf{z}$ points from the collector to the emitter.

To summarize the section, a drift-diffusion simulator would solve Eqn. (3.9) and the drift-diffusion equations simultaneously with the boundary conditions in Eqns. (3.10) and (3.12) in a spontaneously and/or piezoelectrically polarized heterostructure. In homogeneous materials ρ_{pol} is zero, but has to be taken into account in graded layers of semiconductor devices.

3.4 Sign Conventions of Polarization in Nitrides

The previous section presented the theory required for modeling polarization effects in strained and/or intrinsically polarized semiconductors. The sign or polarity of the volume and interface charges distributed in the device can be unambiguously determined, as long as we know what the magnitudes and the directions of $\mathbf{P}_{SP}$ and $\mathbf{P}_{PE}$ are. As mentioned before, the spontaneous polarization and piezoelectric constants of AlN, GaN and InN are calculated by Bernardini *et al.* [49]. One can simply calculate the corresponding values at mole fraction x, by means of linear interpolation for the required quantities.

Observing the calculated spontaneous polarizations in Ref. [49], we can see that the values are negative. Obviously the negative sign must be with respect to a positive direction convention. This is the crystallographic [0001] orientation with the unit vector $\mathbf{c}$ [49], [68]. By convention $\mathbf{c}$ is defined from a cation (Ga for instance) atom towards the nearest neighboring N atom. These are those atoms separated by $u\mathbf{c}$ (u is the anion-cation bond length along the c-axis), as shown in Fig. A.3(b). Recall that the wurtzite structures consist of two hexagonal closed packed (hcp) structures, one with cations and the other with anions, penetrated into each other with a separation of $u\mathbf{c}$. Equivalently one can visualize that in the [0001] direction, the cation atoms are positioned on top of those {0001} cation and anion bilayers which are separated by $(c/2 - u)\mathbf{c}$. That is why the [0001] orientation is also known as Ga-faced in GaN. In our fixed $\mathbf{z}$ coordinate this means that $\mathbf{c} = \mathbf{z}$. However, different materials have different equilibrium separations in the lattice. As discussed in Appendix A, ideally u is equal to $3c/8$, since this is the ratio that annihilates the electric force at each dot-sized ion in the lattice. Because of the nonzero radii of the

ions as well as the effect of electron orbitals, the nitrides have substantially smaller u/c ratios in equilibrium positions (see Table A.1). Thus, since $u \neq c/4$ (neither in reality nor ideally) wurtzite structures do not have inversion symmetry along the **c** direction, *i.e.*, they are polar. In other words, the centers of the positive and negative charges are not at the same point in a primitive unit cell, and it happens to have a negative spontaneous polarization with respect to **c**. Flipping the crystal over in our fixed **z** direction gives the $[000\bar{1}]$ or N-faced polarity. The definition of **c** is unchanged, with regard to the anions and cations, *i.e.*, $\mathbf{c} = -\mathbf{z}$. Also the spontaneous polarization is still negative with respect to **c**. There are still arguments on how to experimentally determine the polarity of nitrides and the chemical properties of the surfaces, as well as which orientation is preferred by different growth techniques [70].

Regarding the piezoelectric polarization, the values of the piezoelectric coefficients are tabulated through a practical sign convention, codified by the IEEE [70]. The piezoelectric axis, **p**, appearing in Eqn. (3.2) is defined in a way that e_{33} is positive. Therefore, the coefficients themselves do not give any correlation between **p** and **c** axes, and it must be figured out independently. This is what is implicitly done in Ref. [49], with the result that $\mathbf{p} = \mathbf{c}$. Thus, the direction of the piezoelectric polarization is decided on the sign of the product of $(e_{31} - e_{33}(C_{13}/C_{33}))$ by ϵ_x in Eqn. (3.2). For AlGaN layers grown pseudomorphically on GaN, the layer is in tensile strain, so ϵ_x is positive. The $(e_{31} - e_{33}(C_{13}/C_{33}))$ factor is negative for any mole fraction in AlGaN layers on GaN (this is not always the case in other nitride heterostructures, *e.g.*, InN on GaN). Consequently, $\mathbf{P}_{PE}$ is in the $-\mathbf{c}$ direction, *i.e.*, in the same direction as $\mathbf{P}_{SP}$.

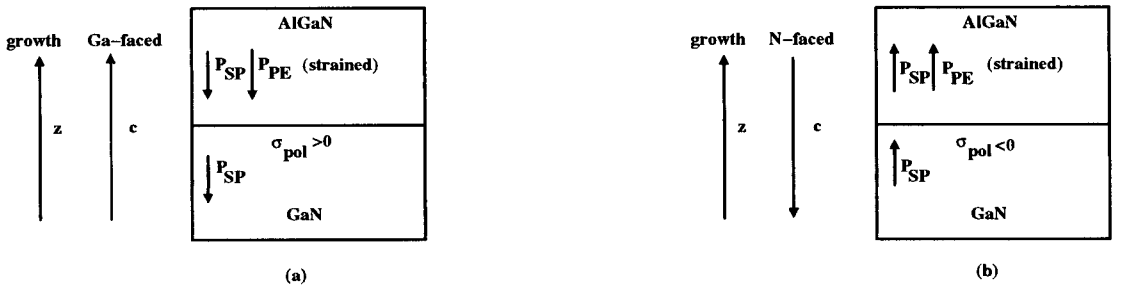

Figure 3.2: Direction of the polar axis, **c**, growth direction, **z**, spontaneous polarization, $\mathbf{P}_{SP}$, and piezoelectric polarization, $\mathbf{P}_{PE}$, in a strained AlGaN layer grown on top of a GaN layer: (a) Ga-faced ([0001]); (b) N-faced ([000$\bar{1}$]). The polarities of the interface charges, σ_{pol}, are also shown.

3.5 Discussion

One can conclude at this point that the above sign conventions are adequate for calculating the interface charges with the correct polarities, by employing Eqns. (3.15) and (3.8). These sign conventions are employed in Refs. [12], [68], [67], [69], and [62]. All of them consider the [0001] case, since it appears to be the more commonly reported direction of the growth by metal-organic chemical vapor deposition [12], [70]. Nevertheless, even though the absolute value of the charges in Eqns. (3.15) and (3.8) can be guessed intuitively, the charges sometimes appear in the literature with the correct polarity, and sometimes with the wrong polarity. This whole critical exercise has been developed in this chapter to resolve the polarity issue.

According to the values of spontaneous polarization of GaN and AlN in Sec. 2.7, we can claim that the polarization charge at the hetero-interface of a strained

AlGaN layer grown on top of a GaN layer is positive if we have [0001] or Ga-faced layers, and is negative if we have a N-faced structure. Fig. 3.2 shows the the polarization vectors, as well as the polarity of the interface charge for these two cases. However, not all the publications are in accord with this picture:

Regarding Eqn. (3.8), the minus sign has been overlooked in Refs. [17] and [67], and even sometimes the incorrect notation ∇P can be found [67], [68], [69]. Regarding Eqn. (3.15), Refs. [68], [62], and [69] are in agreement with us, but Refs. [67] and [12] write the wrong sign, all with the same $\mathbf{z}$ definition as ours. They have all presented some figures, similar to Fig. 3.2, in their publications to indicate the sign of the polarized charges induced at the interfaces. Refs. [68], [69] and [62] are in accord with Eqn. (3.15), and the figures are consistent with this assumption too. Ref. [67] writes the opposite sign for Eqn. (3.15), but through an inconsistency in the paper, the figures appear to be correct. Ref. [12] has the wrong sign of Eqn. (3.15), as well as in the corresponding figure.

Finally, as discussed previously, pyroelectricity usually vanishes at a certain temperature, called the Curie point. The spontaneous polarization constant values in Ref. [49] are for room temperature. To the best of our knowledge, the temperature dependence of the parameter has not been studied yet, however, we expect that the values would be much smaller at the elevated temperatures of interest in nitride HBTs.

3.6 Summary

- Spontaneous polarization is large in nitrides and should be appropriately modeled in AlGaN/GaN HBTs. However, due to the fact that the emitter epilayer

is typically thicker than the critical thickness, the piezoelectric effects do not seem to be relevant in such devices. On the other hand the effect of dislocations in the relaxed emitter is worthy of study regarding the poor performance of the fabricated HBTs.

- Both the spontaneous and/or piezoelectric interface charges can be calculated with justified polarities, provided that the direction of the polarizations are known. The surface and volume charges of abrupt and graded emitter-up AlGaN/GaN HBTs, respectively, are positive for Ga-faced growth, and negative for N-faced growth.

- The precise temperature dependence of the pyroelectric effect in nitrides is unknown. The spontaneous polarization is expected to be reduced at the elevated temperatures of interest in nitride HBTs.

Chapter 4

Static Performance of AlGaN/GaN HBTs

4.1 Overview

The idea of employing band-gap engineering schemes in microelectronic devices is credited to W. Shockley, the father of the transistor. The theoretical aspects of a wide-gap emitter were first studied by H. Kroemer in 1957 [72]. Beyond some scattered publications in the 1970's, due to the lack of promising fabrication techniques *heterojunction bipolar transistors* (HBT) did not emerge until the early 1980's, when Kroemer himself announced its era in an invited paper [73]. The main idea is increasing the emitter injection efficiency, by reducing the minority carrier component of emitter current, through a larger valence band barrier for the holes, a consequence of the difference between the emitter and base band-gaps[1]. As another benefit, one can increase the base doping density to achieve lower base resistance, causing a higher maximum oscillation frequency, in addition to reducing current crowding in high cur-

[1] *n-p-n* transistors are referred to throughout this work.

rent regimes. On the other hand, the emitter doping can be reduced in order to reduce the base-emitter junction capacitance. A mole-fraction-graded emitter buffer between the base and the bulk emitter can remove the abrupt barrier, which blocks the injection of electrons into the base. Band-gap engineering is not necessarily restricted to the emitter, and various structures exist. The classical paper by Kroemer [73] remains a must-read for anybody in the field. A concise review of the state-of-the-art of ultra-fast HBTs can be found in the Wiley Encyclopedia of Electrical and Electronics Engineering [74]. Group III-V device fabrication technology became so mature, and HBTs became so commercial for mobile and optical-fiber communications in the mid-1990's, that HBTs can now be viewed as conventional transistors. Complete textbooks have been devoted to these devices [75], [76].

As discussed in Sec. 1.1.2, nitride HBTs are primarily of interest for high-power, and high-temperature applications in hostile environments like furnaces, motors of automobiles, and all-electrical aircraft technology. The high-power application is a consequence of the high breakdown voltage, since the energy for ionizing collisions that create electron-hole pairs increases with the band-gap in the avalanche process. Sometimes, Johnson's figure of merit for *unipolar* devices, which is proportional to the square of the product of the breakdown critical field and the saturation velocity, is employed to argue the advantage of nitride *bipolar* transistors [77], [21]. This figure for GaN is some hundreds of times superior to that in Si [78], but does not seem to be an appropriate figure of merit, since the saturation velocity is very high in nitrides, and the current associated with it (J_{coll} in Eqn. (4.1)) does not limit the overall current in the way it does for unipolar devices. Gao and Morkoç [79] have developed a reasonable figure of merit for high power class B and C HBT amplifiers, which incorporates the breakdown critical field into the maximum available power of

these amplifiers. They conclude that the AlGaN/GaN system is 19 times superior to AlGaAs/GaAs, and 32 times better than silicon technology.

The interest in nitrides for high-temperature electronics is partly due to chemical stability and inertness at elevated temperatures, but is *not*, as sometimes is thought [77], [81], [83], due to the high thermal conductivity of nitrides. The thermal conductivity of GaN, for instance, is the same as silicon, *i.e.*, 1.3 W/cmK [43]. The figure is higher for SiC, being about 5.0 W/cmK in SiC [24]. It is due instead to the fact that a wide band-gap material has a low intrinsic carrier concentration. Any doped semiconductor is extrinsic at some moderate temperature range above the freeze-out regime, but tends to be intrinsic at adequately high temperatures, when the intrinsic concentration dominates the completely ionized impurity concentration, causing the electron and hole concentrations to be roughly equal. Under such circumstances, obviously, any transistor action relying on having three designated n, p, and n regions collapses. For instance, the electrons are no longer the minority carriers in the base, and the majority carrier concentrations are not dictated by the dopings in the base, emitter or collector. In addition, the high intrinsic concentration will cause a large electron-hole pair generation rate in the collector-base space-charge region, which will dominate the current [80]. This large current will push the device into thermal runaway. With respect to this argument, the smaller the band-gap, the lower the maximum operating temperature. For silicon the limit is about 500 K, and is conventionally avoided by cooling means. It occurs at about 1,000 K in GaAs, but devices are never operated at such high temperatures, because the thermal conductivity is low, and dissociation of the compound semiconductor occurs. The corresponding temperature for the nitrides happens to be in the range of the melting point, *i.e.*, $\sim$2,000 K. Therefore, nitride transistors can operate at very high temperatures, in

principle, without the necessity of expensive and bulky heat sinks. This also makes them suitable for satellite high-power applications, where although the temperature is not too high, large heat sinks are required due to the poor convection properties of the low-pressure atmosphere.

Sec. 4.2 reviews the short history of nitride bipolar transistors. The rest of the chapter is a novel contribution to the literature as it attempts to predict the static performance limits of AlGaN/GaN HBTs. Sec. 4.3 presents the analytic compact expressions for modeling these devices, the results of which lead to the emitter configuration study in Sec. 4.4.1. Aspects of polarization, and high level injection are studied by a drift-diffusion simulator in Sec. 4.4.2, while Sec. 4.4.3 is devoted to the high temperature issues. The high-frequency aspects are discussed in the following chapter.

4.2 History of Nitride HBTs

With respect to other compound material systems, like GaAs, the nitride HBTs are still in their infancy. The first nitride HBT was made by Pankove *et al.* [81], at the University of Colorado in 1994 , by employing the GaN/6H-SiC system, working up to 260°C, with high gains at room-temperature, and a gain record of 100 at 535°C [20], [82]. However, since 1999, Pankove's group has switched to the 4H polytype of SiC for the base, and collector materials [83], [84], [85]. They argue that the high quality, defect-free 6H-SiC layers used in [81], [20], [82], have not been reproducible, and 4H-SiC is more promising in this regard. These HBTs have a modest gain of 15 at room temperature, shrinking to 3 at 300°C.

The band-gaps of 4H-SiC and 6H-SiC are 0.2 and 0.4 eV smaller than GaN

at room temperature, respectively [83]. The AlGaN/GaN system offers higher band-gaps, with the freedom of playing with the band-gap difference via the aluminum mole fraction in the emitter. The development of AlGaN/GaN HBTs was impeded until 1998 though, because of difficulties in obtaining a highly doped p-type base. The first report of a working device is credited to Ren *et al.* [22], at the University of Florida, who achieved a current gain of 10 at 300°C, comparable to the results on the 4H-SiC structure, but with a gain of less than 3 at room temperature [86]. This unexpected behavior was attributed to temperature-dependent base resistance phenomena. The active layers were grown by MBE, but the same group has subsequently reported HBTs grown by MOCVD [86]. They also report on very high Early voltages, and the decrease of the collector-base breakdown voltage from greater than 10 to about 4 V, when the temperature is increased to 250°C from room temperature [88]. Another report also appeared in 1998 by McCarthy *et al.* [92] with the same low current gain of about 3 at room temperature, partly because of the wide 0.2 μm base, and also because of high surface and space-charge recombination, as well as poor contacts [21]. There is also a report of a GaN homojunction bipolar transistor by Yoshida and Suzuki [23], who claim current gains of 10 to 40 at 300°C, which is remarkable for a homojunction device, considering the above-mentioned reports for heterojunction devices. Finally, Shelton *et al.* [87] have achieved the fabrication of a AlGaN/GaN HBT with the encouraging gain of 100.

In terms of theoretical work, there is one publication on the base transit time of GaN/InGaN HBTs [24], which to the best of our knowledge is not a material system that anyone has been developing in terms of device fabrication. Chiu *et al.* [33] do a numerical study of the same quantity in GaN/InGaN/AlGaN, and AlGaN/GaN/AlGaN double hetero-bipolar transistors, focusing on the exit velocity

at the collector-base junction, where they predict a cut-off frequency of about 19 GHz for the latter system. For *n-p-n* AlGaN/GaN HBTs there are three simulation predictions, all appearing in 2000 [77], [89], [90]. The predicted room-temperature current gains are 15 [77], 60 [90], and 1,100 [89]. As we will see, these totally different predictions originate from the dominance of recombination in the base current, and the fact that the three groups used different values of the minority-carrier lifetime. None of the three simulation articles do an analytic study of the bottleneck for carrier transport in different structures, and neither do they consider the effect of incomplete ionization in the base, which is elaborated upon here for the first time. Also, they do not incorporate the effect of polarization in the predictions, as is done here.

As a final note, it must be stressed herein that regardless of technological aspects, the advantage of AlGaN/GaN HBTs over GaN/4H-SiC devices appears to be still controversial, since in addition to the above-mentioned higher thermal conductivity of SiC, the high radiative recombination in the GaN-based devices dominates the base current and the gain, while SiC is an indirect material with longer minority carrier lifetime [83]. In both material systems, problems exist with achieving very high *p*-type doping densities and low sheet resistances (both of which are necessary for the base material). This has led to some preliminary investigations of *p-n-p* AlGaN/GaN transistors as an alternative to the *n-p-n* configuration [91].

4.3 Compact Expressions for Collector and Base Currents

For the purpose of providing an analytic expression for the collector current density, J_C, one of the most recent and comprehensive formulations in the HBT literature

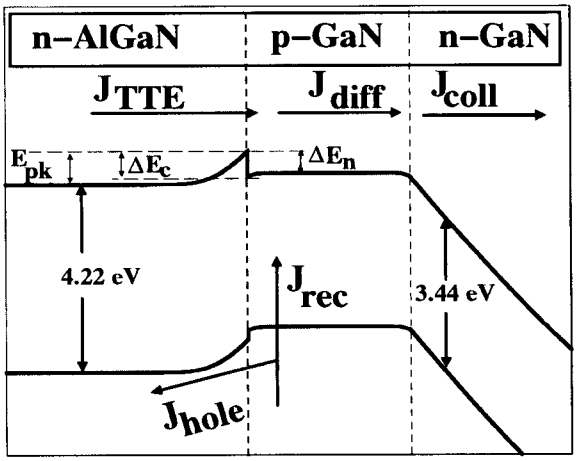

Figure 4.1: Typical band diagrams of an AlGaN/GaN abrupt HBT at 300 K, showing the three components of J_C in Eqn. (4.1), and the two components of J_B in Eqn. (4.9). Also shown are the important energies: E_{pk}, ΔE_C and ΔE_n.

is the cornerstone of our results after performing a few modifications. As depicted in Fig. 4.1, these are three possible major bottlenecks for the transport of carriers, namely: diffusion in the base, J_{diff}; tunneling and thermionic emission at the base-emitter heterojunction, J_{TTE}; and drift in the collector-base space-charge region, J_{coll}. Assuming J_C is much greater than the recombination current in the quasi-neutral base, the three components are incorporated into a *serial flow* model [93],

$$J_C = \frac{1}{\frac{1}{J_{diff}} + \frac{1}{J_{TTE}} + \frac{1}{J_{coll}}} = \frac{-q\overline{n}_{0B}e^{qV_{BE}/k_BT}}{\frac{W_B}{D_{nB}} + \frac{1}{\gamma\delta v_R} + \frac{1}{v_{coll}}} \approx J_{diff}e^{-q\Delta E_{Fn}/k_BT}, \qquad (4.1)$$

where q is the electron charge, v_R is the Richardson velocity in the base, V_{BE} is the base-emitter junction voltage, k_B is the Boltzmann constant, T is the temperature, v_{coll} is the velocity of electrons entering the collector, after diffusing across the quasi-neutral base with width W_B and field-free diffusivity D_{nB}. Typically, J_{coll} is negligible in nitride HBTs, because of the high saturation velocity. $\overline{n}_{0B}$ is the equilibrium value

of electron concentration at the edge of the quasi-neutral base. Multiplying $\overline{n}_{0B}$ by $\delta = \exp(-\Delta E_n/k_BT)$ for calculating J_{TTE} is essential as the concentration of interest is at the top of the barrier in Fig. 4.1, where ΔE_n is the energy difference between the peak of the barrier and the conduction band in the base at the emitter edge of the quasi-neutral base, *i.e.*,

$$\Delta E_n = \Delta E_C - q(1 - N_{rat})(V_{bi} - V_{BE}). \tag{4.2}$$

V_{bi} is the corresponding built-in potential, and ΔE_C is the conduction band offset. $N_{rat} = \epsilon_B N_B/(\epsilon_B N_B + \epsilon_E N_E)$, where N_E and N_B are the emitter- and base-doping concentrations respectively, with corresponding permittivities ϵ_E and ϵ_B.

