Design, Characterisation, and Numerical Simulation of Double Heterojunction Bipolar Transistors for Microwave Power Applications

by

MOHAMMAD SOTOODEH

Submitted in accordance with the requirement for the Degree of Doctor of Philosophy

KING'S COLLEGE LONDON
University of London
Department of Electronic Engineering
2000
ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisor Dr Ali Rezazadeh for his continual advice, encouragement and support throughout this project. Ministry of Culture and Higher Education of Iran is gratefully acknowledged for providing a scholarship. Part of this project was funded by Engineering and Physical Sciences Research Council (EPSRC) of UK. I would like to also acknowledge Epitaxial Products International Ltd. (EPI) of Cardiff for the supply of some of the epitaxial layers, and Dr. A. P. Knights from Surrey University for performing the ion implantation of samples. My thanks to Farid Amin, Hong Sheng, and Dr. Shabbir Bashar for their help and many useful discussions, and especially to Dr. Atta Khalid for fabricating the small-geometry HBTs. Mr Kevin Smith of King’s College London is also thanked for his technical support. Last, but not least, I am grateful to my wife, Afsaneh, for her love, patience and encouragement when it was most needed.

I dedicate this thesis to my mother, Mahindokht Jalali, and my late father, Ali Sotoodeh, to whom it means a lot. God bless them both.
GaAs-based heterojunction bipolar transistors (HBTs) are very attractive candidates for digital, analogue, and power applications due to their excellent switching speed combined with high current driving capability. Recently, there has been a widespread interest in replacing the conventionally used wide bandgap material Al\textsubscript{x}Ga\textsubscript{1-x}As with the more physically and technologically advantageous In\textsubscript{0.49}Ga\textsubscript{0.51}P. Additionally, employing a wide bandgap material as the collector of double HBTs (DHBTs) gives the designer another degree of freedom in tailoring the device behaviour for high power and/or high temperature applications.

In this thesis, InGaP/GaAs DHBTs are studied in details for microwave power applications. A comprehensive numerical simulation code is developed using FORTRAN 90 to investigate some important physical phenomena in HBTs. Particular attention is paid to tunnelling of electrons through conduction band potential barriers, and design issues of base-collector heterojunction in DHBTs. Material properties of a wide range of III-V compounds are studied extensively and very useful empirical relations are presented to model parameters like energy bandgaps, carrier mobilities, and minority carrier lifetimes.

DC and small-signal characterisation methods of fabricated HBTs are also considered in details. A novel technique is proposed to measure the base and collector series resistances of HBTs using only DC measured data. Also a new small-signal parameter extraction procedure is introduced in which all the equivalent circuit elements are extracted analytically without reference to numerical optimisation. This method is shown to have a wide range of applicability, which makes it appropriate for GaAs- and InP-based single and double HBTs. Moreover, a novel approach is presented to accurately determine the HBT total delay time and its components.

Both the vertical layer structure and horizontal layout of HBTs are optimised to improve the cutoff frequency-breakdown voltage product in the fabricated devices. These include design of a new mask set, optimisation of a multiple energy ion implantation process for device isolation to minimise device parasitics, proposing a new structure for the base-collector heterojunction of DHBTs, and improving the processing yield of the existing fabrication technology. As a result, DHBTs with simultaneously high cutoff frequency and maximum frequency of oscillation, large breakdown voltage, and low offset voltage are realised and demonstrated in this work.
# TABLE OF CONTENTS

1. INTRODUCTION ............................................................................................................. 9

**PART I: BACKGROUND THEORY AND LITERATURE SURVEY** ...... 14

2. THEORY OF HETEROJUNCTION BIPOLAR TRANSISTORS . 14
   2.1. The Concepts of Heterojunction Structures ................................................... 14
   2.2. Components of the Base Current ....................................................................... 19
      2.2.1. Base Bulk Recombination ....................................................................... 20
      2.2.2. Back Injection of Holes from Base to Emitter ...................................... 20
      2.2.3. Space-Charge Recombination .................................................................. 20
      2.2.4. Extrinsic Base Surface Recombination .................................................. 21
         2.2.4.a. Emitter-Size Effect ........................................................................... 22
         2.2.4.b. Emitter-Edge Thinning Design .......................................................... 23
   2.3. Abrupt versus Graded Base-Emitter Junctions ............................................ 24
   2.4. The Choice of Base Dopant ............................................................................. 27