γ in Eqn. (4.1) is the tunneling factor, the ratio of the total tunneling and thermionic current, J_{TTE}, over the thermionic current, J_{TE}, in the parallel flow of the two phenomena across the junction. The compact expressions in Searles *et al.* [94], have been incorporated for calculating γ in this work:

$$\gamma = \frac{J_{TTE}}{J_{TE}} = C\mathrm{e}^{\frac{E_{pk}(1-\alpha)}{k_BT}} + 1, \tag{4.3}$$

where $\alpha = \tanh U'_{00}/U'_{00}$, and $U'_{00} = \hbar q\sqrt{N_E/m_{le}\epsilon_E}/k_BT$. m_{le} is the electron longitudinal mass in the emitter, and C is another parameter defined as

$$C = \sqrt{\frac{4\pi E_{pk}U'_{00}\sinh U'_{00}}{k_BT\cosh^3 U'_{00}}}. \tag{4.4}$$

E_{pk} is the peak energy at the top of the abrupt barrier with reference to the conduction band at the emitter contact. As shown in Fig. 4.1, it is the emitter-side portion of the energy drop of the vacuum level across the base-emitter space-charge region, *i.e.*,

$$E_{pk} = qN_{rat}(V_{bi} - V_{BE}). \tag{4.5}$$

Eqn. (4.1) is primarily for an abrupt HBT, but it turns out that with appropriate values for γ, and δ, the same expression can be equivalently employed for

emitter-graded HBTs. The specifications are simply: (i) $\gamma = 1$; (ii) $\Delta E_n = 0$ or $\delta = 1$, since presumably there is no discontinuity in the conduction band. The conventional expressions for V_{bi} can be used for both graded and abrupt cases, if one bears in mind that due to the large effect of incomplete ionization of the Mg dopants in the base of nitride transistors, the hole concentration in equilibrium, $\overline{p}_{0B}$, which is conventionally assumed to be equal to N_B, should be altered to N_B^-, the ionized concentration. Simple expressions can be developed for N_B^- from standard charge-neutrality theory, [95] (Chap. 4, p. 128),

$$N_B^- = -\frac{N_\zeta}{2} + \sqrt{\left(\frac{N_\zeta}{2}\right)^2 + N_\zeta N_B}, \tag{4.6}$$

$$N_\zeta \equiv \frac{N_V}{g_A} \mathrm{e}^{(E_V - E_A)/k_B T}, \tag{4.7}$$

where N_V is the effective density of valence band states, g_A the acceptor degeneracy factor, and $E_A - E_V$ is the acceptor ionization energy, all in the base material. However, one should not argue in the same way, and change N_B in the aforementioned parameter N_{rat} to N_B^-. That is because of the band bending in the space-charge region, where the acceptor level energy will lie close to the hole quasi-Fermi level, causing near-complete ionization in that region. Comparison with the numerical solutions verified this assumption. If N_B^- is used in calculating N_{rat}, there is a large error in the analytic results, basically due to the presence of N_{rat} in the tunneling model. However, using N_B^- for estimating V_{bi} is in agreement with the simulated estimates of the quantity, and also gives the correct collector current density (see Fig. 4.6).

The most right-hand side of Eqn. (4.1) is written with the approximation that the *forward* electron flow in the quasi-neutral base is much larger than the *reverse* electron flow. $J_{diff} = -q(D_{nB}/W_B)\overline{n}_{0B}\exp(qV_{BE}/k_B T)$ is the well-known diffusion

term for homojunction transistors, where the approximations in deriving it are: (i) short-base transistors such that W_B is much greater than the electron diffusion length; (ii) strong reverse bias for the base-collector junction such that the minority electron concentration at the corresponding edge of the quasi-neutral base can be neglected compared to that at the other end of the base. Quasi-Fermi level splitting, ΔE_{Fn}, is a consequence of applying the principle of current balancing across heterojunctions [96]. This splitting brings the non-equilibrium electron concentration at the emitter side of the quasi-neutral base down to $\overline{n}_{0B}\exp(q(V_{BE} - \Delta E_{Fn})/k_BT)$ in the case of HBTs. From Eqn. 4.1, it can be simply inferred that

$$\Delta E_{Fn} = k_BT\ln\Big(1 + \frac{D_{nB}}{W_B}\big(\frac{1}{\gamma\delta v_R} + \frac{1}{v_{coll}}\big)\Big). \tag{4.8}$$

In contrast to the collector current, the base current comprises components which flow in *parallel.* As shown in Fig. 4.1, we consider current densities J_{rec} and J_{hole}, due to quasi-neutral base recombination and back-injection of holes into the emitter, respectively. The hole diffusion length, L_{pE}, is typically of the same order of magnitude as the quasi-neutral emitter width, W_E, in nitride transistors. Therefore, employing the full hyperbolic solution of the continuity equation in this region is more reasonable than the long- or short-emitter approximations. This is the first term in the following expression for the base current density, $J_B = J_{hole} + J_{rec}$, which is applicable to both the graded- and abrupt-emitter cases,

$$J_B = \frac{qD_{pE}n_{iE}^2}{L_{pE}N_E}\coth\Big(\frac{W_E}{L_{pE}}\Big)e^{q\frac{V_{BE}}{k_BT}} + \frac{qW_B\overline{n}_{0B}}{2\tau_n}e^{q\frac{V_{BE}-\Delta E_{Fn}}{k_BT}}. \tag{4.9}$$

D_{pE} is the diffusivity of holes in the emitter, n_{iE} is the intrinsic carrier concentration in the emitter, and τ_n is the minority carrier lifetime in the base. The recombination term, J_{rec}, invokes ΔE_{Fn} in a similar way as in J_{diff} in order to find the concentration of minority carriers at the edge of the quasi-neutral base in an approximately

Table 4.1: Configuration of studied AlGaN/GaN HBTs.

Region	Material	Doping (cm^{-3})	Width (μm)
Emitter	$Al_xGa_{1-x}N$	0.5-5$\times 10^{18}$	0.50
Base	GaN	1×10^{19}	0.10
Collector	GaN	5×10^{16}	0.75
Subcollector	GaN	5×10^{18}	0.75

triangular profile. Physically, the corresponding hole charge that recombines every τ_n seconds, is compensated through supply of holes from the base contact in steady state.

4.4 Results and Discussion

The baseline device of our study has been a n-$Al_xGa_{1-x}N$/p-GaN/n-GaN HBT where the metallurgical emitter, base, collector, and sub-collector widths, and doping densities are shown in Table 4.1. These figures are in accord with the fabricated device reported in Ref. [21], with the exception of the base width, which we take to be half as wide in anticipation of future improvements in the technology. Also, the emitter doping and mole fraction are kept as free parameters, and are subject of an analytic study to find the optimum configuration.

In addition to the analytical computations, numerical simulation results were obtained from the commercial drift-diffusion, finite-element solver, MEDICI [98]. The collector-emitter voltage, V_{CE}, was fixed to 10 V in all the simulations, and incomplete ionization, and Fermi-Dirac statistics were assumed. In graded-emitter HBTs, a 200 Å linear grading in the emitter has been modeled in the simulations, *i.e.*, the

ungraded emitter is 0.48 μm wide. Typically two-dimensional effects like current crowding are negligible in HBTs due to the high doping density in the base, and one-dimensional simulations seem adequate for reliable and fast results. This was verified by comparisons of simulation of the one- and two-dimensional cases. In two-dimensional simulations, 1.0×3.0, and 1.0×4.8 μm^2 planar cross-sections were assumed for the emitter and collector respectively, in an emitter-up configuration. Two symmetrical base contacts with 0.50 μm ledge widths were modeled without any surface recombination feature. With respect to the symmetry of the device geometry, a half-structure was simulated, while all the current density results were normalized to the corresponding emitter cross-section area.

The effort of this static (DC) study was focussed on the Gummel plots, *i.e.*, $\log J_C$ and $\log J_B$ versus V_{BE}, at fixed V_{CE}. Due to the high breakdown electric field, which has not been modeled herein, the above-mentioned bias is below the onset of breakdown in the collector-base junction. It should be noted that all the parameters required for either the numerical or the analytic approaches are taken from Chap. 2. If any property, *e.g.*, band-gap narrowing, is not discussed in that chapter, it means that it has not been incorporated in the models. Second order effects, such as this, are ignored due to the lack of any definite data in the nitride literature at the present time.

4.4.1 Emitter Configuration

This section is devoted to a systematic study of the influence of the emitter structure on the static characteristics of the devices. For this purpose, the analytic expressions in Sec. 4.3 are employed rather than the results of the drift-diffusion simulator. This is primarily due to the interest in gaining a deeper understanding of the various

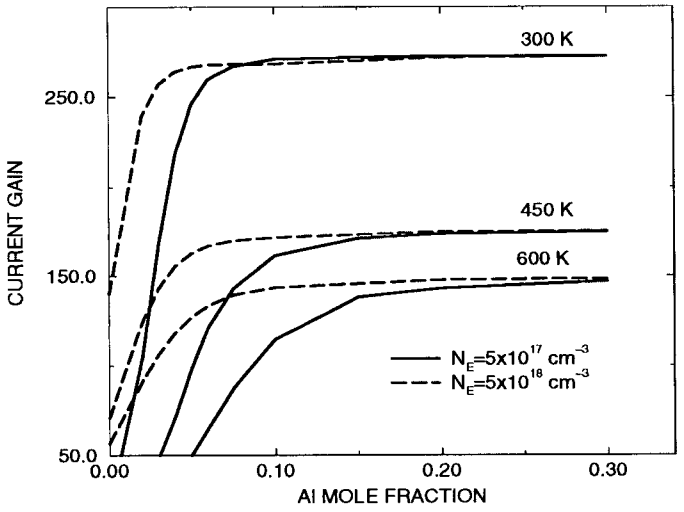

Figure 4.2: β versus x, for 300, 450, and 600 K, calculated analytically through Eqns. (4.1), and (4.9) for two abrupt HBTs with $N_E = 5 \times 10^{17}$ cm^{-3} (solid line), and $N_E = 5 \times 10^{18}$ cm^{-3} (dashed line).

components of the currents, which is not easily obtained by numerical simulation. In all the calculations, W_B has been assumed to be equal to the metallurgical base-width.

The main advantage of using a wide band-gap emitter is the suppression of the back-injection hole current [72]. Therefore, finding the minimal aluminum mole fraction, x, to achieve this suppression appears to be the first step in studying AlGaN/GaN HBTs. Regarding that, the static gain, $\beta = J_C/J_B$, is plotted versus x in Fig. 4.2 for a moderate- (5×10^{17} cm^{-3}) and a high- (5×10^{18} cm^{-3}) emitter doping density in an abrupt device for three temperatures. The results of the graded case are virtually the same (see Fig. 4.8). It is apparent that β increases with x to a maximum of 270 at 300 K, for instance, in either doping case, and is more or less constant beyond $x \approx 0.1$. Examining the two components of the base current in Eqn. (4.9) reveals that this is the mole fraction at which J_{hole} is about three orders of magnitude

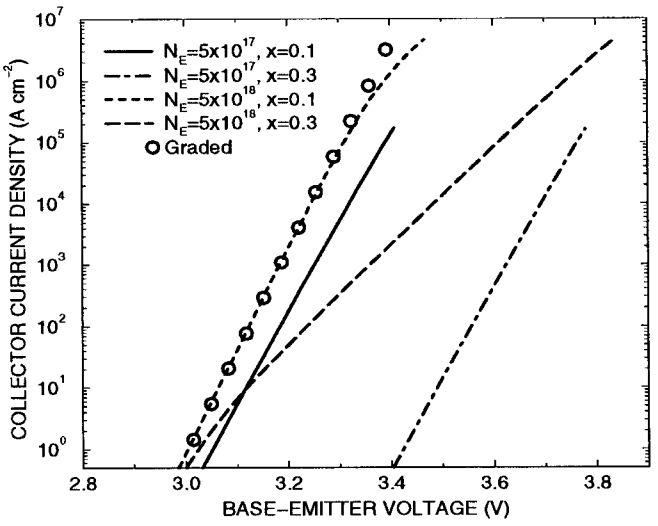

Figure 4.3: J_C versus V_{BE} calculated by Eqn. (4.1) for four abrupt devices with: (a) $N_E = 5 \times 10^{17}$, $x = 0.1$ (solid line); (b) $N_E = 5 \times 10^{17}$, $x = 0.3$ (dot-dashed line); (c) $N_E = 5 \times 10^{18}$, $x = 0.1$ (dashed line); (d) $N_E = 5 \times 10^{18}$, $x = 0.3$ (long-dashed line), and graded-emitter devices (circles). All the dopings are in cm^{-3}.

smaller than J_{rec}. In such circumstances, $\beta \approx \tau_n/\tau_B$, which is a well-known expression for base-limited homojunction bipolar transistors, where the quasi-neutral base signal-delay time is fairly assumed to be $\tau_B = W_B^2/2D_{nB}$ (see Eqn. (5.14)). That is why all the devices have roughly the same gain at high mole fractions. This simple relation also explains why the gain decreases with temperature: although $\tau_B \propto (T\mu_n^e)^{-1}$ decreases slightly with temperature, $\tau_n = 1/Bp_{0B}$ drops more drastically, causing an increase in J_B. The temperature dependence of the gain is discussed further in Sec. 4.4.3. To summarize, in contrast to AlGaAs/GaAs HBTs, where a mole fraction of 0.3 is typical, due to the large band-gaps of AlGaN/GaN heterostructures, 0.1-0.15 appears to be sufficient to suppress J_{hole}. However, it should be stressed here that if the material properties of nitrides are improved in a way that τ_n becomes larger than the value employed in this work, a larger minimal mole fraction would be essential in

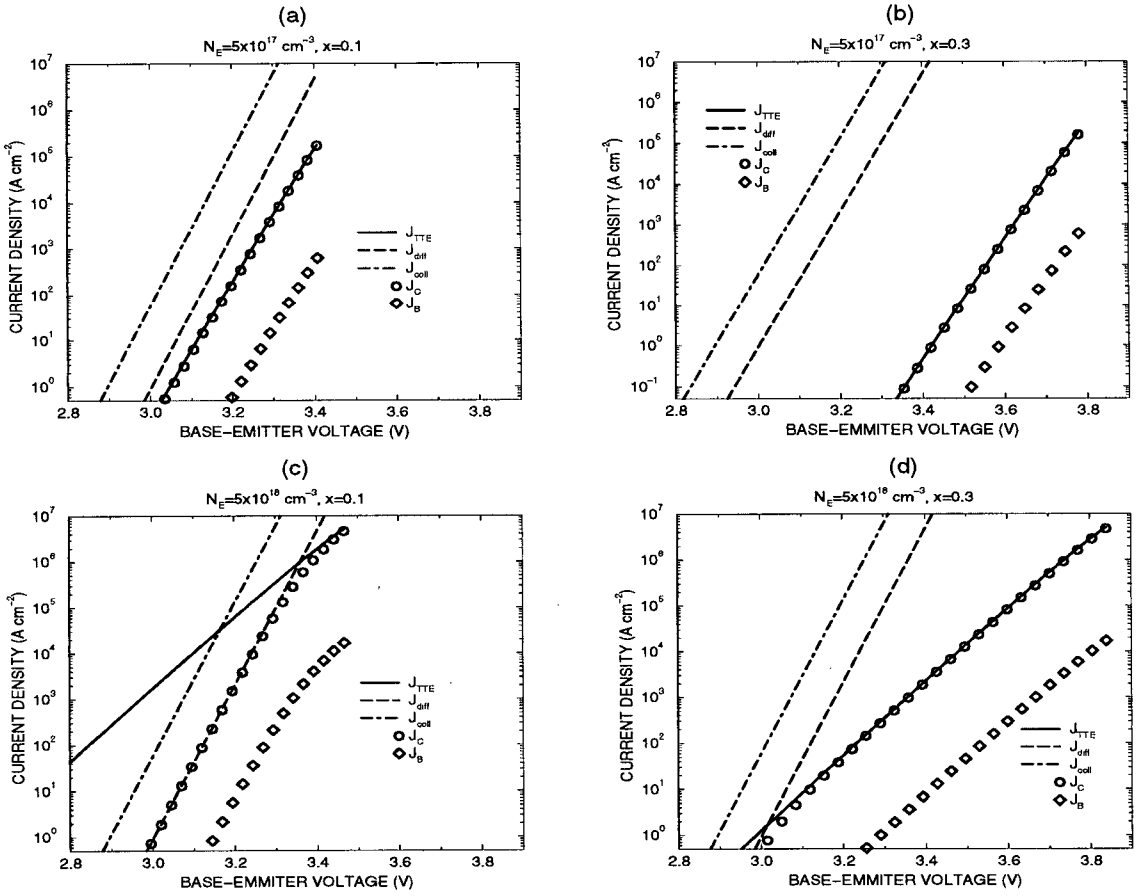

Figure 4.4: Various current densities versus V_{BE} for the four (a)-(d) abrupt HBTs of Fig. 4.3: J_{TTE} (solid line), J_{diff} (dashed line), J_{coll} (dot-dashed line), J_C (circles), and J_B (diamonds).

the design of high-gain AlGaN/GaN HBTs.

To study the emitter configuration further, Fig. 4.3 depicts J_C versus V_{BE} plots for five devices at room temperature. The emitter doping densities are the same as those in the previous figure, but the mole fraction is either 0.1 or 0.3 for either doping, resulting in the four abrupt devices characterized in the figure. The same emitter variations were studied for the graded case, but since their plots are virtually the same, one of them represents them all. The three J_{diff}, J_{TTE}, and J_{coll} components of all the abrupt devices are gathered in Fig. 4.4. As can be seen, each

one has its own peculiar features, and is worthy of a brief discussion. In all of them, J_{coll} is negligible. Device (c), $N_E = 5 \times 10^{18}$ cm^{-3} and $x = 0.1$, is diffusion dominated. The same is true for all the graded cases, as Fig. 4.3 shows the same J_C as device (c). All the three other cases are dominated by J_{TTE}. The two $N_E = 5 \times 10^{17}$ cm^{-3} cases (a), and (b), have different turn-on voltages, due to the higher barrier that has to be overcome by the electrons in the $x = 0.3$ case. The last device, (d), has a diode-ideality factor of over 1.5, while the rest have approximately the same value of unity. The reason for this is quite instructive to elaborate. Ref. [94] has a parallel flow rearrangement of J_{TTE}, where the two ideality factors, $n_1 = U'_{00}/(N_{rat}\tanh(U'_{00}))$, and $n_2 = 1/N_{rat}$, are due to the tunneling and thermionic currents respectively. For device (d), where $N_E = 5 \times 10^{18}$ cm^{-3}, $N_{rat} \approx 0.67$, $n_1 \approx 2$, and $n_2 \approx 1.5$, while for the one-order-of-magnitude lower doping density case, $N_{rat} \approx 0.95$, and both the ideality factors are about unity. So, in the design of any kind of abrupt HBT, one has to bear in mind that if J_C is dominated by J_{TTE}, N_E should be much less than N_B to obtain a diode ideality factor close to unity. This is on top of the conventional interest in having not too high emitter doping, in order to avoid band-gap narrowing in the region, as well as to achieve lower junction capacitance.

Fig. 4.4 also shows J_B versus V_{BE} for all the abrupt devices. It is instructive in the sense that one might think, since the tunneling and thermionic emission is influential on J_C, the high non-ideality factor issue would not appear in J_B. Figs. 4.4(c), and (d) clearly elaborate that this argument is not true. It is the result of the fact that J_B is dominated by recombination. In other words, both J_C and J_B are driven by the same potential difference, $V_{BE} - \Delta E_{Fn}$. Thus, γ strongly affects J_B too, as Eqn. (4.8) demonstrates.

In the following sections, we restrict ourselves to the $x = 0.1$ cases, which have

near-unity ideality factors, and small turn-on voltages. Employing a low mole fraction also minimizes the effect of spontaneous polarization in the graded layer of a graded-junction HBT, and at the hetero-interface in the abrupt case. Between the studied emitter doping densities, the lower one, *i.e.*, 5×10^{17} cm^{-3}, has been picked for the further studies reported here, to achieve smaller base-emitter junction capacitance (see Eqn. (5.12)).

4.4.2 Polarized HBTs

Aspects of polarization in nitrides were discussed thoroughly in Chap. 3. To study the effect of this phenomenon on the static characteristics of AlGaN/GaN HBTs, the developed expressions for the polarization volume charge density, ρ_{pol}, and polarization surface charge density, σ_{pol}, in Eqns. (3.8) and (3.15) were incorporated in the drift-diffusion simulator for the graded and abrupt HBTs respectively. As concluded in the same chapter, the 5000Å wide emitter is considered to be relaxed. Therefore, in contrast to AlGaN/GaN quantum wells, and HFETs, where the piezoelectric effect should be taken into account, it is not relevant in the devices of our interest. The spontaneous polarization is present, though, and as we will see has a considerable effect on the band diagrams and the Gummel plots of the devices. According to the discussions in the previous section device with $N_E = 5 \times 10^{17}$ cm^{-3} and $x = 0.1$, is the subject of this one-dimensional study at room temperature. The magnitude of $\sigma_{pol} = P_{SP}^{GaN} - P_{SP}^{AlGaN}$, applicable at the hetero-interface of the abrupt HBT, is 5.2×10^3 C m^{-2}, and the constant polarization charge in the graded region is $\rho_{pol} = -\nabla.\mathbf{P}_{SP}$, which has the magnitude of 2.6×10^5 C m^{-3}, and is the derivative of the linear spontaneous polarization with respect to the fixed coordinate, **z**, with the positive direction being from the collector to the emitter. According to the dis-

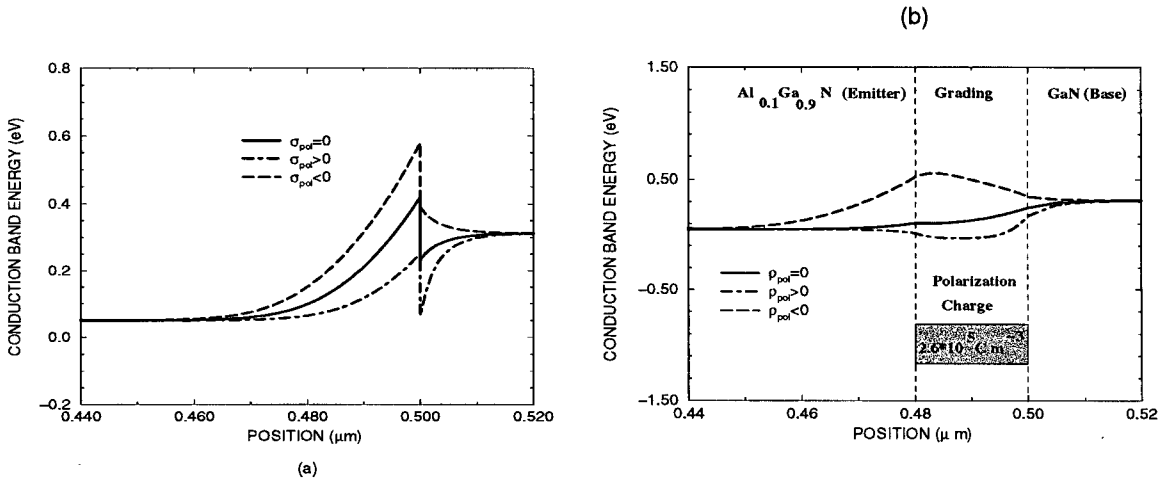

Figure 4.5: Effect of spontaneous polarization charges on the conduction band of (a) abrupt-emitter and (b) graded-emitter HBTs with $N_E = 5 \times 10^{17}$ cm^{-3}, $x = 0.1$, $V_{BE} = 3.0$ V: zero charge (solid line); positive charge (dot-dashed line); and negative charge (dashed line).

cussion in Sec. 3.5, both these charges are positive if the growth direction is [0001] or Ga-faced, and are negative if we have [000$\bar{1}$] or N-faced HBTs.

Figs. 4.5 (a), and (b) show the huge effect of the polarization charges on the conduction bands of the abrupt and graded cases, respectively, for $V_{BE} = 3.0$ V. The plots can be simply understood by recalling the boundary conditions required to solve Poisson's equation, as thoroughly discussed in Sec. 3.3. For instance, in the abrupt case, instead of $\epsilon_1 d\psi_1/dz = \epsilon_2 d\psi_2/dz$, the boundary condition $\epsilon_1 d\psi_1/dz - \epsilon_2 d\psi_2/dz = \sigma_{pol}$ must be satisfied at the hetero-interface for the vacuum potential ψ, where 1 refers to the base, and 2 to the emitter materials. If $\sigma_{pol} > 0$, the slope of the conduction band energy would be larger on the base side, compared to the emitter side of the junction[2]. This argument is verified by Fig. 4.5(a). To satisfy that, the band has to bend more steeply in the base side, causing a significant decrease in the barrier. On the other hand, for a negative σ_{pol}, the conduction band slope turns out to be

[2]Recall that the conduction band energy, $E_C = -q\psi - \chi$, where χ is the affinity.

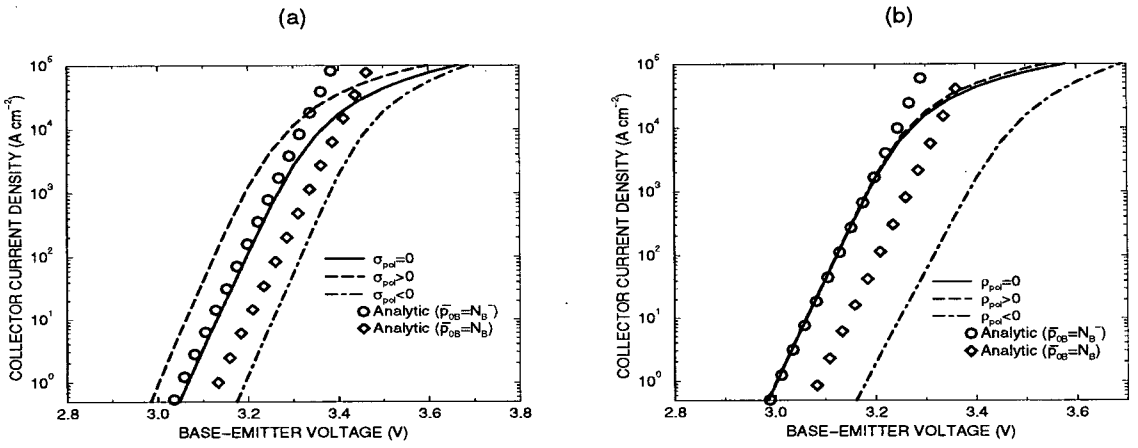

Figure 4.6: Effect of spontaneous polarization on J_C versus V_{BE} plots for (a) the abrupt and (b) the graded devices of Fig. 4.5: zero charge (solid line); positive charge (dashed line); negative charge (dot-dashed line); Analytic values from Eqn. (4.1) for the cases of incomplete- (circles), and complete- (diamonds) ionization are also shown.

negative on the base-side, resulting in a larger barrier for the transport of electrons. Similar arguments are applicable for interpreting the graded-emitter band diagrams in Fig. 4.5(b).

The J_C Gummel plots for negative-, positive-, and zero-charge cases are shown in Figs. 4.6 (a), and (b) for abrupt and graded HBTs, respectively. The smaller abrupt barrier in the positive case has caused an increase in J_C of about one order of magnitude in the linear region compared to the $\sigma_{pol} = 0$ case, while the negative charge degrades the current by two orders of magnitude. In the graded device, the smaller hump in the conduction band of the positive case does not give any improvement in the current compared to the zero charge case, since the effect of the barrier has been already removed by the grading. However, the big conduction band hump of the $\rho_{pol} < 0$ case, causes a reduction in the current of about three orders of magnitude, with respect to the zero-charge case.

The analytic J_C results are also presented in Figs. 4.6(a), and (b) to demon-

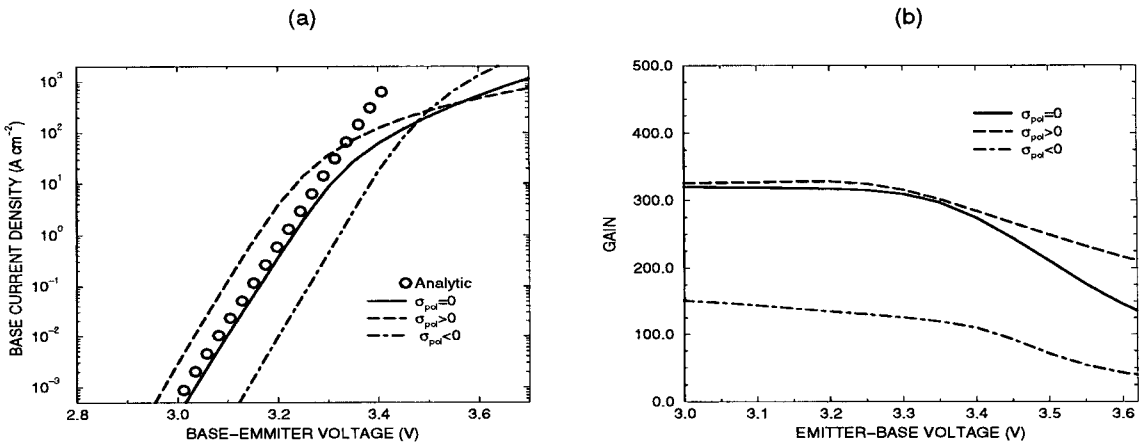

Figure 4.7: Effect of spontaneous polarization on (a) J_B and (b) β versus V_{BE} for the abrupt device of Fig. 4.5(a): zero charge (solid line); positive charge (dot-dashed line); negative charge (dashed line). Also shown in (a) are the analytic values of Eqn. (4.9) (circles).

strate the good agreement with the numerical results, if polarization effects are not considered. As can be seen in both the abrupt and the graded cases, the discrepancy between the analytic and simulation results becomes larger at higher biases, due to the onset of high-level injection in the quasi-neutral base, which is ignored in the analytic expressions. This is a classical behavior of homojunction bipolar transistors, where the base doping is smaller than the emitter doping, but it also happens at a somewhat lower collector current density of about $3 - 5 \times 10^3$ A cm^{-2} in nitride HBTs. The reason for this is the large ionization level of Mg dopants, which leads to incomplete ionization in the base. High-level injection occurs when the injected electron concentration is comparable with the hole concentration in the base, and electrons cannot be assumed to be minority carriers anymore. An electric field arises in response to these carrier concentrations, causing a drift current to flow. A rough estimate of the collector current density for the appearance of this phenomena in homojunction devices, J_{HLI}, is proportional to $T\mu_n N_B^-$ [99] (Chap. 10, p. 253), where

μ_n is the electron mobility in the base. The majority hole carrier in the base of the nitride transistor $N_B^- = 4.4 \times 10^{17}$ cm^{-3}. Since this value is about 20 times smaller than N_B, the onset of high-level injection is expected to happen at low current densities, when compared to GaAs HBTs for instance.

The polarization charges influence the valence bands in a way that resembles the conduction band shifts, with the same trends for positive and negative cases. The total hole barrier is not affected by polarization, though, and one expects that J_{hole} is unchanged. As discussed before, J_{rec} is dominant in AlGaN/GaN HBTs, and with similar arguments to the one in the previous subsection regarding the coupling of J_C and J_{rec} through ΔE_{Fn}, we expect that the polarization should affect J_B, as Fig. 4.7 (a) elaborates for the abrupt HBT. The figure looks very similar to the J_C plot in Fig. 4.6(a), but of course with a relative shift to lower current densities for zero, positive, and negative σ_{pol}. The amount of the shifts are charge dependent. To see the effect of polarization on the gain, the gain plots versus bias are presented in Fig. 4.7(b) for the same abrupt device. $\sigma_{pol} > 0$ has caused a small improvement in the gain compared to the zero-charge case, but for $\sigma_{pol} < 0$, the gain has been roughly halved. That is because J_B in the $\sigma_{pol} < 0$ device is no longer dominated by the quasi-neutral base recombination, and the contribution of J_{hole} is considerable. To elaborate on this, we examined the numerical quasi-Fermi level plots at $V_{BE} = 3.0$ V. While the zero-charge case has $\Delta E_{Fn} = 0.047$ eV, due to the large barrier of the negative-charge device ΔE_{Fn} is increased to 0.19 eV. According to the corresponding hyperbolic term in Eqn. (4.9), $J_{hole} = 1.1 \times 10^{-6}$ A cm^{-2}, which is not a function of ΔE_{Fn}. However, J_C and J_{rec} are proportional to $\exp(-q\Delta E_{Fn}/k_BT)$. By picking the analytic values, $J_C = 2.4 \times 10^{-1}$ A cm^{-2} and $J_{rec} = 8.7 \times 10^{-4}$ A cm^{-2}, which are valid for $\sigma_{pol} = 0$ case. If, according to the above-mentioned numerical values for ΔE_{Fn}, we modify them for

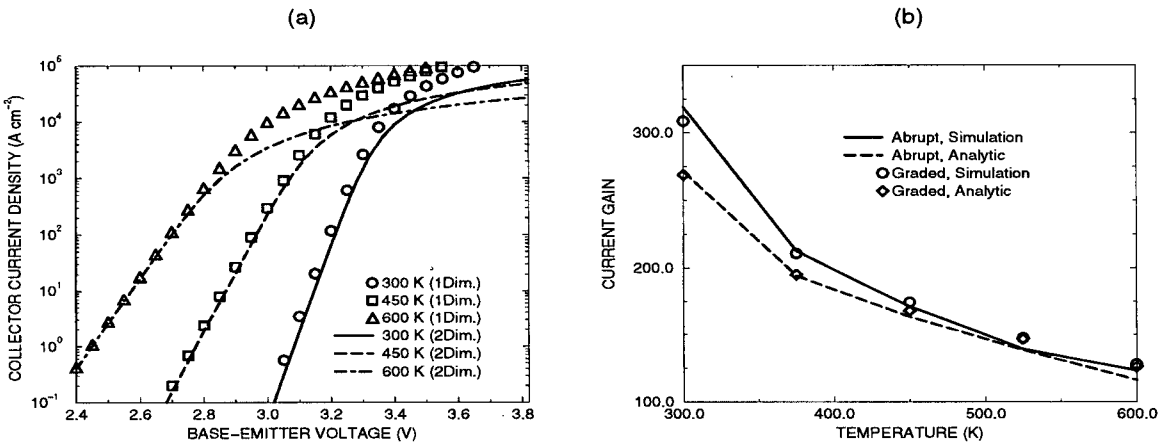

Figure 4.8: (a) J_C versus V_{BE} for the abrupt HBT at three temperatures without any polarization charge effects, with two-dimensional simulations: 300 K (solid line); 450 K (dashed line); 600 K (dot-dashed line); and also one-dimensional simulations: 300 K (circles); 450 K (squares); and 600 K (triangles); (b) β versus T for AlGaN/GaN HBTs without any polarization charge effects: simulated abrupt (solid line); analytic abrupt (dashed line); simulated graded (circles); and analytic graded (diamonds).

the $\sigma_{pol} < 0$ device by applying the factor of $(q(0.047 - 0.19)/k_BT)$, it turns out that J_{rec} is only about three times larger than J_{hole}. Consequently, $\beta = J_C/(J_{rec} + J_{hole})$ can be estimated to be 155 for the negative-charge device, rather than 270, which is in excellent agreement with the results in Fig. 4.7(b). Similar arguments apply for the effect of volumetric polarization charge on the gain of graded-emitter HBTs.

4.4.3 High-Temperature HBTs

As discussed in Sec. 4.1, nitride transistors are among the candidates for high-temperature electronics, since they remain extrinsic at elevated temperatures. In investigating the high-temperature performance, polarization effects are not considered. This is because of the lack of data on the temperature-dependence of the spontaneous polarizations, and also because the importance of polarization will diminish as the temperature approaches the Curie point. On the other hand, the gain

study of the nitride HBTs is very dependent on the radiative recombination lifetime, since it dominates J_B. As explained in Sec. 2.6, we have picked the experimental value of $B = 6.6 \times 10^{-10}$ cm^{3}s^{-1} at 300 K, and done a theoretical extrapolation which gives $B = 2.2 \times 10^{-10}$ at 600 K, for instance. So, using a temperature-dependent B for estimating $\tau_n = 1/B\overline{p}_{0B}$, will compensate the 7-8 times increase of $\overline{p}_{0B}$ at 600 K to some extent.

Fig. 4.8(a) presents the Gummel plots of J_C at three temperatures without any polarization charge features for the abrupt device. To stress the validity of one-dimensional simulations in our sample devices, the two-dimensional results are also added. Discrepancies only arise in the very high-current regimes. As can be seen, departure from the ideal regime occurs at lower currents at higher temperatures (about 10^3 A cm^{-2}). This is on top of the previously discussed high-level injection phenomenon. It was found out that this is due to the large influence of recombination in the quasi-neutral base on the collector current, violating the constancy of the hole quasi-Fermi level in the base, which is an essential assumption for the applicability of the one-dimensional approximation, as long established by Gummel [97]. The high-level injection occurs at roughly the same current densities for the three temperatures. That might seem odd, since the ionization factor increases from 4% at 300 K to 33% at 600 K. However, the shrinkage of mobility counteracts that to some extent, keeping J_{HLI} somewhat invariant.

By viewing Fig. 4.8 (a), one should not be tempted into thinking that the ideality factor increases with temperature, since the decrease in the slope of the plots is basically due to the T dependence of $\exp(qV_{BE}/nk_BT)$, where the ideality factor, n, remains close to unity at all temperatures. As can be seen, a major effect of elevated temperature is to reduce the turn-on voltage. At 600 K, V_{bi} decreases by 0.2

V from its 3.46 V value at room temperature. This will decrease the height of the barrier in Fig. 4.1. Also $\overline{n}_{0B}$ increases enormously in higher temperatures, causing the thermionic emission current to increase about 2×10^{10} times from 300 to 600 K at $V_{BE} = 2.5$ V, for instance. Somewhat similar orders of magnitude increases are applicable to the other components of J_C. It explains why the current density is much higher for a given bias in the ideal regime at higher temperatures, or, in other words, operating at elevated temperatures has the advantage of lower turn-on voltages.

Finally, Fig. 4.8(b) shows the analytic and simulated gains versus temperature for the graded and abrupt devices. With respect to the dominance of recombination in the base current, we see that all the cases have more or less the same value at a given temperature, starting at about 320 at room temperature (abrupt simulation), and degrading to the values of about 120 at 600 K. This might seem optimistic regarding the best experimental reported to date of 100 at 300 K [87]. However, recalling that our B at 300 K is an experimental value, found useful in modeling UV-photodiodes, we can conclude that a current gain of about 300 at room temperature can be claimed as a reasonable performance *limit* for future AlGaN/GaN HBTs. Meanwhile, analytic studies show that grading the base can cause significant improvement of the gain to about 2,000 at room temperature [100]. An attainable gain of over 100 at 600 K is also an encouraging figure of merit for the future of these devices for high-power, high-temperature electronic applications.

4.5 Summary

A comprehensive study of the static performance of AlGaN/GaN HBTs was performed, and the following conclusions can be drawn:

- Due to the large band-gap difference, a mole fraction of 0.1 seems to be sufficient for obtaining the benefits of a wide band-gap emitter in terms of suppression of the back-injection hole current into the emitter. In such circumstances, the abrupt devices are tunneling- and thermionic-emission-limited for low emitter doping densities. For high-emitter doping, abrupt (low mole-fraction) and graded devices are limited by diffusion in the base. In all cases, the high radiative recombination in the base dominates the base current.

- Polarization charges have a large effect on the characteristics of the devices. Generally speaking, the negative charge of the N-faced material causes degradation in the collector current and the gain for either graded or abrupt cases, while the positive charge of Ga-faced material improves the collector current of abrupt HBTs, but does not affect the graded devices.

- High level injection in the base limits the maximum available current densities to about 3-5 kA cm^{-2} in nitride HBTs. At room temperature this is due to the incomplete ionization of deep acceptors. At higher temperatures, the degradation in the hole mobility compensates higher ionization factors, making the high-level injection current density only weakly dependent on temperature.

- The temperature-dependence study of the current gain, with the employed recombination lifetimes in this work, predicts that uniform-base AlGaN/GaN HBTs can attain a current gain as high as 300 at room temperature, shrinking to about 120 at 600 K.

Chapter 5

Regional Signal-Delays in HBTs

5.1 Overview

In order to study the high-frequency performance of HBTs, the common-emitter unity current gain frequency, or the cut-off frequency, has been the figure of merit considered in the present work. The other important dynamic figure of merit, the maximum oscillation frequency, or unity power gain frequency, is related to it through a well-known square-root expression [99] (Chap. 7, p. 177). The charge control theory offers a powerful cornerstone for both numerical and analytical calculations of the emitter-to-collector signal delay time, which is inversely proportional to the cut-off frequency. Traditional compact expressions for estimating the different signal-delay times have been long established for homojunction devices, and are widely used as design tools. In contrast, numerical drift-diffusion simulators can easily give an exact evaluation of the cut-off frequency. Some efforts have also been implemented to tailor the analytical expressions for the case of HBTs in the last two decades [108], [110]. To the best of our knowledge, an exact regional comparison between the two approaches has not been done for these new high-frequency devices. Due to the high doping

density in the base, the band-gap narrowing phenomenon has a significant effect on the transit time of that region. In addition, for short-base devices, quasi-ballistic transport, and hot-electron injection play roles as far as the base delay is concerned. Grading the emitter-base heterojunction is also an issue in real devices. All these features can be easily implemented in simulations, but cannot be simply taken into account in the compact analytic expressions, making the exactness of the traditional analytic approach questionable. This has been the motivation for this study.

Sec. 5.2 is a critical review of the charge control theory. Sec. 5.3 discusses the problems with the available schemes in partitioning the device into appropriate regions to facilitate comparison with the compact expressions, where a new scheme called the integrated charge method is proposed. Sec. 5.4 reviews the compact expressions for the various delay times. It is believed that the limitations of the classical expressions are not widely known, even for conventional material systems [109]. Therefore, the results in Sec. 5.5 relate to mature-technology AlGaAs/GaAs HBTs, to more clearly promote the issue. In contrast, not all the subtle modelings, like band-gap narrowing, are available for the AlGaN/GaN system. Regarding the high frequency performance of AlGaN/GaN HBTs, Ref. [77] employs first-order analytic expressions for estimating the cut-off frequency with an estimate of 44 GHz. However, incomplete ionization of acceptors in AlGaN/GaN HBTs, and high level injection at quite low biases, makes the comparisons more complicated in these devices. Therefore, a separate section on the aspects of charge control theory in nitrides is the last part of this work. This work is the first in terms of studying the non-conventional aspects of incomplete ionization in the charge control theory of these novel devices.