3. HBT STRUCTURES ..................................................................................................... 29
   3.1. InGaP/GaAs HBTs ............................................................................................. 29
   3.2. Various HBT Configurations .......................................................................... 31
      3.2.1. Double HBTs ............................................................................................ 31
      3.2.2. Pnp HBTs .................................................................................................. 36
      3.2.3. Collector-Up HBTs .................................................................................. 37
      3.2.4. Graded Base HBTs ................................................................................... 37
   3.3. Alternative Materials for HBTs ....................................................................... 38
      3.3.1. InP/In(0.53)Al(0.47)As HBTs .................................................................. 38
      3.3.2. In(0.55)Al(0.45)As/In(0.55)Ga(0.47)As HBTs ......................................... 38
      3.3.3. InP/In(0.55)Al(0.25)Ga(0.75)As Continuous-Conduction-Band HBTs .... 39
      3.3.4. SiGe HBTs ................................................................................................ 39
      3.3.5. InP/Ga(0.5)Sb(0.4)/InP DHBTs ................................................................. 40
      3.3.6. Nitride-Based HBTs ............................................................................... 40
   3.4. Alternative Heterojunction Bipolar Device Structures .................................. 41
      3.4.1. Tunnelling Emitter Bipolar Transistors (TEBTs) .................................... 41
      3.4.2. Heterostructure-Emitter Bipolar Transistors (HEBTs) ......................... 41
      3.4.3. Other Heterojunction Bipolar Device Structures ................................... 43

**PART II- NUMERICAL DEVICE MODELLING** .......................................................... 44

4. BASIC SEMICONDUCTOR EQUATIONS ..................................................................... 45
   4.1. Basic Equations .................................................................................................. 45
   4.2. Boundary Conditions ....................................................................................... 51
      4.2.1. Emitter and Collector Ohmic Contacts .................................................. 51
      4.2.2. Base Contact .............................................................................................. 53
5. **PHYSICAL PARAMETERS IN III-V SEMICONDUCTORS**

5.1. Interpolation Schemes ................................................................. 54

5.2. Lattice Parameters and Dielectric Constant ............................... 56

5.2.1. Lattice Constant ........................................................................ 56

5.2.2. Dielectric Constant .................................................................. 57

5.3. Band Parameters ........................................................................... 58

5.3.1. Energy Bandgaps ....................................................................... 58

5.3.1.a. GaAs ......................................................................................... 58

5.3.1.b. AlAs and Al$_x$Ga$_{1-x}$As .......................................................... 60

5.3.1.c. InAs and In$_{1-x}$Ga$_x$As ............................................................ 63

5.3.1.d. In$_{1-x}$Al$_x$As ......................................................................... 65

5.3.1.e. InP ................................................................................................ 66

5.3.1.f. GaP and In$_{1-x}$Ga$_x$P ............................................................... 67

5.3.1.g. In$_{1-x}$Ga$_x$As$_y$P$_{1-y}$ ................................................................. 70

5.3.2. Unified Model for Electron Affinity .............................................. 72

5.3.2.a. GaAs and Al$_x$Ga$_{1-x}$As .......................................................... 74

5.3.2.b. In$_{1-x}$Ga$_x$As ......................................................................... 75

5.3.2.c. InP ................................................................................................ 76

5.3.2.d. In$_{1-x}$Ga$_x$P ......................................................................... 76

5.3.2.e. In$_{1-x}$Ga$_x$As$_y$P$_{1-y}$ ................................................................. 77

5.3.2.f. In$_{1-x}$Al$_x$As ......................................................................... 79