5.2 Signal-Delay Times

The common-emitter unity current gain frequency, or the cut-off frequency, f_T, is an important figure of merit in high-frequency bipolar transistors. According to the hybrid-π model, in the absence of significant parasitic resistances, it is proved in any standard textbook, [99] (Chap. 7, p. 175), that,

$$f_T = \frac{1}{2\pi} \cdot \frac{g_m}{C_\pi + C_\mu} \tag{5.1}$$

where $g_m = (\partial I_C / \partial V_{BE})|_{V_{CE}}$ is the transconductance, and $C_\pi = C_{je} + C_S$. C_{je} and C_μ are the junction capacitances of the base-emitter, and base-collector *space charge regions* respectively, and $C_S = (\partial Q_S / \partial V_{BE})|_{V_{CE}} = g_m(\partial Q_S / \partial I_C)|_{V_{CE}}$ is the capacitance due to the stored charges in the device, in response to a small variation in the base-emitter voltage, V_{BE}, while the collector-emitter voltage, V_{CE}, is held constant. Usually, it is assumed that Q_S is merely due to the variation of the minority carrier charge Q_B in the *quasi-neutral base*[1] [99] (Chap. 7, p. 172). Considering a *n-p-n* transistor, and assuming that the carrier current across the quasi-neutral base is merely the diffusion of minority carriers, both Q_B and collector current, I_C, will be proportional to the minority carrier concentration at the base side of the base-emitter space-charge region, $n(x_{pE})$, and therefore $C_S = C_B = g_m(Q_B / I_C)$. In other words, both quantities are exponential functions of V_{BE}, *i.e.*, $n(x_{pE}) \propto \exp(qV_{BE}/k_BT)$, where q is the electron charge, T is the temperature, and k_B is the Boltzmann constant. Herein, we equivalently write $C_B = g_m(\Delta Q_B / \Delta I_C)|_{V_{CE}}$ for a small (much less than k_BT/q) incremental ΔV_{BE} in the bias.

On the other hand, analyzing a one-dimensional device in the x direction[2], we

[1] Quasi-neutral base is that region inside the base, which is outside the emitter and collector space-charge regions, and in which the electric field is negligible. The phenomenological definition assumed in this work appears in Sec. 5.3.3.

[2] This is the $-z$ axis of Chaps. 3, and 4.

can define the average base transit (or signal-delay) time, τ_B, as

$$\tau_B = \int_{x_{pE}}^{x_{pE}+W_B} dx/v_n(x) = qA \int_{x_{pE}}^{x_{pE}+W_B} n(x)dx/I_C = Q_B/I_C$$
$$= \left.\frac{\Delta Q_B}{\Delta I_C}\right|_{V_{CE}} = q \left.\int_{x_{pE}}^{x_{pE}+W_B} \frac{\Delta n(x)}{\Delta J_C} dx\right|_{V_{CE}}, \tag{5.2}$$

where $v_n(x)$ is the diffusive velocity of electrons, flowing in the cross-section area A, with the current density J_C. Consequently, assuming that C_B is dominant in Eqn. (5.1), we can write

$$\frac{1}{2\pi f_T} \simeq \tau_B = q \left.\int_{x_{pE}}^{x_{pE}+W_B} \frac{\Delta n(x)}{\Delta J_C} dx\right|_{V_{CE}}. \tag{5.3}$$

This is a rather interesting result, since it links the high-frequency figure of merit f_T, with the base signal-delay time, as well as the DC variation of the electron charge in the quasi-neutral base. Physically, one can imagine that the base-emitter and base-collector junction capacitances can be similarly related to the corresponding signal delay times in the space-charge regions. Mathematically, for any region i of the device, with capacitance C_i, which has a charge variation ΔQ_i in response to ΔV_{BE}, through the chain rule we have

$$C_i = \left|\frac{\partial Q_i}{\partial V_{BE}}\right| = \left|\frac{\partial I_C}{\partial V_{BE}} \cdot \frac{\partial Q_i}{\partial I_C}\right| = g_m \left.\left(\frac{\Delta Q_i}{\Delta I_C}\right)\right|_{V_{CE}}. \tag{5.4}$$

Finally, if we define the signal-delay time of region i as $\tau_i = \Delta Q_i/\Delta I_C\big|_{V_{CE}}$ $= q \int_i (\Delta n(x)/\Delta J_C)dx$, and the emitter-to-collector signal-delay time, τ_{EC}, as the summation of all these delays, with respect to the linear dependence of f_T^{-1} to various capacitances in Eqn. (5.1), we can generalize Eqn. (5.3) to

$$\frac{1}{2\pi f_T} = \tau_{EC} = \sum_i \tau_i = \sum_i \left.\frac{\Delta Q_i}{\Delta I_C}\right|_{V_{CE}}$$
$$= q \int_0^L \left.\frac{\Delta n(x)}{\Delta J_C} dx\right|_{V_{CE}} = q \left.\int_0^L \frac{\Delta p(x)}{\Delta J_C} dx\right|_{V_{CE}}, \tag{5.5}$$

where L is the length of the one-dimensional device. The last identity for hole concentration $p(x)$ is valid, because of the neutrality of the transistor as a whole, *i.e.*, one can expect that the total variation in electron concentration is equal to the corresponding one for holes, as far as we assume complete ionization of dopants.

The last formulation in Eqn. (5.5) is known as the *charge control* representation of τ_{EC}. Mathematically, $C_\pi + C_\mu$ of the hybrid-π model can be directly linked to $\Delta Q/\Delta I_C\big|_{V_{CE}}$ in a general sense, where Q is the total electron or hole charge in the transistor [93]. The approach presented here has been followed to emphasize the physical interpretation of the theory in terms of the transit times[3]. It should be borne in mind that Eqn. (5.1) is an approximation, and does not take into account the parasitics. However, they can be incorporated in the charge control approach as shown in Eqn. (5.21), and the breaking-up of the signal delay into various regions of the device remains a valid concept.

5.3 Partitioning the Device into Different Regions

As discussed in the previous section, τ_{EC} can be calculated numerically by means of Eqn. (5.5). But for getting an idea of the relative contribution of different regions of the device, a partitioning scheme must be employed. The partitioning scheme should not only have a well-defined criterion for implementing the numerical approach, but also should have a justifiable physical basis. Such a scheme will simplify the numerical calculation, and the comparison with the traditional analytical expressions. Unfortunately there is no widely-accepted scheme that satisfies both of these conditions. Conventionally the device is split into three regions designated as emitter, base, and

[3]Another interesting approach is by Moll and Ross [101], who link the base transit time to the frequency at which very short pulses of carriers can be injected into the emitter, and still be distinguishable when they reach the collector contact.

collector, while the signal-delay time of each is defined by

$$\tau_E = \left.\frac{\Delta Q_E}{\Delta I_C}\right|_{V_{CE}}, \tag{5.6}$$

$$\tau_B = \left.\frac{\Delta Q_B}{\Delta I_C}\right|_{V_{CE}}, \tag{5.7}$$

$$\tau_C = \left.\frac{\Delta Q_C}{\Delta I_C}\right|_{V_{CE}}. \tag{5.8}$$

τ_B is the quasi-neutral base delay time, τ_E and τ_C are the quasi-neutral and space-charge region delay times of emitter and collector, respectively. Usually τ_E is split into two parts τ_{EB}, and τ_{QNE}, denoting the space-charge and quasi-neutral signal-delay times respectively [4]. These are schematically shown in Fig. 5.1(a) for a simplified pedagogical device. To contrast the subtle features of a real device, simulated plots of $q\Delta n(x)/\Delta J_C\big|_{V_{CE}}$ and $q\Delta p(x)/\Delta J_C\big|_{V_{CE}}$ versus J_C for a typical graded GaAs HBT, discussed in Sec. 5.5, are presented in Fig. 5.2.

It is difficult to locate the edges of the space-charge regions (x_{nE} and x_{pE}) in forward bias, since there is no sharp transition between the space-charge regions and quasi-neutral regions. In terms of analytic theories, the only available method to partition the device is the depletion approximation, which as we will see (see Fig. 5.3) is not in accord with the quasi-neutrality concept. Any suggested numerical method in the literature has its own advantages and disadvantages. Some of these methods, as well as the one attempted in this study are discussed in the following subsections.

[4] The same argument can be utilized for τ_C, in principle. However, since the quasi-neutral collector and sub-collector signal-delay times are virtually zero, it is not common in the literature.

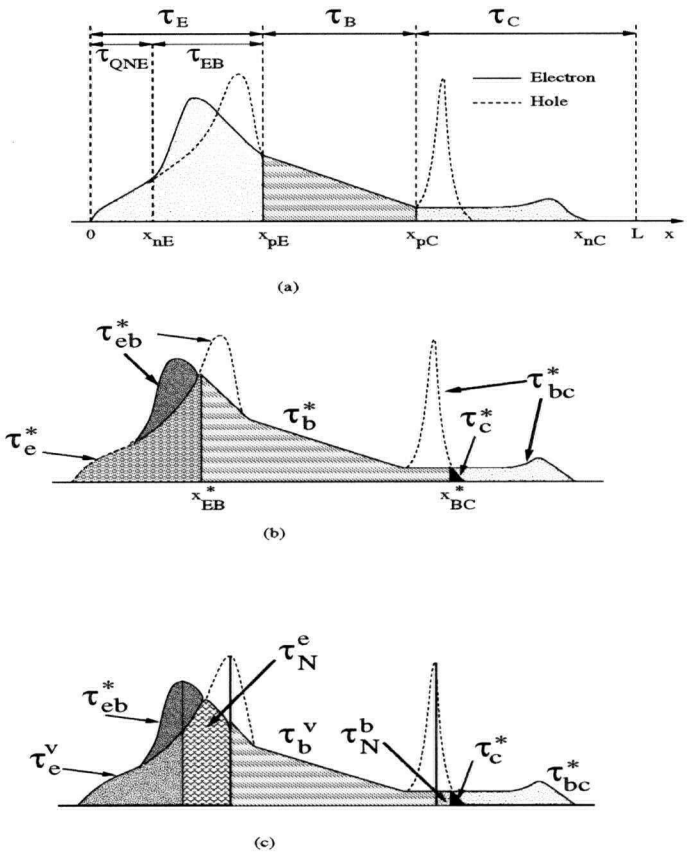

Figure 5.1: Schematic of $q\Delta n(x)/\Delta J_C$ (solid line), and $q\Delta p(x)/\Delta J_C$ (dashed line), versus x. (a) The conventional partitioning based on space-charge and quasi-neutral concepts. (b) van den Biesen's numerical scheme with x^*_{eb} and x^*_{bc} as boundaries, and five asterisked delay time components. (c) van den Biesen's boundaries for analytical purposes.

5.3.1 Metallurgical Boundaries

The simplest method that comes to mind and can be simply employed numerically is choosing the metallurgical junctions between the different emitter, base and collector regions as the boundaries. One should be careful when employing this scheme that the electron charge should be chosen for different regions (in *n-p-n* transistors), since the quasi-neutrality concept is overlooked in this scheme, and for example in the base region, the $q\Delta p(x)/\Delta J_C$ spikes of the space-charge regions should not be attributed to the base delay time. This method has been employed in Refs. [102] and [103].

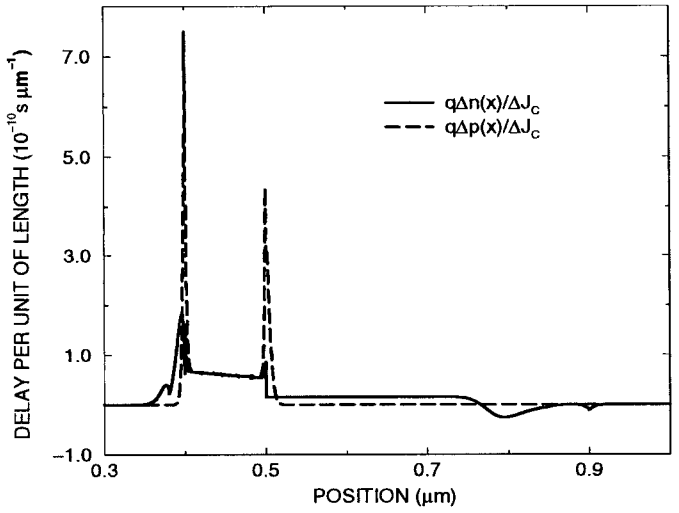

Figure 5.2: Plot of $q\Delta n(x)/\Delta J_C$ (solid line), and $q\Delta p(x)/\Delta J_C$ (dashed line) versus x for the studied graded GaAs HBT (see Sec. 5.5) at $J_C \approx 1 \times 10^4$ A cm^{-2}.

The problem with this approach is the fact that these boundaries do not partition the device into regions with a good electrical basis, *i.e.*, it is difficult to interpret the results with analytic expressions, since the boundaries are inside the two space-charge regions, and the related delay times which can be manipulated analytically are distributed in two different numerical regions.

5.3.2 van den Biesen's Regional Scheme

Another scheme which seems at first sight to be promising to reconcile the partitioning scheme issue is due to van den Biesen [104]. He uses the intersections of $q\Delta p(x)/\Delta J_C\big|_{V_{CE}}$ and $q\Delta n(x)/\Delta J_C\big|_{V_{CE}}$ as new boundaries, suggested earlier by Kwok [105]. As Fig. 5.1(b) depicts, there are two points for these intersections, x^*_{EB} and x^*_{BC}, close to the emitter-base and base-collector metallurgical junctions, respectively. The asterisks are to emphasize that these points are different from the metallurgical

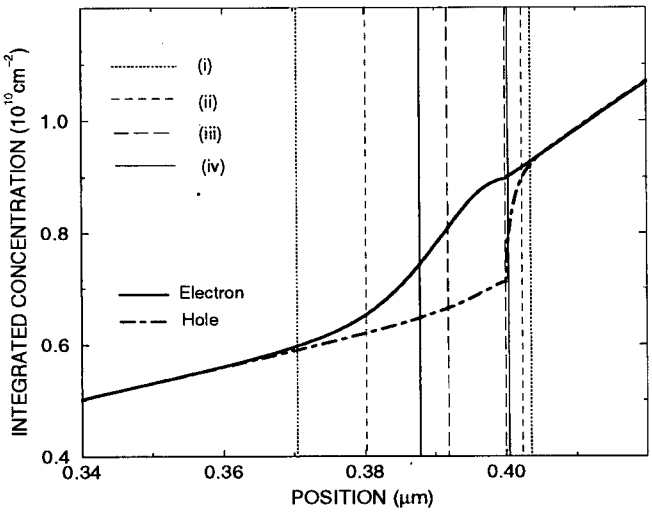

Figure 5.3: The cumulative integrals $\int_0^x \Delta n(x')dx'$ (solid curve) and $\int_0^x \Delta p(x')dx'$ (dot-dashed curve) of the studied abrupt HBT (see Sec. 5.5) for $J_C \approx 1 \times 10^4$A cm^{-2} around the base-emitter junction at $x = 0.4\mu$m, with the estimates of the space-charge region width by different methods: (i) 1% criteria of the integrated charge method (dotted line); (ii) 5% criteria of the same method of part (i) (dashed line); (iii) van den Biesen's maxima (long-dashed line); (iv) Depletion approximation (solid line).

boundaries. Only in the case of symmetrical junctions do the two methods give identical boundaries. For asymmetrical step junctions, the asterisked boundaries have an obvious tendency to be in the lightly doped material. By means of these two new boundaries, van den Biesen proposes five new delay time components, where the addition of all is $\tau_{EC} = \tau_{eb}^* + \tau_{bc}^* + \tau_e^* + \tau_b^* + \tau_c^*$. These are schematically shown in Fig. 5.1(b), and can be defined mathematically in terms of $\Delta p(x)/\Delta J_C$, and $\Delta n(x)/\Delta J_C$ [104]. τ_{eb}^* and τ_{bc}^* are the delay times due to the uncompensated charges in the space-charge regions, *i.e.*, the amount of electron incremental charge which is in excess of the hole incremental charge in the region. Because of the duality of electrons and holes, it has an equivalent definition in terms of holes (see the two equi-area uncom-

pensated humps in Fig. 5.1(b)). Since $\Delta p(x)/\Delta J_C \approx 0$ in the reverse-biased collector space-charge region, $\tau_c^* \approx 0$. For the same reason τ_{bc}^* is essentially identical with the ordinary collector delay time defined by Eqn. (5.8), provided that the asterisked boundary is utilized in estimating the space-charge region width.

Although this partitioning scheme is unambiguous, and splits τ_{EC} neatly into five components, it encounters serious problems when interpreting the results in terms of analytical expressions (except for the two collector related delay times, τ_c^* and τ_{bc}^*, as discussed in the previous paragraph). The boundaries of τ_b^* are far from the edges of the quasi-neutral base, when compared to Fig. 5.1(a). To circumvent this problem, in addition to the asterisked boundaries for distinguishing the three regions, van den Biesen proposes picking the two maxima of $q\Delta p(x)/\Delta J_C\big|_{V_{CE}}$ as the edges of the quasi-neutral base for analytical purposes, as shown in Fig. 5.1(c). Recalling that for an ideal junction, these maxima coincide with the delta functions of the depletion approximation, this seems to have physical justification. But beyond that, he has to add some fraction, about half, of the so-called compensated charge in the base-emitter space-charge region, τ_N^e (see Fig. 5.1(c)), to be consistent with his numerical τ_b^*, which he claims can be neglected for homojunction silicon devices. He also implicitly neglects what we have shown as τ_N^b in the figure.

With regard to τ_e^*, the quasi-neutral emitter edge is similarly selected at the maxima of $q\Delta n(x)/\Delta J_C\big|_{V_{CE}}$ for analytical purposes. Once again, he has to add the remaining part of τ_N^b, to be consistent with his numerical partitioning scheme. Finally, in another set of articles [106], [107], he has tried to develop some analytical expressions for the capacitance corresponding to τ_{eb}^*.

Since his partitioning scheme has no ambiguity in terms of numerical analysis (only using intersections and extrema), it seems to be a promising way of resolving

the old issue of regionalizing the device in the charge control approach. But, after considerable work on this method, it was found out that it is not a practical approach, especially for HBTs. The reasons are:

1) Defining the edges of quasi-neutral regions at the mentioned maxima is reasonable for biases around equilibrium (J_C small). It was verified numerically that these boundaries have good agreement with the depletion approximation and the integrated charge method (to be discussed later) definition of quasi-neutrality in low biases, but it fails to identify a definite boundary in the high biases of practical interest, especially for the emitter edge of the base-emitter space-charge region, where there is a smooth transition in $q\Delta n(x)/\Delta J_C$ (see Fig. 5.3).

2) If the collector doping density, N_C, is greater than the corresponding N_B value for the base, x^*_{BC} is located in the base material. However, for the case of HBTs, where $N_B \gg N_C$, x^*_{BC} is deeply located in the collector material (see Fig. 5.2). This violates using van den Biesen's approach for interpreting τ_b^* analytically, since there is considerable amount of collector space-charge region delay time in τ_b^*.

3) With regard to τ^*_{eb}, the approximate analytical expressions are applicable for reverse biases and very high biases, or for all voltage ranges of a symmetrical homojunction [108], but, they are not generally useful for asymmetrical heterojunctions.

4) Finally, it was found out that one cannot simply ignore the two portions of τ^e_N in modifying τ_b^* and τ_e^* analytically for very high-frequency devices.

5.3.3 Integrated Charge Method

The proposed partitioning approach, is called the *integrated charge method* [109], and is based on the following physical observations: (i) $\Delta n(x) = \Delta p(x)$ in the quasi-neutral regions; (ii) the variation of majority-carrier concentration greatly exceeds

the variation of minority-carrier concentration near the edges of space-charge regions; (iii) the variation of charges calculated in terms of electrons or holes are not only equal for the whole device, but also in any region designated as a quasi-neutral region or a space-charge region, *i.e.*, $qA\int_i \Delta n(x)dx = qA\int_i \Delta p(x)dx$, where i can be any quasi-neutral region or space-charge region. These observations can be visualized more sensibly by careful study of Figs. 5.2, 5.3, and 5.9 for different quasi-neutral and space-charge regions throughout the device. These observations are in accord with the conventional picture in Fig. 5.1(a), giving consistency for region to region comparison with the traditional compact expressions.

With respect to this observation, we have defined the boundaries of the different regions at the positions where the relative deviation of cumulative delays calculated by means of electrons and holes is 1%, *i.e.*, the positions x throughout the device where

$$\left| \frac{\int_0^x \Delta p(x')dx'\Big|_{V_{CE}} - \int_0^x \Delta n(x')dx'\Big|_{V_{CE}}}{\int_0^x \Delta p(x')dx'\Big|_{V_{CE}}} \right| = 0.01 \tag{5.9}$$

This approach defines the edges of the quasi-neutral base such that there is little variation when the criterion is changed slightly. However, the integrated charge method cannot define a clear boundary between the quasi-neutral emitter and the base-emitter space-charge region. That is because of the slow transition of the quantity in the region, *i.e.*, changing the criteria from 0.01 to a slightly different one, *e.g.*, 0.05 has tremendous effect in estimating the width of quasi-neutral emitter. The same argument is true for all the other observable physical quantities, *e.g.*, conduction band, electric field, $(N_B + p - n)/N_B$, or $q(\Delta n(x)/\Delta J_C - \Delta p(x)/\Delta J_C)$.

To summarize the discussed partitioning schemes, Fig. 5.3 shows the integrated concentrations with regard to electrons and holes for a typical abrupt HBT (the structure of explained in Sec. 5.5), as well as the edges of the base-emitter quasi-

neutral emitter by different methods, discussed in this section. For the case of the integrated charge method two different criteria (0.01 and 0.05) are investigated. It is seen that the estimates for the width of the quasi-neutral emitter using these two criteria are about 100Å apart, but the deviation on the position of the base side of the the base-emitter space-charge region is just about 10Å. One can see how poor is the estimate of the depletion approximation. Also, worthy of note is that at this rather high bias, van den Biesen's x^*_{BE} lies exactly on the metallurgical junction at $x = 0.40\mu$m. Its estimate on the emitter side is even smaller than the depletion approximation.

The ambiguity in defining the edge of quasi-neutral emitter by any applied method, strongly supports the idea that there is no way to define the width of this region. It is basically because of the violation of the depletion approximation in the emitter side of the base-emitter space-charge region, a consequence of currents far from equilibrium. As seen later, fortunately it is not important to know this boundary for the case of practically useful graded HBTs, since the quasi-neutral emitter delay time is virtually zero. Finally, it is worthy of note that although the integrated charge method can be employed for estimating the edge of the base-collector space-charge region, allowing τ_C to be treated as two separate space-charge and quasi-neutral delays, since the reverse-biased quasi-neutral collector and sub-collector delay times are also virtually zero, there is no necessity to do this.

5.4 Compact Expressions for Signal-Delay Times

5.4.1 Emitter Delay Time

With regard to the employed integrated charge method regional scheme, τ_E consists of two components τ_{QNE} and τ_{EB} (see Fig. 5.1(a)):

Quasi-Neutral Emitter Delay Time

By employing Lundstrom's boundary conditions for heterojunctions [96], the evaluation of τ_{QNE} is straightforward. This has been done by Gao *et al.* [110], with the result that

$$\tau_{QNE} = G\Big(\frac{W_E W_B}{2D_{nB}}\Big), \tag{5.10}$$

where W_E and W_B are the widths of the quasi-neutral emitter and quasi-neutral base respectively, D_{nB} is the zero-field electron diffusivity in the base, and G is a weighing factor. For a homojunction device $G = 1$. For graded HBTs, $G = N_B/N_E e^{-q(E_{gE}-E_{gB})/k_B T}$, where N_E is the emitter doping density, E_{gE} and E_{gB} are the band-gap energies in the emitter and base, respectively. For abrupt HBTs, the same expression is valid for G, but with the multiplication of another factor, namely S_{dn}/S_{en}. S_{dn}, the electron minority carrier diffusion velocity in the base, is equal to D_{nB}/W_B. S_{en} is a velocity related to the drift and diffusion of electrons across the space-charge region barrier. Lundstrom has developed an analytical expression for this velocity, provided that the base and emitter portions of the space-charge region widths are much greater than the corresponding extrinsic Debye lengths in the regions. Otherwise, S_{en} needs to be calculated numerically [96].

The above formulation assumes a linear hole profile in the quasi-neutral emitter, *i.e.*, W_E is a small fraction of L_p, the hole diffusion length in the emitter. When

these lengths are comparable, one can employ the ordinary hyperbolic solution of the steady-state continuity equation governing the injected holes in the emitter. By integration of the profile, it can be shown in a straight-forward way that

$$\tau_{QNE} = \frac{qL_p n_{iE}^2}{g_m k_B T}\Big(\frac{1-\cosh(W_E/L_p)}{(1+\frac{S_{dp}}{S_{ip}}+\frac{S_{dp}}{S_{ep}})\sinh(W_E/L_p)}\Big)e^{qV_{BE}/k_BT}. \tag{5.11}$$

n_{iE} is the intrinsic carrier concentration in the emitter, and V_{BE} is the effective junction voltage. S_{dp} and S_{ep} are the hole effective junction velocities corresponding to those for electrons defined earlier. Since the valence band can be assumed continuous, S_{ip}, the interface velocity of holes, is equal to v_{th}, the thermal velocity, which is sufficiently large that S_{dp}/S_{ip} can be ignored.

Emitter-Base Space-Charge-Region Delay Time

Conventionally, the depletion approximation formulated by Shockley in his historical article [111] is used widely for τ_{EB} in both homojunction and heterojunction devices, *i.e.*,

$$\tau_{EB} = \frac{\epsilon_E A}{g_m W_{BE}}. \tag{5.12}$$

Physically, it is the parallel-plate capacitance formula, the separation of the imaginary plates, W_{BE}, calculated from the depletion approximation. For the case of heterojunctions, a modified version that takes the difference of permittivities into account is available [7] (Chap. 9, p. 286), which is what meant by the (ordinary) depletion approximation in the present work. Nevertheless, it is well known that this approach is only a good approximation for reverse biases or low forward biases [113]. The reason for the violation of this approximation for moderate to high biases is the presence of large amount of mobile carriers in the so called "depleted" region. Therefore, it cannot be assumed depleted anymore, and "space charge region" seems

to be a more reasonable term for this region. Enormous efforts have been made to take into account this charge in estimating the corresponding delay time. They have been summarized in a tutorial by Liou and Lindholm [108]. Basically, it is difficult to deal with this delay time analytically, and most of the published works are numerical. Any kind of analytic approach has to use some simplifying assumptions, and often the resulting error is of the same order as the charge concerned.

In the following discussion, adapting Lindholm's expression [112], we write

$$C_{je} = C_D + C_F \equiv FC_D, \tag{5.13}$$

where C_D is the traditional depletion-region capacitance due to the modulation of the space-charge region width, and C_F is due to the free carriers in the region, and F is a correction factor of this last quantity.

One of the very few published works to take into account the free charge in an abrupt heterojunction is due to Liou *et al.* [114]. With the assumption of a piece-wise linear intrinsic Fermi-level, E_i, they find a bulky expression for the corresponding C_F applicable for moderate biases. It is worth mentioning that Shirts and Gordon [115], have done another approximate analytic improvement for the case of abrupt heterojunctions by means of the unconventional "dielectric response function". One can imagine how complicated any attempt would be to deal with the free carriers in the space-charge region of a graded heterojunction. The only work for graded structures is by Liou [75] (Chap. 3, p. 72), who has only managed to modify the depletion approximation by taking into account the effect of linear permittivity in solving Poisson's equation.

5.4.2 Base Delay Time

With regard to the law of junctions, and the assumption of linear electron profile in the quasi-neutral base, *i.e.*, neglecting recombination, it is quite straightforward to show that the quasi-neutral base delay time is [99] (Chap. 9, p. 236),

$$\tau_B = \frac{W_B^2}{2D_{nB}} + \frac{W_B}{v_{coll}}. \tag{5.14}$$

Basically this is the variation in the integration of the trapezoidal-like profile in Figs. 5.1(a) and 5.2. v_{coll} is the collecting velocity of electrons at the collector edge of the quasi-neutral base, *i.e.*, $J_C/qn(x_{pC})$. x_{pC} is the position of the relevant quasi-neutral base edge. For long-base devices, the second term can be ignored, but it is the dominant term in moderate and short base devices. This is both a simple and exact expression, as long as the correct v_{coll} is substituted. For homojunction devices, v_{coll} is identical with v_{sat}, the saturation velocity. But it has quite different values when band-gap narrowing, the nature of the emitter-base heterojunction, quasi-ballistic and hot electron effects in HBTs are considered (see Sec. 5.5.2).

5.4.3 Collector Delay Time

Similar to τ_{EB}, the simplest estimate of the collector delay time, τ_C, invokes the depletion approximation. This seems to be promising since in the active mode of transistor operation, the base-collector junction is reversed biased. However, this is not actually a good estimate for moderate to high biases. The reason for this is, again, the presence of a considerable electron charge in the space-charge region to provide $J_C \approx qn_C v_{sat}$. n_C is the electron concentration in the region. Since, it is assumed that the drift velocity in the region is the saturation velocity, n_C is constant with respect to position too. Regarding this mobile charge, another delay time component, namely

$W_{BC}/2v_{sat}$ (W_{BC} is the width of the region) is added to the depletion approximation expression. In the development of this additional term, it is conventionally assumed that the terminal base-collector voltage, V_{CB}, is held constant [99] (Chap.9, p. 238). This approach is not followed here, so allowing the complete expression for τ_C to be developed, including the effect of parasitics.

The total net charge in the space-charge region is $Q_C = qA(N_C - n_C)W_{BC}$, where N_C is the collector doping concentration. It has been assumed that the minority carrier concentration $p(x)$ in the region is negligible, which is reasonable for a reverse-biased junction. We can write the variation of Q_C in response to $\delta V_{BE} = \Delta V_{BE}$ as

$$\delta Q_C = qA(-W_{BC}\delta n_C + (N_C - n_C)\delta W_{BC}). \tag{5.15}$$

Integrating Poisson's equation, $\nabla^2\psi = -q(N_C - n_C)/\epsilon_C$, twice from the base-collector junction to W_{BC}, *i.e.*, neglecting the voltage drop in the base, we can write

$$V_{bi} + V_{CB} - J_C(W_C - W_{BC})/\sigma_C = q(N_C - n_C)W_{BC}^2/2\epsilon_C, \tag{5.16}$$

where ϵ_C is the permittivity in the collector, W_C is the physical length of the collector, and σ_C is the conductivity of the collector material, which allows the ohmic drop in the un-depleted collector, $J_C(W_C - W_{BC})/\sigma_C$, to be taken into account. Thus

$$\delta W_{BC} = \frac{\delta V_{CB} + qW_{BC}^2\delta n_C/2\epsilon_C - (W_C - W_{BC})\delta J_C/\sigma_C}{-J_C/\sigma_C + q(N_C - n_C)W_{BC}/\epsilon_C}. \tag{5.17}$$

If we assume that the ohmic voltage drop in the space-charge region is negligible compared to the junction potential difference, *i.e.*, $J_C W_{BC}/\sigma_C \ll q(N_C - n_C)W_{BC}^2/2\epsilon_C$, we can ignore the J_C/σ_C term in the denominator of the last equation, and conclude

$$|\delta Q_C| = \frac{qAW_{BC}\delta n_C}{2} - \frac{\epsilon_C A}{W_{BC}}\Big(\delta V_{CB} - (W_C - W_{BC})\frac{\delta J_C}{\sigma_C}\Big). \tag{5.18}$$

Furthermore, taking the emitter parasitic resistor into account, $V_{CE} = V_{CB} + V'_{BE} + (J_C + J_B)W_E/\sigma_E$, where V'_{BE} is the base-emitter *junction* voltage. So,

$$\frac{\delta V_{CB}}{\delta J_C} = \frac{-\delta V'_{BE}}{\delta J_C} - \frac{W_E \delta(J_B + J_C)}{\sigma_E \delta J_B} = -A/g_m - \frac{W_E}{\sigma_E}(1 + \frac{1}{h_{fe}}). \tag{5.19}$$

Assuming $h_{fe} \gg 1$, we conclude

$$\tau_C = \left|\frac{\delta Q_C}{\delta I_C}\right| = \frac{\epsilon_C A}{W_{BC}}\Big(\frac{1}{g_m} + \frac{W_E}{A\sigma_E} + \frac{W_C - W_{BC}}{A\sigma_C}\Big) + \frac{W_{BC}}{2v_{sat}}. \tag{5.20}$$

The $\epsilon_C A/g_m W_{BC} \equiv g_m^{-1} C_\mu$ term is the only one appearing in Eqn. (5.1) as the depletion approximation. The collector and emitter parasitic resistors, R_C, and R_E, are usually defined as $(W_C - W_{BC})/A\sigma_C$, and $W_E/A\sigma_E$, respectively, leading to the compact form

$$\tau_C = C_\mu\Big(\frac{1}{g_m} + R_E + R_C\Big) + \frac{W_{BC}}{2v_{sat}}. \tag{5.21}$$

If the depletion approximation is not valid for the biases of interest, its estimation for the corresponding width, W_{BC}, is not either, *i.e.*, the effect of the mobile carriers in the space-charge region should be considered. This has been done thoroughly by Kirk [116], in the early 1960's, and the result is as follows:

By defining two threshold currents $J_1 = qN_C v_{sat}$ and $J_2 = qN_C \mu_n V_{CB}/W_C$ (μ_n is the mobility of electrons) one can find different regimes of operation according to the relative magnitude of these currents and J_C [117]. J_1 is the threshold current density at which the ionized static charges in the space-charge region are completely compensated by electrons. J_2 is the threshold current density at which the bipolar transistor will enter the quasi-saturation regime. For typical devices of our interest $J_C < J_1 < J_2$ or $J_1 < J_C < J_2$. For the latter case which is commonly known as the Kirk effect, it is assumed that $W_{BC} = W_C$. For the former regime, the modification of the depletion approximation is

$$W_{BC} = W^0_{BC}\sqrt{(1 - J_C/J_2)/(1 - J_C/J_1)}. \tag{5.22}$$

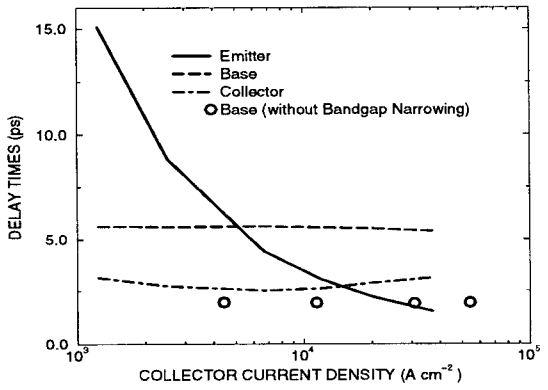

Figure 5.4: Simulated τ_E (solid line), τ_B (dashed line),and τ_C (dot-dashed line) delay times versus J_C for the studied graded HBT. The huge effect on τ_B, by ignoring band-gap narrowing is also shown (circles).

W^0_{BC} is the space-charge region thickness when $J_C = 0$. One should be careful as to what it means. It does not mean the width in equilibrium, but the ordinary thickness due to the depletion approximation at the related voltage bias. Typically $J_2 \gg J_1$, therefore one can drop the J_2 term in the numerator of Eqn. (5.22), or more simply, just modify N_C to $N_C - J_C/qv_{sat}$ in the ordinary depletion approximation formulation [118].

5.5 Analytic versus Numerical Delay Times

For the purpose of comparing analytical and simulated delay times n-$Al_{0.3}Ga_{0.7}As$/p-GaAs/n-GaAs HBTs with the following configuration were studied: the emitter, base, collector and sub-collector dopings were 6×10^{17}, 3×10^{19}, 4×10^{16}, and 4×10^{18} cm^{-3} respectively. The corresponding metallurgical widths were 0.40, 0.10, 0.40, and 0.30 μm. For the case of graded-emitter devices, a 200 Å linearly graded layer was modeled in the simulations, *i.e.*, the ungraded emitter width was 0.38 μm . For all the

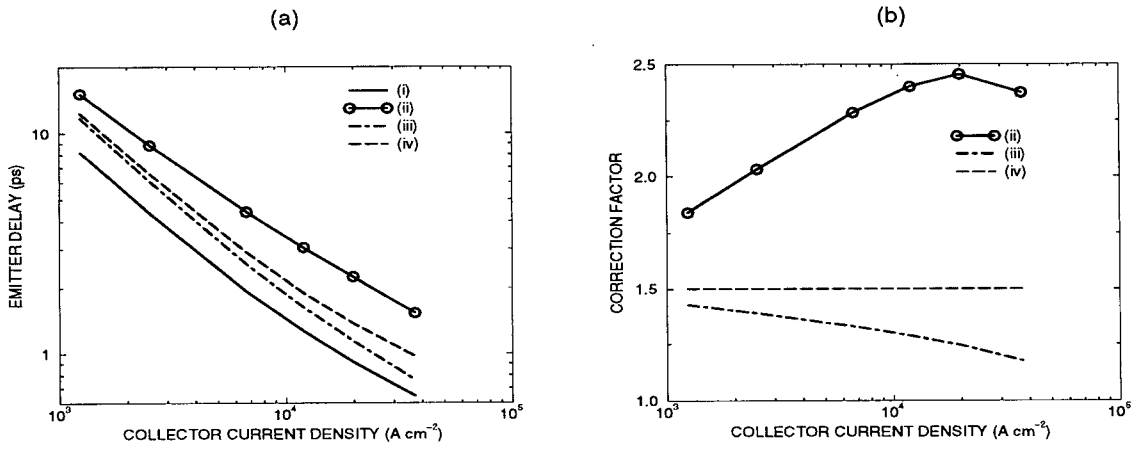

Figure 5.5: (a) τ_E versus J_C plots for a graded HBT: (i) Eqn. (5.12) (solid line); (ii) Simulation (solid-circle line); (iii) Depletion approximation for graded junctions (dot-dashed line), ([75], p. 72); (iv) Analytic expression to add the free electron charge in the space-charge region to (i) (dashed line), [114]. (b) Correction factor, F, imposed over the ordinary depletion approximation ($F = 1$), applied to (ii), (iii), and (iv) as noted in part (a).

simulations, V_{CE} was held constant at 2.0 V, and J_C was varied typically in the range of 1×10^3 to 1×10^5 A cm^{-2}. The ΔV_{BE} imposed for variational analysis was 0.0005 V (less than 2.0% of $k_B T/q$ at room temperature). Fermi-Dirac statistics, and complete ionization were assumed in all the simulations. With regard to band-gap narrowing, field-dependent mobility, and recombination, the enhanced models described in Ref. [102] were utilized. This is not true for the saturation velocity where the default value of the employed drift-diffusion simulator, MEDICI [98], was used. Regarding the g_m required for analytic calculations (see Eqn. (5.20) for instance), no analytic model was employed to calculate the current. Instead, for consistency of the two approaches, the numerically calculated g_m was utilized in the analytic expressions. Unless otherwise stated, the 1% integrated charge method criteria was used to regionalize the device numerically.

To have an idea of the relative magnitudes of τ_E, τ_B and τ_C, the simulated

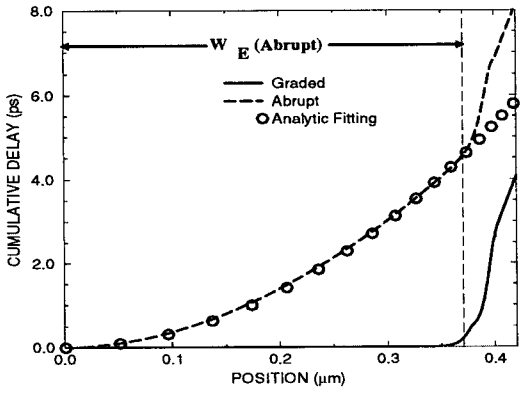

Figure 5.6: Cumulative delay $q\int_0^x \Delta n(x')dx'/\Delta J_C$ versus x, in the τ_E regime of graded (dashed line) and abrupt (solid line) HBTs for $J_C \approx 1\times 10^4$A cm^{-2} with the analytic expression for $q\int_0^x \Delta p(x')dx'/\Delta J_C$ form Eqn. (5.23) (circles), and with W_E evaluated by the 1% integrated charge method.

results of the graded HBT are plotted in Fig. 5.4. As can be seen, for current densities above about 1×10^4 A cm^{-2}, the dominant term is τ_B. The maximum cut-off frequency, calculated by Eqn. (5.5) is about 18 GHz, and occurs at $J_C \approx 5\times 10^4$ A cm^{-2}. This is in reasonable agreement with the experimental results of a somewhat similar structure [102]. As shown in the same figure, switching the band-gap narrowing feature off in the simulation causes a dramatic decrease (over 60%) of τ_B, but basically does not affect the other two signal-delay times. This will be discussed in Sec. 5.5.2.

5.5.1 Emitter Delay Time

Fig. 5.5(a) shows the simulated plot of τ_E versus J_C for the graded device calculated by means of the integrated charge method. τ_{QNE} is virtually zero, as observed from the cumulative delay time in Fig. 5.6. This is in agreement with the analytic expression in Eqn. (5.10), for graded devices. Hence essentially all the emitter delay

time is contributed by τ_{EB}. Three available analytical expressions are plotted in Fig. 5.5(a), namely: the ordinary depletion approximation in Eqn. (5.12); the depletion approximation taking into account the position-dependent permitivitty in the graded region [75]; and also the analytic expression for abrupt heterojunctions by Liou *et al.* [114], which takes the mobile carrier in the space-charge region into account. One can observe that the use of the last two unwieldy expressions does not contribute any encouraging improvement after all. This leads us to the idea that it seems more convenient to use the widely-accepted ordinary depletion approximation, but with the correction factor F introduced earlier. The resulting numerical F values, as well as the ones for the two mentioned analytic approaches, are shown in Fig. 5.5(b). As more clearly depicted in this figure, the two unwieldy analytical approaches do not yield more than 50% improvement on the depletion approximation. However, it can be seen that the numerical F is generally speaking between 1.8 and 2.5. These values are in accord with the numerical and analytical calculations by Negus and Roulston [119]. Although their work is based on a completely different linearly-graded homojunction device, this coincidence suggests that values for $F \approx 2$ are typical for any kind of device. This simple conclusion seems to be an interesting result for practical engineering purposes, where simple and easy-to-use expressions like Eqn. (5.12), are preferred for quick, first-order designs.

For abrupt HBTs, τ_{QNE} is not negligible. The physical reason is that due to the blocking effect of the discontinuity in the conduction band of the abrupt HBT, it is necessary to increase V_{BE} compared with the graded device to pump the same amount of current into the collector. This causes a decrease in the barrier of the valence band, and, consequently, facilitates the injection of holes into the quasi-neutral emitter. The cumulative delay time of the abrupt HBT is also shown in Fig. 5.6.