5.3.3. Effective Mass and Effective Density of States .......................... 80

5.3.3.a. Effective Mass of Binary Compounds ........................................ 81

5.3.3.b. Effective Mass of Ternary Compounds ..................................... 81

5.3.3.c. External Perturbations on Effective Mass ................................... 82

5.3.3.d. Conduction and Valence Band Density of States ...................... 83

5.4. Carrier Mobility ............................................................................ 85

5.4.1. Low-Field Mobility Model ......................................................... 85

5.4.2. Limitations of the Present Mobility Model ................................. 88

5.4.3. Fitting Parameters for Low-Field Mobility ................................. 90

5.4.3.a. GaAs ......................................................................................... 90

5.4.3.b. AlAs and Al$_x$Ga$_{1-x}$As .......................................................... 92

5.4.3.c. InP ................................................................................................ 94

5.4.3.d. GaP ............................................................................................ 94

5.4.3.e. In$_{1-x}$Ga$_x$P ......................................................................... 95

5.4.3.f. InAs ............................................................................................ 97

5.4.3.g. In$_{1-x}$Ga$_x$As ......................................................................... 97

5.4.3.h. In$_{1-x}$Al$_x$As ......................................................................... 97

5.4.3.i. In$_{1-x}$Ga$_x$As$_y$P$_{1-y}$ ................................................................. 98

5.4.4. High-Field Mobility Model ......................................................... 99

5.5. Recombination and Generation Mechanisms ............................... 103

5.5.1. Band-to-Band (Radiative) Recombination-Generation .................. 103

5.5.2. Auger Recombination ................................................................. 104

5.5.3. Indirect Thermal Recombination-Generation .............................. 105

5.5.4. Fitting Parameters for Recombination Lifetime ......................... 107

5.5.5. Interface Recombination ............................................................. 112

5.6. Bandgap Narrowing (BGN) .......................................................... 113

5.7. Ionised Doping Concentration ...................................................... 116
6. NUMERICAL METHODS

6.1. Normalisation of the Variables and Parameters

6.2. Discretisation of the Equations

6.2.1. Mesh Generation

6.2.2. Discretisation Scheme

6.2.3. Discretisation of the Boundary Conditions

6.3. Solution Strategies

6.4. Linearisation of the Equations

6.4.1. Newton and Newton-like Methods

6.4.2. Matrix Form of the Equations

6.4.3. The Initial Guess Problem

6.5. Solution of the Linear System of Equations

6.6. Terminal Characteristics

6.6.1. Calculation of Terminal Currents

6.6.2. Small-Signal Equivalent Circuit Parameters

6.7. Flowchart of the Completed Program

7. NUMERICAL SIMULATION RESULTS

7.1. Driving Force for Velocity Saturation

7.2. Thermionic-Field Emission (TFE)

7.3. Interface Recombination: Abrupt versus Graded HBTs

7.4. Temperature Dependence of Current Gain: HEBT versus HBT

7.4.1. Analysis of Temperature Dependence of β

7.4.2. Discussion of Various Base Current Components

7.4.2.a. SCR Recombination

7.4.2.b. Extrinsic Base Surface Recombination

7.4.2.c. Reverse Hole Injection

7.4.2.d. Base Bulk Recombination

7.4.3. HEBT versus HBT

7.4.4. Conclusions

7.5. Graded Base HBTs

7.5.1. Graded InGaAs Base

7.5.2. Graded AlGaAs Base

7.6. Capacitance Calculations

7.7. Kirk Effect

PART III- EXPERIMENTAL RESULTS

8. FAMILIARISATION WITH DEVICE TECHNOLOGY

8.1. Processing Steps for Large-Geometry HBT Fabrication

8.2. Ohmic Contact Formation and Characterisation

8.2.1. Ohmic Contacts

8.2.2. Transmission Line Model (TLM)

8.3. Chemical Wet Etching

8.4. Small-Geometry HBT Fabrication

8.4.1. Implant Isolation
8.4.2. Polyimide Processing ................................................................. 191
8.4.3. Mask Design ............................................................................. 192
8.4.4. Test Structures .......................................................................... 196