Interpreting τ_E of the abrupt HBT is more difficult. The reason for this is two-fold. The ambiguity of defining a well-behaved boundary between the quasi-neutral emitter and the base-emitter space-charge region was described thoroughly before and is not discussed further here. With regard to the analytic expressions, we found out that there is a serious problem in estimating τ_{QNE}. Both the analytical expressions presented in this work in Eqn. (5.11), and by Gao *et al.* in Eqn. (5.10), are a function of W_E. But, we already know that there is no exact value for this width. Scrutinizing both the G factor expression for such devices and Eqn. (5.11) shows that they are functions of the drift-diffusion velocities in the space-charge region, namely S_{en} and S_{ep}. The analytical expressions developed by Lundstrom for these velocities are actually orders of magnitude lower than what is needed to give a reasonable fit between the analytical cumulative delay time curve and the simulated one. It was found out that a value of about 2×10^2 cm s^{-1} for S_{en} is adequate for a G factor in the biases around $J_C = 1 \times 10^4$ A cm^{-2}, whereas Lundstrom's analytic expression yields 2×10^4 cm s^{-1} for the same velocity. As mentioned before, this huge discrepancy is due to the assumption made in the derivation, *i.e.*, that the base and emitter portions of the space-charge region width be orders of magnitude larger than the corresponding extrinsic Debye lengths. This assumption is enormously violated in our cases. For instance, the ratio of the base-side width of the space-charge region to the corresponding extrinsic Debye length, $\sqrt{k_B T \epsilon_B / 2q^2 N_B}$, is 0.6, rather than being sufficiently greater than one, as required for the validity of the analytic expression.

Although one can deal with S_{en} and S_{ep} numerically, instead of by compact expressions (a fact that Lundstrom mentions himself too [96]), one can take the velocities in question as free parameters in employing either of the two analytic expressions in Eqns. (5.10) and (5.11), and have an estimate of them by fitting the analytical

and simulated cumulative delay times. For instance, if an estimate of S_{ep} is desired, by the valid assumption that $W_E \ll L_p$, the quadratic cumulative hole density that yields the delay in Eqn. (5.11) can be written as

$$\int_0^x \Delta p dx' = \left(\frac{n_{iE}^2}{2W_E N_E} . \frac{e^{qV_{BE2}/k_BT} - e^{qV_{BE1}/k_BT}}{(1 + \frac{S_{dp}}{S_{ip}} + \frac{S_{dp}}{S_{ep}})} \right) x^2. \tag{5.23}$$

V_{BE1}, and V_{BE2} are the primary and secondary base-emitter voltages, with the difference of ΔV_{BE}. The fitting of the corresponding cumulative delay with $S_{ep} = 6.7 \times 10^4$ cm s^{-1} is also shown in Fig. 5.6, where W_E in Eqn. (5.23) has been estimated by the 1% integrated charge method. It is clear that changing the criteria, will affect the estimate on S_{ep}. Obviously, because of these ambiguities in calculating τ_{QNE} for abrupt HBTs, trying to develop a correction factor for τ_{EB} in a similar manner to that presented for graded HBTs would be very questionable and better to be avoided.

5.5.2 Base Delay Time

As mentioned before, Eqn. (5.14) is an appropriate expression for τ_B, provided that the correct collection velocity at the quasi-neutral base edge is picked. The different conduction band diagram scenarios are shown schematically in Fig. 5.7. Diagram (a) is for a device without band-gap narrowing features. For this case, it was verified for the corresponding data presented in Fig. 5.4 that $v_{coll} = v_{sat}$. The saturation is because of the high electric field in the space-charge region, as first suggested by Roulston for homojunction devices [120]. This is true for either an abrupt- or a graded-emitter device. However, when band-gap narrowing is taken into account, the collection velocity is less than v_{sat}. The high doping of the base material causes the band-gap of the region be a few k_BT smaller than the corresponding one in the collector. Consequently, a pseudo-heterojunction is induced, creating a bottleneck

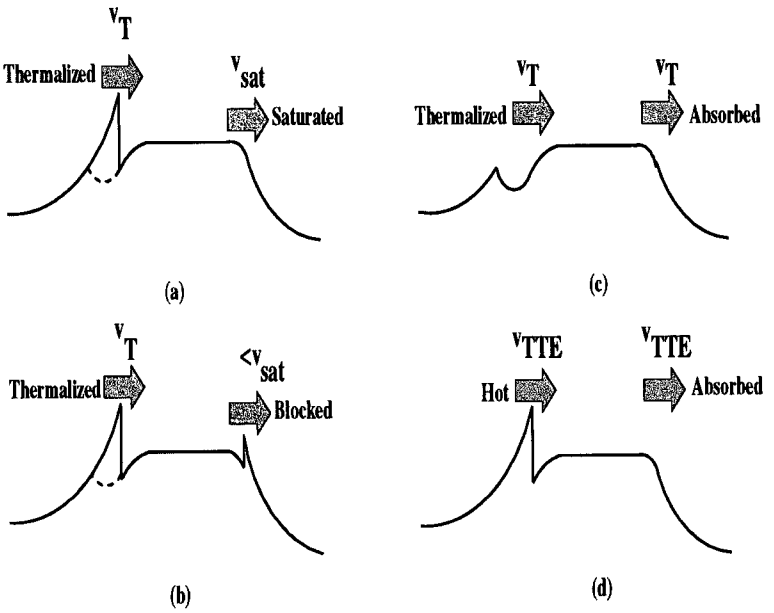

Figure 5.7: Different scenarios for the base collection velocity: (a) Long-base graded or abrupt HBT without band-gap narrowing effect, $v_{coll} = v_{sat}$; (b) Long-base graded or abrupt HBT with band-gap narrowing effect, $v_{coll} < v_{sat}$; (c) Short-base graded HBT, $v_{coll} \approx v_T$; (d) Short-base abrupt HBT, $v_{coll} \approx v_{TTE}$.

for the flow of the carriers. Therefore, the collection velocity is reduced significantly. This scenario is schematically elaborated in Fig. 5.7(b). The effect of the band-gap narrowing pseudo-heterojunction on the $\Delta n(x)$ profile is apparent in Fig. 5.2. For the case of the graded device, it turns out numerically that the velocity at the base edge of the space-charge region is about $0.28v_{sat}$, which gives an excellent agreement for evaluating τ_B, if this collection velocity is employed in the analytic expression.

As a final note on τ_B, it must be stressed herein that a drift-diffusion simulator, such as MEDICI, does not implement hot-electron and quasi-ballistic effects, which are important for short-base modern HBTs. The two right-hand side conduction band diagrams in Fig. 5.7 are for such devices. Diagram (c) is for the graded case, where the electrons are injected as usual into the base at the emitter-base heterojunction with the thermionic injection velocity $v_T = \sqrt{2k_BT/\pi m_e^*}$, where m_e^* is the electron

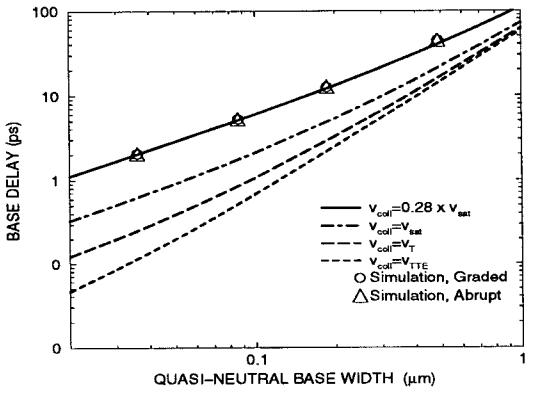

Figure 5.8: τ_B versus W_B: Simulation of graded HBT (circles), and abrupt HBT (triangles). Also shown are Eqn. (5.14) with different collection velocities: $v_{coll} = 0.28v_{sat}$ (solid line); $v_{coll} = v_{sat}$ (dotted-dashed line); $v_{coll} = v_T$ (long-dashed line); $v_{coll} = v_{TTE}$ (dotted line) [109].

effective mass. However, since the base width is comparable to the average scattering (or mean-free path) length [121], the electrons transit the base quasi-ballistically, and are collected with roughly the same velocity. For an abrupt short-base device, Fig. 5.7(d), where the electrons are injected into the base with the "hot" velocity v_{TTE} (determined by quantum-mechanical tunneling and thermionic emission), the collection velocity lies closer to the same value, rather than v_{sat}, again due to the quasi-ballistic transport across the base [121]. The plots of τ_B versus quasi-neutral base width calculated by MEDICI are presented in Fig. 5.8 [109], as well as the analytical signal-delay times with different velocities introduced here, namely v_{sat}, v_T, and v_{TTE}.

5.5.3 Collector Delay Time

The current-correction factor for the depletion-region width in Eqn. (5.22), developed for homojunction transistors, provides excellent agreement with the simulated results

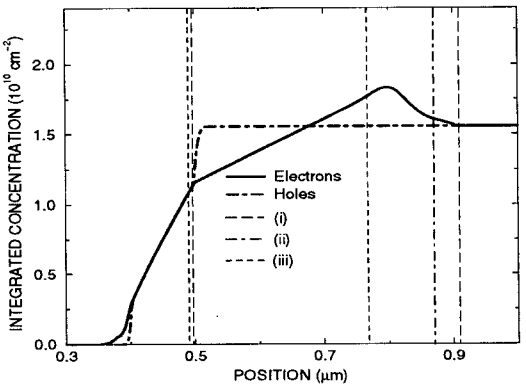

Figure 5.9: The cumulative integrals $\int_0^x \Delta n(x')dx'$ (solid curve) and $\int_0^x \Delta p(x')dx'$ (dot-dashed curve) of the studied graded HBT for $J_C \approx 2 \times 10^4$A cm^{-2}, with the estimates of the collector-base space-charge region width by different methods: (i) 1% criteria of the integrated charge method (long-dashed line); (ii) Eqn. (5.22) (dotted-dashed line); (iii) Depletion approximation (dotted line). Note: method (ii) and method (iii) give the same predictions at the base side of the space-charge region.

of HBTs, either graded or abrupt. The cumulative delay time of the graded device at a bias close to $J_C = 2\times10^4$ A cm^{-2} is shown in Fig. 5.9. Also added are the space-charge region edges predicted by the depletion approximation, the one modified by the Kirk effect, and the numerical integrated charge method. It can be seen that the depletion approximation underestimates the width, but the modified one is acceptably in accord with the numerical approach. It must be noted that the depletion approximation estimates a small shrinkage in the width as the bias is increased, but this is not actually the case in the simulation. Eqn. (5.17) is a mathematical explanation for that. Physically, since the increase in the current causes a decrease in the net charge, $q(N_C - n_C)$, in the space-charge region, W_{BC} has to increase for the satisfaction of Poisson's equation. This increase is also the reason for the negative integrand portion in the $\Delta n(x)/\Delta J_C$ plots in Fig. 5.2, another subtle feature of that plot. In low biases, this effect is not so apparent (as ignored in the simple picture of Fig. 5.1), but the

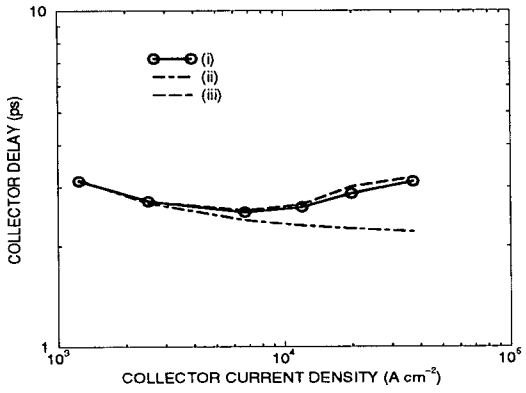

Figure 5.10: τ_C versus J_C plot for the graded HBT: (i) Simulation (solid line with circles). (ii) Eqn. (5.20) with the depletion approximation (dot-dashed line). (iii) Eqn. (5.20) with Kirk effect in Eqn. (5.22) (long-dashed line) [109].

negative values can be seen at moderate to high biases.

Fig. 5.10 [109], shows the numerical plots of τ_C versus bias. Also shown are the analytic plots when using either the depletion approximation or Eqn. (5.22), which takes into account the effect of n_C in estimating W_{BC}. This modification is often overlooked in calculating W_{BC}, assuming that it has no influence until after the onset of the Kirk effect at $J_C = J_1 \approx 6 \times 10^4$A cm^{-2}. However, the presented results show that the effect of n_C cannot be simply ignored at lower biases.

5.6 Signal-Delay Times in Nitride HBTs

The $\Delta n(x)$ and $\Delta p(x)$ plots versus x for the baseline AlGaN/GaN abrupt HBT of Sec. 4.4.2 are shown in Fig. 5.11(a) at room temperature, without any polarization feature and with $V_{BE} = 3.3$ V. Comparing with Fig. 5.2 leads us to observe that: (i) Observation (i) in Sec. 5.3.3 is violated, *i.e.*, $\Delta n(x) \neq \Delta p(x)$ in the quasi-neutral base; (ii) Observation (iii) in the same section is also violated. The cumulative charges due

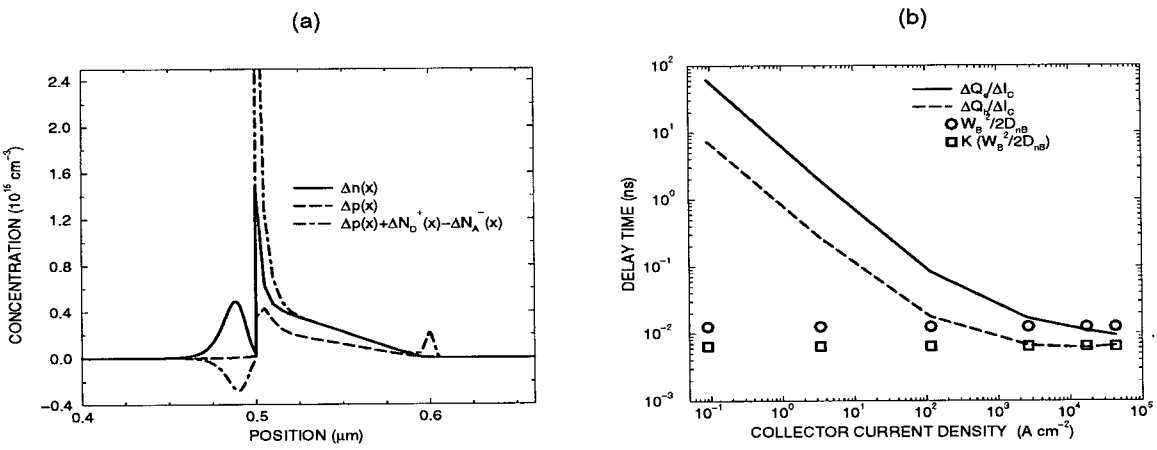

Figure 5.11: Charge control approach in the abrupt AlGaN/GaN HBT of Sec. 4.4.2 at room temperature. (a) Three incremental concentrations versus x at $V_{BE} = 3.3$ V : $\Delta n(x)$ (solid line); $\Delta p(x)$ (dashed line); $\Delta(p(x) + N_D^+(x) - N_A^-(x))$ (dot-dashed line). (b) Delay times versus J_C by: electron incremental charge, ΔQ_e, (solid line); hole incremental charge, ΔQ_h, (dashed line); Eqn. (5.14) (circles); Eqn. (5.14) times the K factor in Eqn. (5.27) (squares).

to holes and electrons are not equal in the quasi-neutral base, and the space-charge regions. For instance, the discussed equivalent interpretation of van den Biesens's τ_{eb}^* in terms of holes and electrons in Fig. 5.1(a) is no longer valid, as the relevant humps do not contain the same amount of charge. In addition, the emitter-to-collector signal-delay time depends on whether electrons or holes are picked. This is more clearly elaborated upon in Fig. 5.11(b), where $\Delta Q_e/\Delta I_C$ and $\Delta Q_h/\Delta I_C$ versus J_C are shown for the same device, where Q_e and Q_h are the electron and hole charges in the transistor, respectively. In other words, the most right-hand side identity in Eqn. (5.5) is not applicable. So, why is that and what is the delay of these devices?

The origin of all the observations is simply the incomplete ionization of dopants. As the bias is incrementally increased by ΔV_{BE}, more electrons are fed from the emitter lead, and more holes provided by the base lead are piled in the transistor, with distributions being dictated by Poisson's equation. Since the continuity equations are

solved simultaneously, charge neutrality of the whole device is a consequence of,

$$\int_0^L (p(x) - n(x) + N_D^+(x) - N_A^-(x))dx = 0. \tag{5.24}$$

If complete ionization of impurities are assumed, Eqn. (5.24) leads to $\int_0^L \Delta n(x) = \int_0^L \Delta p(x)$ as employed in Eqn. (5.5). However, if incomplete ionization of acceptors in the base is taken into account[5], $\Delta Q_e = \Delta Q_h - q\Delta N_A^-$, where

$$N_A^-(x) = \frac{N_A}{1 + g_A e^{(E_A(x) - F_p(x))/k_B T}} = \frac{N_A}{1 + \alpha p(x)} \tag{5.25}$$

and $\alpha \equiv g_A e^{(E_A(x) - E_V(x))/k_B T}/N_V$ is a constant defined for the simplicity of the following derivations. N_A^- changes if the hole quasi-Fermi level, $F_p(x)$, has a variation in response to ΔV_{BE}. Since the acceptors are only interacting with holes, as the right hand side of Eqn. (5.25) demonstrates, the increase in $p(x)$ will cause a negative ΔN_A^-, or $\Delta Q_e > \Delta Q_h$, as Fig. 5.11(b) verifies. It is the electron delay time which gives the correct τ_{EC}, and the equal amount of holes injected into the device are partly neutralized by ΔN_A^-. If one wants to use holes in the charge control theory, the ionized dopant charge must also be considered. Fig. 5.11(a) shows that if ΔN_A^- is subtracted from $\Delta p(x)$ in the base, the charge-neutrality concept is restored. Also, the addition of ΔN_D^+ to $\Delta p(x)$ shows that the contribution of the ionized donor charge is not very small in the emitter side of the emitter-base space charge region[6].

It is possible to find the ratio of the increment of hole charge in the quasi-neutral base, $\Delta Q_{h,B}$, over the electron charge, $\Delta Q_{e,B}$. Taking the differential of Eqn. (5.24), and assuming $N_D = 0$ in the base, gives

$$\Delta Q_{h,B} - \Delta Q_{e,B} = -q\alpha N_A \int_{x_{pE}}^{x_{pE}+W_B} \frac{\Delta p(x)}{(1 + \alpha p(x))^2} dx. \tag{5.26}$$

[5]The change in the donor ionized concentration, ΔN_D^+, should be incorporated in principle too.

[6]Actually, the negative hump should be subtracted from $\Delta n(x)$, if one wants to do a more exact analysis of the delay times.

With the approximation that the majority carrier concentration, $p(x) = \overline{p}$, is constant in the quasi-neutral base for a given bias, we can write the ratio K as

$$K = \frac{\Delta Q_{h,B}}{\Delta Q_{e,B}} = \left(1 + \frac{\alpha N_A}{(1+\overline{p})^2}\right)^{-1}. \tag{5.27}$$

In addition, if we assume that the majority carrier concentration does not differ from the equilibrium value $\overline{p}_{0B}$, *i.e.*, assuming $\overline{p} - \overline{p}_{0B} \ll \overline{p}_{0B}$, $K \approx 0.5$ for our structure. To compare that with the numerical results, the integrated charge method was applied to Δn and $\Delta(p - N_A^-)$. Eqn. (5.14) gives the same delay time as numerically calculated by $\Delta Q_{e,B}/\Delta J_C$, while $\Delta Q_{h,B}/\Delta J_C$ is half of that value, in excellent agreement with the above approximate analytic estimate of K. Qualitatively, similar arguments apply to the effect of incomplete ionization on the space-charge region delay times. There is one publication available on the influence of incomplete ionization on p-n symmetrical homojunction capacitances [122], but it appears to be difficult to quantize our non-symmetrical heterojunction case, in terms of analytic expressions.

The v_{coll} term in Eqn. (5.14) is negligible, since there is no band-gap narrowing feature in our nitride simulations. $W_B^2/2D_{nB}$ and $KW_B^2/2D_{nB}$ are also added to Fig. 5.11(b). The metallurgical width is substituted for W_B. As can be seen, they give reasonable agreement with the numerical results at high biases, where τ_B dominates τ_{EC}. Recall from Chap. 4 that the high level injection occurs at about 3-5 kA cm^{-2}, which is roughly at the same point that τ_B starts to dominate τ_{EC}. Although the high-level injection can give an improvement in τ_B [99] (Chap. 10, p. 254), to gain the best frequency performance in the ideal regime, the device should operate at a bias very close to the onset of high-level injection, *i.e.*, at $J_C = 2.6 \times 10^3$ A cm^{-2} in Fig. 5.11(a). The nitride HBT of our study has a cut-off frequency of about 10 GHz at this bias, which should be sufficient for some of the communications applications envisaged for AlGaN/GaN HBTs [24].

5.7 Summary

A comparison between the results of a drift-diffusion simulator for regional signal delays and traditional compact expressions of HBTs was performed.

- The suggested integrated charge method was found to be the most appropriate way of partitioning the device in terms of the quasi-neutrality definition, although no partitioning scheme to designate the boundary of the emitter space charge region and the quasi-neutral emitter exists.

- With regard to emitter signal-delay time, imposing a correction factor of about 2 on the ordinary depletion approximation appears to be the most convenient way to analytically estimate the emitter signal-delay time in a graded-emitter HBT. For an abrupt-emitter HBT, it is not feasible to separate the quasi-neutral emitter and base-emitter delay times, neither numerically nor analytically.

- The ordinary equation of the base signal-delay time gives an exact estimate of this component in the charge control theory for graded or abrupt HBTs, as long as the appropriate exit velocity is picked.

- The current-dependent modification of the depletion approximation should be taken into account for the collector signal-delay time, even at biases below the onset of the Kirk effect.