9. DC CHARACTERISATION OF HBTs .................................................. 199
  9.1. Room Temperature DC Characteristics of HBTs ......................... 200
    9.1.1. Output Characteristics ............................................................. 200
    9.1.2. Gummel Plots ......................................................................... 201
    9.1.3. Variation of Offset Voltage with $I_B$ ........................................ 201
    9.1.4. Extraction of the Series Resistances ......................................... 206
    9.1.5. Forward and Reverse Gummel Plots ........................................ 212
  9.2. Variation of HBT DC Characteristics with Temperature .............. 216
    9.2.1. Gummel Plot .......................................................................... 217
    9.2.2. Turn-on Voltage ...................................................................... 218
    9.2.3. Output Characteristics ............................................................. 220
    9.2.4. Breakdown Voltage .................................................................. 222
    9.2.5. Ideality Factors and Saturation Currents .................................... 222
    9.2.6. Current Gain ........................................................................... 224

10. HIGH FREQUENCY CHARACTERISTICS OF HBTs ....................... 228
  10.1. Small-Signal Equivalent Circuit of HBTs ..................................... 229
  10.2. High Frequency Small-Signal Measurement .................................. 232
  10.3. Parameter Extraction Procedure .................................................. 236
    10.3.1. Parasitic Pad Capacitance Determination .................................... 237
    10.3.2. Z-Parameters and Second-Order Approximations ..................... 241
    10.3.3. Cold-HBT Measurement ......................................................... 244
    10.3.4. Hot-HBT Measurement ............................................................ 245
    10.3.5. Comparison between the Measured and Calculated Data .......... 248
  10.4. High Frequency Figures of Merit ($f_T$ and $f_{max}$) ..................... 250
    10.4.1. Forward Transit Time .............................................................. 253
    10.4.2. Cutoff Frequency and Total Delay Time .................................... 257
    10.4.3. Unilateral Power Gain ............................................................ 259
  10.5. Bias and Geometry Dependence .................................................... 261
    10.5.1. Bias Dependence of the Parameters ........................................ 261
    10.5.2. Geometrical Scaling Rules ...................................................... 265
  10.6. Sensitivity Analysis ..................................................................... 270

11. OPTIMISATION OF THE B-C HETEROSTRUCTURE OF
    InGaP/GaAs DHBTs ........................................................................ 276
  11.2. Numerical Simulation of the Delay Times ..................................... 278
  11.3. Experimental Verification of an Improved Design InGaP DHBT ....... 283
    11.3.1. DC Results ........................................................................... 284
    11.3.2. High Frequency Performance ................................................ 286
12. SUMMARY AND SUGGESTIONS FOR FUTURE WORK...... 289

APPENDIX A. NETWORK PARAMETERS AND RELATIONSHIPS ........................................................................ 292

APPENDIX B. CALCULATION OF TUNNELLING INTEGRAL. 296

APPENDIX C. ANALYTICAL APPROXIMATIONS FOR FERMI-DIRAC INTEGRALS ........................................... 299

LIST OF PUBLICATIONS ......................................................................................................................... 301

REFERENCES .............................................................................................................................................. 302
CHAPTER 1. INTRODUCTION

The concept of heterojunction bipolar transistor (HBT) is almost as old as the bipolar transistor itself. The basic idea was described by Shockley in his patent for bipolar junction transistor (BJT) in 1948 (U.S. patent 2,569,347). Although the great advantages of such a design have long been recognised, no technology existed to build practically useful transistors of this kind until the early 70's, when liquid-phase epitaxy (LPE) emerged as a growth technology for III-V compound semiconductor heterostructures. Then in the mid-70's, two additional heterostructure technologies appeared: molecular beam epitaxy (MBE), and metal-organic chemical vapour deposition (MOCVD). Both technologies are capable of growing epitaxial layers with high crystalline perfection and purity, and high control over doping levels. Most important in the context of heterostructures, is the possibility in both technologies to change from one III-V semiconductor to a different (lattice-matched) one. Finally, in both techniques the growth rates and hence the layer thicknesses can be controlled very precisely, so that extremely thin layers of only few atomic monolayers can be achieved.