- The quasi-neutrality definition is not applicable for nitride HBTs, unless the variation of ionized acceptors in the base is also added to the variation of hole charge. Abrupt-emitter AlGaN/GaN HBTs are predicted to have cut-off frequencies as high as 10 GHz.

Appendix A

Wurzite Crystal Structure

Since the ordinary materials employed in the electronic and opto-electronic technology usually have diamond, *e.g.*, Si, or zincblende, *e.g.*, GaAs, structures, the crystallographic aspects of wurtzite materials are not widely known to the workers in this field, and a description of them is presented in this appendix.

The wurtzite structure is the penetration of two hexagonal closed-packed (hcp) structures of different atoms. It is also called the hexagonal zinc sulfide structure, since ZnS is one of the first materials known to have the structure. To understand the hcp structure in the first instance, we start with an analogy to the game of pool, or snooker. Imagining the two-dimensional arrangement of the balls that are placed as close as possible, *i.e.*, minimum interstitial area of the top view cross-section circles, in the triangle at the beginning of the game as a basal layer, let us stack some of these layers in three dimensions. To be more specific, we want to arrange identical spheres of atoms in that positioning which minimizes the interstitial volume, or in other words, makes a closely packed arrangement. One can perceive intuitively that this is the arrangement to which the interacting bonding forces are inclined to. This can be more technically justified in terms of minimizing the bonding energy per unit

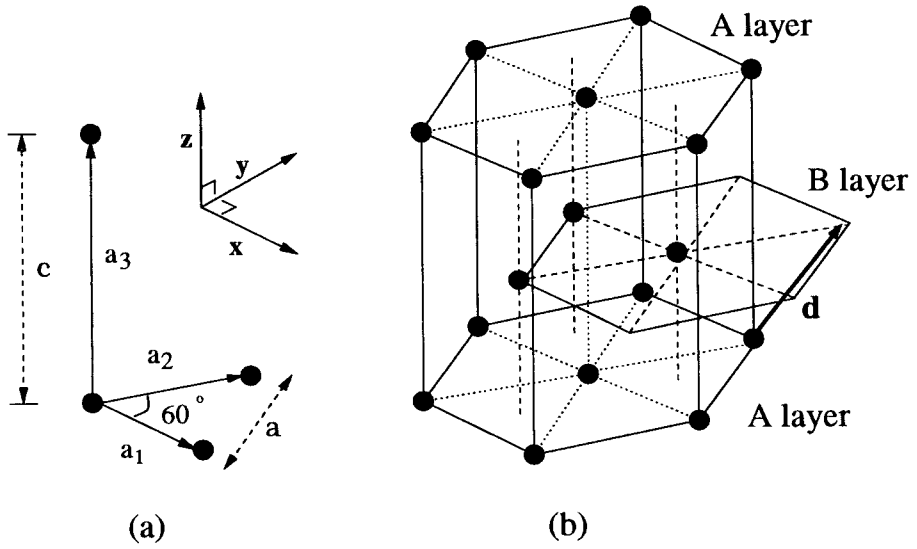

Figure A.1: (a) simple hexagonal primitive cell and vectors; (b) schematic of hexagonal close-packed lattice of two simple hexagons displaced by **d**, also showing two A and one B layers.

volume as a physical principle, provided that the forces are non-directional [2] (Chap. 2, p. 34). Going back to our billiards balls, one can simply visualize how the second layer of spheres (B) will slip in the depression in between of the first layer spheres (A) to satisfy this minimization. These overlayed depressions are not all the ones available between any three contacting spheres of the first layer, but half of them. Therefore, when we try to add a third layer we have two choices: one is on top of the other half depressions of the first layer (C), the other exactly on top of the first layer spheres (A). For the fourth layer we merely have one of the previous three choices A, B, or C. The random sequence of these triple layers can lead to an infinite number of arbitrary stacking orders. However, not all of them are periodic (leading to a repetitive unit cell), and not all of the periodic orders are common. The most common ones are .. AB AB .. (called hcp or 2H), and ... ABC ABC ... which is called face-centered cubic (fcc) or 3C, and ..... ABCBAC ABCBAC ..... simply called six-hexagonal order (6H). All of these structures have the minimum filling factor of the interstitial

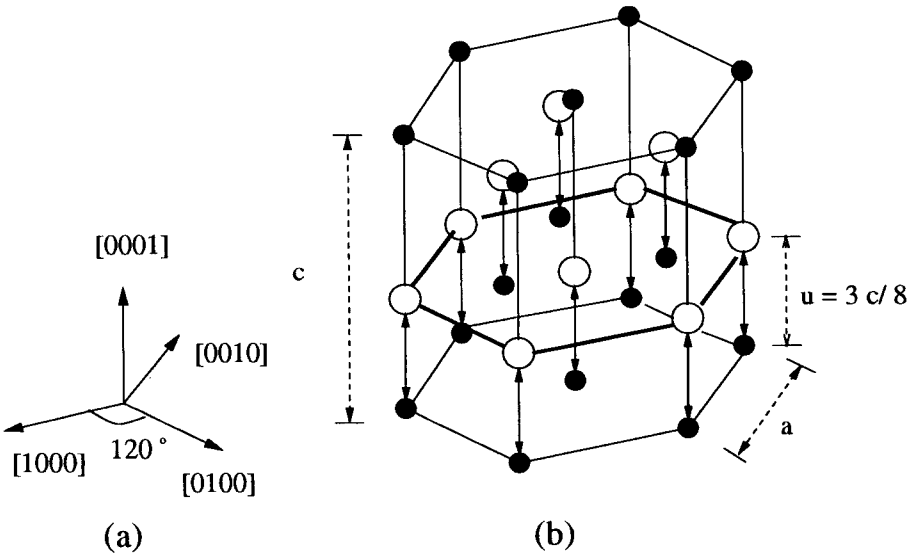

Figure A.2: (a) Four direction indices in hexagonal lattices; (b) wurtzite structure with two different atom sites (filled and hollow spheres) in hcp lattice, inter-penetrating with a separation u in the [0001] direction.

volume, which can be geometrically figured out as 74% [6] (Chap. 1, p. 23). fcc and hcp are the basis for the primitive cell of almost any semiconductor of interest. Two inter-penetrating fcc Bravais lattices lead to the diamond and zincblende structures, or rock-salt (NaCl) structure, depending on the displacement between the penetrating primitive cells. As mentioned, having nitrides in zincblende and rock-salt structures is feasible, but is not of device-engineering interest. Therefore, they are not discussed here further.

In contrast to fcc, the lattice of the hcp structure is not a Bravais one, *i.e.*, it is not the simplest repetitive unit of the lattice. The primitive unit cell is actually the simple hexagonal Bravais lattice, which has the primitive vectors $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$, with lattice sites at the corners of an equivalent triangle with side a, and another site at the perpendicular direction of the plane of the triangle above one of the other sites at distance c, as shown in Fig. A.1(a). hcp is basically the penetration of two of these Bravais lattices displaced by $\mathbf{d} = \mathbf{a}_1/3 + \mathbf{a}_2/3 + \mathbf{a}_3/2$, as elaborated in Fig.

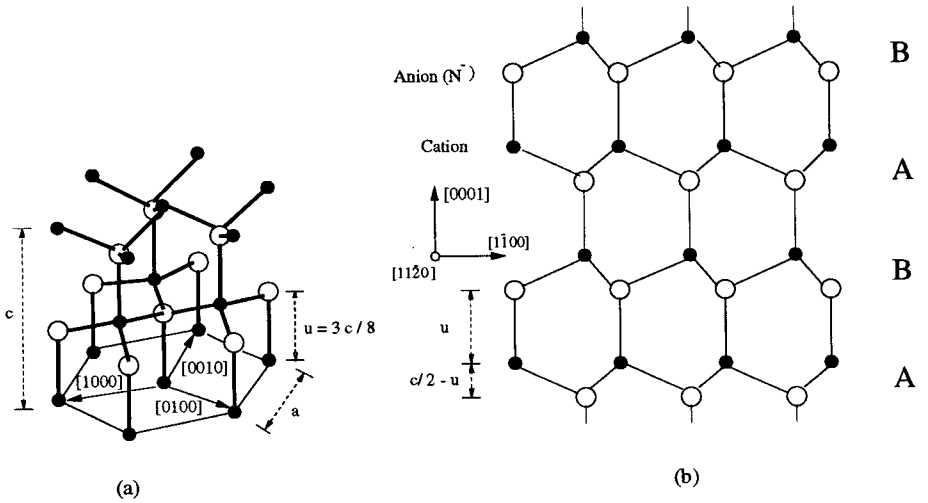

Figure A.3: Wurtzite structure: (a) tetrahedral bonds of the wurtzite lattice; (b) cross-section of wurtzite lattice with emphasis on [0001] convention in nitrides.

A.1(b). This is the one found by the AB AB order in our analogy of stacked billiard balls. To be consistent with that picture, there should be a certain ratio between a and c which turn outs to be $c/a = \sqrt{8/3} = 1.6333$ [6] (Chap. 1, p. 33).

Finally, the complex wurtzite structure consists of two inter-penetrating hcp lattices of different atoms, separated by $u\mathbf{z}$, as shown in Fig. A.2(b) . Ideally $u = 3c/8$ [7] (Chap. 1, p. 9), which is the displacement annihilating the electric force induced on any atomic site, provided that we assume dot-size atoms. Obviously, since $u \neq c/4$, the wurtzite structure does not have symmetry in the $\mathbf{a}_3$ direction. It is common practice in wurtzite literature *not* to use the three primitive vectors of Fig. A.1(a), but an equivalent four-component set shown in Fig. A.2(a). As can be seen, although two vectors are sufficient to describe any direction in the xy plane, three separated by 120° seem to be more convenient. The corresponding direction indices are [1000], [0100], and [0010], and of course have a 120° rotational symmetry. The c-axis is [0001], where due to the mentioned non-symmetrical property, is not equivalent to [000$\bar{1}$]. In other words, the c-axis is a polar axis in the wurtzite structure which causes

Table A.1: Lattice parameters of wurtzite nitrides.

Material	a(Å)	c(Å)	u(Å)	c/a	u/c
AlN	3.112	4.982	0.380	1.6010	0.076
GaN	3.186	5.185	0.376	1.6259	0.072
InN	3.54	5.705	0.377	1.6116	0.066
Ideal	–	–	–	1.6333	0.375

non-conventional properties in these materials. Fig. A.3(a) shows the same wurtzite lattice as Fig. A.2, but with a schematic of the tetrahedral bonds. Any atom has four covalent bonds with the atoms of the different type. Only one of the bonds is in the [0001] direction, which has the separation of u. One might be tempted to say that this is the same as the cubic structure, since these tetrahedral bonds remind one of the well-known similar ones in diamond. Nevertheless, the stacking order is very important here. The top layer of the three filled sphere layers in Fig. A.3(a) is identical with the bottom (both are A layers). To depict more clearly, Fig. A.3(b) presents a simple two-dimensional cross-section of the wurtzite structure. The AB AB layer stacking order is also shown, with each layer being in this case an actual bilayer of Ga and N atoms. This would not be the case in the diamond or the zincblende structures, where the stacking order is ABC ABC.

To become more specific in terms of nitrides, Fig. A.3 elaborates the symbols for the anion (N^{3-}) and cation (for example Ga^{3+}). This has been assumed in consistency with the direction *defined* as [0001] in the nitride literature. As can be seen, this is the one in which the cations have a single bond in that direction. In other words, *e.g.*, gallium atoms are on top of the close bilayer ideally separated by $c/2 - u = c/8$.

This is why the [0001] direction is also called *Ga-faced*, if we consider GaN. Flipping the crystal over leads us to the $[000\bar{1}]$ direction, also known as *N-faced*. The lattice constants a and c, and the anion-cation bond length in the [0001] direction, u, in unstrained equilibrium circumstances, are summarized in Table A.1 for various nitrides, borrowed from the review article by Ambacher [12]. As observed, there is a slight deviation in c/a from the ideal 1.6333 value. However, u/c is significantly smaller than the ideal 0.375 value. This is the effect of the non-zero radii of the atoms, as well as the electron orbitals, which are ignored in the geometrical, ideal calculation. This argument is more justified if we note that the ratio decreases as the atomic number of the cation increases. Meanwhile, by calculating the lattice mismatch between various nitrides, it can be deduced that it is large compared to other III-V compounds like the AlAs/GaAs system. Therefore, the heterostructures made out of nitrides could be strongly strained, provided that the lattice is not relieved. This fact, plus the instantaneous polarization (the effect of the aforementioned non-symmetry), and also the typically large piezoelectric constants in wurtzite structures, cause unusual effects on the characteristics of devices made out of nitrides, as discussed in Chap. 3.

Bibliography

[1] K. P. O'Donnell, "Beyond Silicon: The Rise of Compound Semiconductors," *Group III Nitride Semiconductor Compounds*, B. Gil Editor, Oxford University Press, Oxford, UK, pp. 1-18 (1998).

[2] S. Wang, *Fundamentals of Semiconductor Theory and Device Physics*, Prentice-Hall, Englewood Cliffs, NJ, (1989).

[3] J. I. Pankove, and T. D. Moustakas, "Introduction: A Historical Survey of Research on Gallium Nitride," *Gallium Nitride (GaN) I*, Semiconductors and Semimetals, vol. 50, J. I. Pankove, and T. D. Moustakas Editors, Academic Press, San Diego, CA, pp. 1-10 (1998).

[4] A. Trampert, O. Brandt, and K. H. Ploog, "Crystal Structure of Group III Nitrides," *Gallium Nitride (GaN) I*, Semiconductors and Semimetals, vol. 50, J. I. Pankove, and T. D. Moustakas Editors, Academic Press, San Diego, CA, pp. 167-192 (1998).

[5] G. Popovoco, H. Morkoç, and S. N. Mohammad, "Deposition and Properties of Group III Nitrides by Molecular Beam Epitaxy," *Group III Nitride Semiconductor Compounds*, B. Gil Editor, Oxford University Press, Oxford, UK, pp. 19-69 (1998).

[6] C. Kittel, *Introduction to Solid State Physics*, 5th Edition, John Wiley & Sons, New York, NY, (1976).

[7] C. M. Wolfe, Jr. N. Holonyak, and G. E. Stillman, *Physical Properties of Semiconductors*, Prentice-Hall, Englewood Cliffs, NJ, (1989).

[8] R. J. Molnar, "Hydride Vapor Phase Epitaxial Growth of III-V Nitrides," *Gallium Nitride (GaN) II*, Semiconductors and Semimetals, vol. 57, J. I. Pankove, and T. D. Moustakas Editors, Academic Press, San Diego, CA, pp. 1-31 (1999).

[9] S. P. DenBaars, and S. Kellar, "Metalorganic Chemical Vapor Deposition (MOCVD) of Group III Nitrides," *Gallium Nitride (GaN) I*, Semiconductors and Semimetals, vol. 50, J. I. Pankove, and T. D. Moustakas Editors, Academic Press, San Diego, CA, pp. 11-37 (1998).

[10] O. Briot, "MOVPE Growth of Nitrides," *Group III Nitride Semiconductor Compounds*, B. Gil Editor, Oxford University Press, Oxford, UK, pp. 70-122 (1998).

[11] T. D. Moustakas, "Growth of III-V Nitrides by Molecular Beam Epitaxy," *Gallium Nitride (GaN) II*, Semiconductors and Semimetals, vol. 57, J. I. Pankove, and T. D. Moustakas Editors, Academic Press, San Diego, CA, pp. 33-128 (1999).

[12] O. Ambacher, "Growth and Application of Group III-Nitrides," *Journal of Physics D: Applied Physics*, vol. 31, pp. 2653-2709 (1998).

[13] H. P. Maruska, and J. J. Tietjen, "The Preparation and Properties of Vapor-Deposited Single-Crystalline GaN," *Applied Physics Letters*, vol. 15, pp. 327-328 (1969).

[14] F. A. Ponce, "Structural Defects and Materials Performance of the III-V Nitrides," *Group III Nitride Semiconductor Compounds*, B. Gil Editor, Oxford University Press, Oxford, UK, pp. 123-157 (1998).

[15] O. Brandt, "Cubic Group III Nitrides," *Group III Nitride Semiconductor Compounds*, B. Gil Editor, Oxford University Press, Oxford, UK, pp. 417-459 (1998).

[16] K. Doverspike, and J. I. Pankove, "Doping in the III-Nitrides," *Gallium Nitride (GaN) I*, Semiconductors and Semimetals, vol. 50, J. I. Pankove, and T. D. Moustakas Editors, Academic Press, San Diego, CA, pp. 259-277 (1998).

[17] E. T. Yu, G. J. Sullivan, P. M. Asbeck, C. D. Wang, D. Qiao, and S. S. Lau, "Measurements of Piezoelectrically Induced Charge in GaN/AlGaN Heterostructure Field-effect Transistors," *Applied Physics Letters*, vol. 71, pp. 2794-2796 (1997).

[18] W. Götz, N. M. Johnson, C. Chen, H. Liu, C. Kuo, and W. Imler, "Activation energies of Si Donors in GaN," *Applied Physics Letters*, vol. 68, pp. 3144-3146 (1996).

[19] H. Amona, M. Kito, K. Hiramatsu, and I. Akasaki, "*p*-type Conduction on Mg-Doped GaN Treated with Low-Energy Electron Beam Irradiation (LEEBI)," *Japanese Journal of Applied Physics* , vol. 28, pp. L2112-2114 (1989).

[20] S. S. Chang, J. Pankove, M. Leksono, and B. Van Zeghbroeck, "500°C Operation of a GaN/SiC Heterojunction Bipolar Transistor," *Proceedings of Device Research Conference, 53rd Annual*, pp. 106-107 (1995).

[21] L. S. McCarthy, P. Kozodoy, M. J. Rodwell, S. P. DenBaars, and U. K. Mishra, "AlGaN/GaN Heterojunction Bipolar Transistor," *IEEE Electron Device Letters*, vol. 20, pp. 277-279 (1999).

[22] F. Ren, C. R. Abernathy, J. M. Van Hove, P. P. Chow, R. Hickman, J. J. Klaasen, R. F. Kopf, H. Cho, K. B. Jung, J. R. la Roche, R. G. Wilson, J. Han, R. J. Shul, A. G. Baca, and S. J. Pearton, "300°C GaN/AlGaN Heterojunction Bipolar Transistors," *MRS Internet Journal of Nitride Semiconductor Research*, vol. 3, Art. 41 (1998).

[23] S. Yoshida, and J. Suzuki, "High-temperature Reliability of GaN Metal Semiconductor Field-Effect Transistors and Bipolar Junction Transistors," *Journal of Applied Physics*, vol. 85, pp. 7931-7934 (1999).

[24] S. N. Mohammad, A. A. Salvador, and H. Morkoç, "Emerging Gallium Nitride Based Devices," *Proceedings of the IEEE*, vol. 83, pp. 1306-1355 (1995).

[25] S. Nakamura, T. Mukai, and M. Senoh, "High Power GaN *p-n* Junction Blue Light-Emitting Diode," *Japanese Journal of Applied Physics*, vol. 30, pp. L1998-2000, (1991).

[26] S. Nakamura, "III-V Nitride-based Short-wavelength LEDs and LDs," *Group III Nitride Semiconductor Compounds*, B. Gil Editor, Oxford University Press, Oxford, UK, pp. 391-416 (1998).

[27] J. J. Kuek, M. A. Wong, T. A. Fisher, B. D. Nener, and D. L. Pulfrey, "Modelling AlGaN Heterojunction Solar-blind Photodetectors," *Proceedings of Conference on Optoelectronic and Microelectronic Materials Devices*, IEEE, Perth, Australia, pp. 407-410 (1998).

[28] B. Monemar, "Fundamental Energy Gap of GaN From Photoluminescence Excitation Spectra," *Physical Review B*, vol. 10, p. 676 (1974).

[29] S. Strite, M. E. Lin, and H. Morkoç, "Progress and prospects of GaN and III-V Nitride Semiconductors," *Thin Solid Films*, vol. 231, pp. 197-210 (1993).

[30] V. W. Chin, T. L. Tansley, and T. Osotchan, "Electron Mobilities in Gallium, Indium, and Aluminum Nitrides," *Journal of Applied Physics*, vol. 75, pp. 7365-7372 (1994).

[31] S. N. Mohammad, and H. Morkoç, "Base Transit Time of GaN/InGaN Heterojunction Bipolar Transistors," *Journal of Applied Physics*, vol. 78, pp. 4200-4205 (1995).

[32] P. Kozodoy, M. Hansen, S. P. DenBaars, and U. K. Mishra, "Enhanced Mg Doping Efficiency in $Al_{0.2}Ga_{0.8}N$ Superlattices," *Applied Physics Letters*, vol. 74, pp. 3681-3683 (1999).

[33] S. Y. Chiu, A. F. M. Anwar, and S. Wu, "Base Transit Time in Abrupt GaN/InGaN/GaN and AlGaN/GaN/AlGaN HBTs," *MRS Internet Journal of Nitride Semiconductor Research*, vol. 4S1, Art. G6.7 (1999).

[34] W. Götz, N. M. Johnson, "Characterization of Dopants and Deep Level Defects in Gallium Nitride," *Gallium Nitride (GaN) II, Semiconductors and Semimetals*, vol. 57, J. Pankove, and T. Moustakas Editors, Academic Press, San Diego, CA, pp. 185-207 (1999).

[35] S. Dhar, and S. Ghosh, "Low Field Electron Mobility in GaN," *Journal of Applied Physics*, vol. 86, pp. 2668-2676 (1999).

[36] D. M. Caughey, and R. E. Thomas, "Carrier Mobilities in Silicon Empirically Related to Doping and Field," *Proceedings of the IEEE*, vol. 55, pp. 2192-2193 (1967).

[37] J. D. Albrecht, P. P. Ruden, E. Bellotti, and K. F. Brennan, "Monte Carlo Simulation of Hall Effect in n-Type GaN,". *MRS Internet Journal of Nitride Semiconductor Research*, vol. 4S1, Art. G6.6 (1999).

[38] T. Tanaka, A. Watanabe, H. Amano, Y. Kobayashi, I. Akasaki, S. Yamazaki, and M. Koike, "p-Type Conduction in Mg-Doped GaN and $Al_{0.08}Ga_{0.92}N$ Grown by Metalorganic Vapor Phase Epitaxy," *Applied Physics Letters*, vol. 65, pp. 593-594 (1994).

[39] M. Asif Khan, J. N. Kuznia, D. T. Olson, W. J. Schaff and J. W. Burm, and M. S. Shur, "Microwave Performance of a 0.25 μm Gate AlGaN/GaN Heterostructure Field Effect Transistor," *Applied Physics Letters*, vol. 65, pp. 1121-1124 (1994).

[40] U. V. Bhapkar, and M. S. Shur, "Monte Carlo Calculation of Velocity-Field Characteristics of Wurtzite GaN," *Journal of Applied Physics*, vol. 82, pp. 1649-1655 (1997).

[41] J. J. Barnes, R. J. Lomax, and G. I. Haddad, "Finite-Element Simulation of GaAs MESFET's with Lateral Doping Profiles and Sub-Micron Gates," *IEEE Transactions on Electron Devices*, vol. 23, pp. 1042-1048 (1976).

[42] M. Suzuki, J. Nishio, M. Onomura, and C. Hongo, "Doping Characteristics and Electrical Properties of Mg-Doped AlGaN Grown by Atmospheric-Pressure MOCVD," *Proceedings of 2nd International Conference on Nitride Semiconductors*, Tokushima , Japan, S-7 (1997).

[43] S. Strite, and H. Morkoç, "GaN, AlN, and InN: A Review," *Journal of Vacuum Science and Technology B*, vol. 10, pp. 1237-1266 (1992).

[44] W. R. L. Lambrecht, K. Kim, S. N. Rashkeev, and B. Segall, "Electronic and Optical Properties of the Group-III Nitrides, their Heterostructures and Alloys," *Material Research Society Symposium, Proceeding 395*, pp. 455-466 (1996).

[45] J. A. Majewski M. Stadele, and P. Vogl, "Electronic Structure of Biaxially Strained Wurtzite Crystals of GaN and AlN," *Material Research Society Symposium, Proceeding 449*, pp. 887-892 (1997).

[46] M. S. Shur, and M. Asif Khan, "GaN/AlGaN Heterostructure Devices: Photodetectors and Field-Effect Transistors," *Material Research Bulletin*, vol. 22, p. 44 (1997).

[47] D. L. Pulfrey, and B. D. Nener, "Suggestion for the Development of GaN-Based Photodiodes," *Solid-State Electronics*, vol. 42, pp. 1731-1736 (1998).

[48] Y. C. Yeo, T. C. Chong, and M. F. Li, "Electronic Band Structures and Effective-Mass Parameters of Wurtzite GaN and InN," *Journal of Applied Physics*, vol. 83, pp. 1429-1436 (1998).

[49] F. Bernardini, V. Fioreniti, and D. Vanderbilt, "Spontaneous Polarization and Piezoelectric Constants of III-V Nitrides," *Physical Review B*, vol. 56, R10024-10027 (1997).

[50] A. V. Dmitriev, A. L. Oruzheinikov, and M. V. Lomonosov, "The Rate of Radiative Recombination in the Nitride Semiconductor Alloys," *MRS Internet Journal of Nitride Semiconductor Research*, vol. 1, Art. 46 (1996).

[51] R. N. Hall, "Recombination Processes in Semiconductors," *Proceedings of the Institution of Electrical Engineers*, vol. 106B (Suppl. 17), pp. 923-931 (1960)

[52] W. Van Roosbroeck, and W. Shockley, "Photon-Radiative Recombination of Electrons and Holes in Germanium," *Physical Review*, vol. 94, pp. 1558-1560 (1954).

[53] O. Brandt, J. Ringling, K. H. Ploog, H. J. Wünsche and F. Henneberger, "Temperature Dependence of the Radiative Lifetime in GaN," *Physical Review B*, vol. 58, pp. R15977-15980 (1998).

[54] D. A. Ahmari, M. T. Fresina, Q. J. Hartmann, D. W. Barlage, P. J. Mares, M. Feng, and G. E. Stillman, "High-Speed InGaP/GaAs HBT's with a Strained $In_xGa_{1-x}As$ Base," *IEEE Electron Device Letters*, vol. 17, pp. 226-228 (1996).

[55] A. Hangleiter, J. S. Im, H. Kollmer, S. Heppel, J. Off, and F. Scholz, "The Role of Piezoelectric Fields in GaN-Based Quantum Wells," *MRS Internet Journal of Nitride Semiconductor Research*, vol. 3, Art. 15 (1998).

[56] T. Yamanaka, Y. Yoshikuni, K. Yokoyama, W. Lui, and S. Seki, "Theoretical Study on Enhanced Differential Gain and Extremely Reduced Linewidth Enhancement Factor in Quantum-Well Lasers," *IEEE Journal of Quantum Electronics*, vol. 29, pp. 1609-1616 (1993).

[57] A. Bykhovski, B. Gelmont, and M. Shur, "The Influence of the Strained-Induced Electric Field on the Charge Distribution in GaN-AlN-GaN Structure," *Journal of Applied Physics*, vol. 74, pp. 6734-6739 (1993).

[58] M. S. Shur, R. Gaska, and B. Bykhovski, "GaN-Based Electronic Devices," *Solid-State Electronics*, vol. 43, pp. 1451-1458 (1999).

[59] C. Kim, I. K. Robinson, J. Myoung, K. Shim, M. C. Yoo, and K. Kim, "Critical Thickness of GaN Thin Films on Sapphire," *Journal of Applied Physics*, vol. 69, pp. 2358-2360 (1996).

[60] H. Sugahara, J. Nagano, T. Nittono, and K. Ogawa, "Improved Reliability of AlGaAs/GaAs Heterojunction Bipolar Transistors with a Strain-Relaxed Base," *GaAs IC Symposium, 15th Annual*, Technical Digest, pp. 115-118 (1993).

[61] R. D. King-Smith, and D. Vanderbilt, "Theory of Polarization in Solids," *Physical Review B*, vol. 47, pp. 1651-1654 (1993).

[62] B. E. Foutz, O. Ambacher, M. J. Murphy, V. Tilak, and L. F. Eastman, "Polarization Induced Charge at Heterojunctions of the III-V Nitrides and their Alloys," *Physica status solidi (b)* vol. 216, pp. 415-418 (1999).

[63] A. Bykhovski, B. Gelmont, M. Shur, and A. Khan, "Current-Voltage Characteristics of Strained Piezoelectric Structures," *Journal of Applied Physics*, vol. 77, pp. 1616-1620 (1995).

[64] F. Jona, C. Shirane, *Ferroelectric Crystals*, Pergamon Press, New York, NY, (1962).

[65] J. F. Nye, *Physical Properties of Crystals*, Oxford University Press, Oxford, UK, (1957).

[66] J. R. Reitz, and F. J. Milforld, *Foundations of Electromagnetic Theory*, Addison Wesley, Reading, MA, (1960).

[67] 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. Reiger,

and J. Hilsenbeck, "Two-Dimensional Electron Gases Induced by Spontaneous and Piezoelectric Polarization Charges in N- and Ga-face AlGaN/GaN Heterostuctures," *Journal of Applied Physics*, vol. 85, pp. 3222-3233 (1999).

[68] O. Ambacher, R. Dimitrov, M. Stutzmann, B. E. Foutz, M. J. Murphy, J. A. Smart, J. R. Shealy, N. G. Weimann, W. J. Schaff, and L. F. Eastman, "Role of Spontaneous and Piezoelectric Polarization Induced Effects in Group-III Nitride Based Heterostructure and Devices," *Physica status solidi (b)*, vol. 216, pp. 381-389 (1999).

[69] O. Ambacher, B. Foutz, J. Smart, J. R. Shealy, N. G. Weimann, K. Chu, M. Murphy, W. J. Schaff, L. F. Eastman, R. Dimitrov, A. Mitchel, and M. Stutzmann, "Two-Dimensional Electron Gases Induced by Spontaneous and Piezoelectric Polarization Charges in Undoped and Doped AlGaN/GaN Heterostuctures," *Journal of Applied Physics*, vol. 87, pp. 334-344 (2000).

[70] E. S. Hellman, "The Polarity of GaN: a Critical Review," *MRS Internet Journal of Nitride Semiconductor Research*, vol. 3, Art. 11 (1998).

[71] R. Held, G. Nowak, B. E. Ishaug, S. M. Seutter, A. Parkhomovsky, A. M. Dabiran, P. I. Cohen, I. Grzrgory, and S. Porowski, "Structure and Composition of GaN(0001) A and B Surfaces," *Journal of Applied Physics*, vol. 85, pp. 7697-7704 (1999).

[72] H. Kroemer, "Theory of a Wide-Gap Emitter for Transistors". *Proceedings of the IRE*, vol. 45, pp. 1533-1537 (1957).

[73] H. Kroemer, "Heterojunction Bipolar Transistors and Integrated Circuits," *Proceedings of the IEEE*, vol. 70, pp. 13-25 (1982).

[74] D. L. Pulfrey, "Heterojunction Bipolar Transistors," *Wiley Encyclopedia of Electrical and Electronics Engineering*, vol. 8, J. G. Webster Editor, John Wiley & Sons, New York, NY, pp. 690-706 (1999).

[75] J. J. Liou, *Principles of AlGaAs/GaAs Heterojunction Bipolar Transistors*, Artech House, Boston, MA (1996).

[76] W. Liu, *Handbook of III-V Heterojunction Bipolar Transistors*, John Wiley & Sons, New York, NY, (1998).

[77] E. Alekseev, and D. Pavlidis, "DC and High-Frequency Performance of AlGaN/GaN Heterojunction Bipolar Transistors," *Solid-State Electronics*, vol. 44, pp. 245-252 (2000).

[78] T. P. Chow, and R. Tyagi, "Wide Bandgap Semiconductors for Superior High-Voltage Unipolar Power Devices," *IEEE Transactions on Electron Devices*, vol. 41, pp. 1481-1483 (1994).

[79] G. B. Gao, and H. Morkoç, "Material-Based Comparison for Power Heterojunction Bipolar Transistors," *IEEE Transactions on Electron Devices*, vol. 38, pp. 2410-2416 (1991).

[80] K. Shenai, R. S. Scott, and B. J. Baliga, "Optimum Semiconductors for High-Power Electronics," *IEEE Transactions on Electron Devices*, vol. 36, pp. 1811-1823 (1989).

[81] J. Pankove, S. S. Chang, H. C. Lee, R. J. Molnar, T. D. Moustakas, and B. Van Zeghbroeck, "High-Temperature GaN/SiC Heterojunction Bipolar Transistor with High Gain," *Proceedings of International Electron Device Meetings, 1994*, IEEE, pp. 389-392 (1994).

[82] J. I. Pankove, M. Leksono, S. S. Chang, C. Walker, and B. Van Zeghbroeck, "High-Power High-Temperature Heterobipolar Transistor with Gallium Nitride Emitter," *MRS Internet Journal of Nitride Semiconductor Research*, vol. 1, Art. 39 (1996).

[83] J. T. Torkiv, M. Leksono, J. I. Pankove, and B. Van Zeghbroeck, "A GaN/4H-SiC Heterojunction Bipolar Transistor with Operation up to 300°C," *MRS Internet Journal of Nitride Semiconductor Research*, vol. 4, Art. 3 (1999).

[84] B. Van Zeghbroeck, S. S. Chang, R. L. Waters, J. T. Torkiv, and J. I. Pankove, "GaN/SiC HBTs and Related Issues," *Solid-State Electronics*, vol. 44, pp. 265-270 (2000).

[85] J. T. Torkiv, J. I. Pankove, and B. Van Zeghbroeck, "GaN/SiC Heterojunction Bipolar Transistors," *Solid-State Electronics*, vol. 44, pp. 1229-1233 (2000).

[86] F. Ren, J. Han, R. Hickman, J. M. Van Hove, P. P. Chow, J. J. Klaasen, J. R. La Roche, K. B. Jung, H. Cho, X. A. Cao, S. M. Donovan, R. F. Kopf, R. G. Wilson, A. G. Baca, C. R. Abernathy, and S. J. Pearton, "GaN/AlGaN HBT Fabrication," *Solid-State Electronics*, vol. 44, pp. 239-244 (2000).

[87] B. S. Shelton, J. J. Huang, D. J. H. Lambert, T. G. Zhu, M. M. Wong, C. J. Eiting, H. K. Kwon, M. Feng, and R. D. Dupuis, "AlGaN/GaN Heterojunction Bipolar Transistors Grown by Metal Organic Chemical Vapour Deposition," *Electronics Letters*, vol. 36, pp. 80-81 (2000).

[88] X. A. Cao, J. M. Van Hove, J. J. Klaasen, C. J. Polley, A. M. Wowchack, P. P. Chow, D. J. King, F. Ren, G. Dang, A. P. Zhang, C. R. Abernathy, and S. J. Pearton, "High Temperature Characteristics of GaN-Based Heterojunction

Bipolar Transistors and Bipolar Junction Transistors," *Solid-State Electronics*, vol. 44, pp. 649-654 (2000).

[89] D. J. H. Lambert, D. E. Lin, and R. D. Dupuis, "Simulation of the Electrical Characteristics of AlGaN/GaN Heterojunction Bipolar Transistors," *Solid-State Electronics*, vol. 44, pp. 253-257 (2000).

[90] X. A. Cao, J. M. Van Hove, J. J. Klaasen, C. J. Polley, A. M. Wowchack, P. P. Chow, D. J. King, A. P. Zhang, G. Dang, C. Moniar, and S. J. Pearton, "Simulation of GaN/AlGaN Heterojunction Bipolar Transistors: Part I-*npn* Structure," *Solid-State Electronics*, vol. 44, pp. 1255-1259 (2000).

[91] X. A. Cao, J. M. Van Hove, J. J. Klaasen, C. J. Polley, A. M. Wowchack, P. P. Chow, D. J. King, A. P. Zhang, G. Dang, C. Moniar, and S. J. Pearton, "Simulation of GaN/AlGaN Heterojunction Bipolar Transistors: Part II-*pnp* Structure," *Solid-State Electronics*, vol. 44, pp. 1261-1265 (2000).

[92] L. S. McCarthy, P. Kozodoy, S. P. DenBaars, and U. K. Mishra, "A First Look at AlGaN/GaN HBTs," *25th International Symposium on Compound Semiconductors*, vol. 1, Nara, Japan, pp. 16-18 (1998).

[93] D. L. Pulfrey, "Modeling High-Frequency HBTs," *High-Speed Semiconductor Devices*, P. Roblin, and H. Rohdin Editors, Chap. 18, Cambridge University Press, New York, NY (to be published).

[94] S. Searles, D. L. Pulfrey, and T. C. Klencher, "Analytic Expressions for the Tunnel Current at Abrupt Semiconductor-Semiconductor Heterojunctions," *IEEE Transactions on Electron Device*, vol. 44, pp. 1851-1856 (1997).

[95] R. F. Pierret, " Advanced Semiconductor Fundamentals," *Modular Series on Solid-State Devices*, vol. VI, R. F. Pierret, G. W. Neudeck Editors, Addison-Wesley, Reading, MA, (1983).

[96] M. S. Lundstrom, "Boundary Conditions for *pn* Heterojunctions," *Solid-State Electronics*, vol. 27, pp. 491-496 (1984).

[97] H. K. Gummel, "A self-Consistent Iterative Scheme for One-Dimensional Steady State Transistor Calculations," *IEEE Transactions on Electron Device*, vol. 11, pp. 455-465 (1964).

[98] *Medici^{TM}, Release 4.1*, Avant! Corporation, Fremont, CA, (1998).

[99] D. J. Roulston, *Bipolar Semiconductor Devices*, McGraw-Hill, New York, NY, (1990).

[100] D. L. Pulfrey, and S. Fathpour, "Performance Predictions for *npn* $Al_xGa_{1-x}N$/GaN HBTs," *IEEE Transactions on Electron Devices*, submitted July 5, 2000.

[101] J. L. Moll, and I. M. Ross, "The Dependence of Transistor Parameters on the Distribution of Base Layer Resistivity," *Proceedings of the IRE*, vol. 44, pp. 72-78 (1956).

[102] D. L. Pulfrey and A. St. Denis, "Technical Report," The University of British Columbia and NORTEL Networks, (unpublished), July 24, 1996.

[103] J. Yoshida, M. Kurata, K. Morizuka and A. Hojo, "Emitter-Base Bandgap Grading Effects on GaAlAs/GaAs Heterojunction Bipolar Transistor Characteristics," *IEEE Transactions on Electron Devices*, vol. 32, pp. 1714-1721 (1985).

[104] J. J. van den Biesen, "A Simple Regional Analysis of Transit Times in Bipolar Transistors," *Solid-State Electronics*, vol. 29, pp. 529-534 (1986).

[105] C. Y. Kwok, "Space Charge and Injection Capacitance of *p-n* Junctions from Small Signal Numerical Solutions," *Journal of Physics D*, vol. 16, pp. L263-267 (1983).

[106] J. J. van den Biesen, "*P-N* Junction Capacitances, Part I: The Depletion Capacitance," *Phillips Journal of Research*, vol. 40, pp. 88-102 (1985).

[107] J. J. van den Biesen, "*P-N* Junction Capacitances, Part II: The Neutral Capacitance," *Phillips Journal of Research*, vol. 40, pp. 103-113 (1985).

[108] J. J. Liou and F. A. Lindholm, "Capacitance of Semiconductor *p-n* Junction Space-Charge Layers: an Overview," *Proceedings of the IEEE*, vol. 76, pp. 1406-1422 (1988).

[109] D. L. Pulfrey, S. Fathpour, A. St. Denis, M. Vaidyanathan, W. A. Hagley, and R. K. Surridge, "Applicability of the Traditional Compact Expressions for Estimating the Regional Signal-Delay Times of Heterojunction Bipolar Transistors," *Journal of Vacuum Science and Technology A*, vol. 18, pp. 775-779 (2000).

[110] G. B. Gao, J. I. Chyi, J. Chen, and H. Morkoç, "Emitter Region Delay Time of AlGaAs/GaAs Heterojunction Bipolar Transistors," *Solid-State Electronics*, vol. 33, pp. 389-390 (1990).

[111] W. Shockley, "The Theory of *p-n* Junctions in Semiconductors and *p-n* Junction Transistors," *Bell Systems Technical Journal*, vol. 28, pp. 435-489 (1949).

[112] F. A. Lindholm, "Simple Phenomenological Modeling of Transition-Region Capacitance of Forward-Biased *p-n* Junction Diodes and Transistor Diodes," *Journal of Applied Physics*, vol. 53, pp. 7606-7608 (1982).

[113] C. T. Sah, "Effects of Electrons and Holes on the Transition Layer Characteristics of Linearly Graded *P-N* Junctions," *Proceeding of the IRE*, vol. 49, pp. 603-618 (1961).

[114] J. J. Liou, F. A. Lindholm, and D. C. Malocha, "Forward-Voltage Capacitance of Heterojunction Space-Charge Regions," *Journal of Applied Physics*, vol .63, pp. 5015-5022 (1988).

[115] R. B. Shirts, and R. G. Gordon, "Improved Approximate Analytic Charge Distributions for Abrupt *p-n* Junctions," *Journal of Applied Physics*, vol .50, pp. 2840-2847 (1979).

[116] J. C. T. Kirk, "A Theory of Transistor Cutoff Frequency (f_T) Falloff at High Current Densities," *IRE Transactions on Electron Devices*, vol. 9, pp. 164-174 (1962).

[117] D. L. Bowler, and F. A. Lindholm, "High Current Regimes in Transistor Collector Regions," *IEEE Transactions on Electron Devices*, vol. 16, pp. 257-263 (1973).

[118] D. Rosenfeld, and S. A. Alterovitz, "The Composition Dependence of the Cut-off Frequency of Ungraded $Si_{1-x}Ge_x/Si_{1-y}Ge_y/Si_{1-x}Ge_x$ HBTs," *Solid-State Electronics*, vol. 38, pp. 641-651 (1995).

[119] K. J. Negus, and D. J. Roulston, "Simplified Modelling of Delays in the Emitter-Base Junctions," *Solid-State Electronics*, vol. 31, pp. 1464-1466 (1988).

[120] D. J. Roulston, "Low Current Base-Collector Boundary Conditions in GHz Frequency Transistors," *Solid-State Electronics*, vol. 18, pp. 845-847 (1988).

[121] D. L. Pulfrey, A. St. Denis, and M. Vaidyanathan, "Compact Modeling of High-Frequency Small-Dimension Bipolar Transistors," *Proceedings of Conference on Optoelectronic and Microelectronic Materials Devices*, IEEE, Perth, Australia, pp. 81-85 (1998).

[122] W. Nuyts, and R. Van Overstraeten, "Influence of the Incomplete Ionization of Impurities on the Capacitance of *p-n* Junctions," *Journal of Applied Physics*, vol. 42, pp. 3988-3990 (1971).