While the III-V HBT technology is relatively less mature in comparison with that of the Si BJT and GaAs field-effect transistor (FET) technologies, it offers unique performance advantages over both of these devices. Compared to Si BJTs, the GaAs- and InP-based HBTs present the following benefits: (1) high Early voltage due to higher base doping; (2) small base resistance as a result of higher base doping, leading to larger power gains and negligible emitter current crowding; (3) high electron mobility and velocity overshoot, reducing electron transit times; (4) low base-emitter junction capacitance, since emitter doping is lowered; (5) reduced parasitic capacitances, owing to the availability of a semi-insulating substrate; and (6) negligible high injection effects inside the base. The main area of concern for GaAs substrates, however, is the low thermal conductivity, which limits the maximum dissipated power density and integration level of GaAs circuits, and also raises some reliability issues.

HBTs also have many intrinsic advantages over FETs. These include: (1) The key distances that govern electron transit time are established by epitaxial growth, not by lithography. This allows high cutoff frequency, $f_T$, with modest processing requirements. (2) The entire emitter area conducts current, as opposed to thin horizontal channel of FETs, leading to high current-handling capability per unit area. (3) The threshold voltage for current flow is governed by built-in junction potential, leading to well-matched and uniform characteristics. (4) The transistor is well-shielded from traps in the bulk and surface regions, contributing to low 1/f noise and the absence of trap-induced frequency dispersion behaviour. (5) Breakdown voltage is directly controllable by the epitaxial structures of the device. (6) High transconductance, $g_m$, is achievable due to the exponential relationship between input voltage and output current. (7) Low output conductance, $g_o$, can be obtained owing to the high base doping concentration. This results in very high values of amplification factor, $g_m/g_o$. 

9
As a result, HBT technology has emerged as a key competitive and enabling technology for high speed digital and analogue applications. HBTs with peak $f_T$ in excess of 250 GHz [1] and maximum frequency of oscillation, $f_{max}$, approaching 1 THz [2] have already been demonstrated. Apart from military applications, the main areas where HBTs are competing for commercial products are as follows:

- **Microwave Power Amplifiers**: Owing to their high efficiency, high power density, high linearity, and low intermodulation distortion, HBTs are ideal candidates for power amplifiers in wireless handheld units as well as base-stations [3-4].

- **Optical Communication Systems**: High bit-rate fibre-optic communication systems require very high speed technologies for their implementation. At bit rates $>10$ Gb/s, HBTs are very much capable of offering a full chip-set solution [5-7]. These chip-sets typically include amplifiers, laser drivers, multiplexer/demultiplexers, and clock and data recovery circuits, which are already demonstrated for 40 Gb/s bit rates in several companies (for reviews, see [7-8] and references therein).

- **Data Conversion**: The large Early voltage, high $f_T$, excellent device matching, and compatibility with high performance Schottky diodes have made HBTs excellent candidates for analogue-to-digital (ADC), digital-to-analogue converters (DAC) and direct digital synthesisers [7, 9-10].

- **RF and Microwave Instrumentation**: HBT technology is well-suited for instrumentation applications, since it can simultaneously achieve broad bandwidth, low phase noise, excellent reliability, high gain and linearity [11]. Hewlett Packard has recently reported the use of a Darlington feedback amplifier based on InGaP/GaAs HBTs in numerous instrument applications, resulting in significant improvement of performance [11].

Conventionally HBTs were made based on the AlGaAs/GaAs material system, mainly due to the easy lattice-matching of $\text{Al}_x\text{Ga}_{1-x}\text{As}$ to GaAs with any mole fraction, $x$, of Al. Although early devices showed promising results, there are inherent disadvantages with this material system. More recently, $\text{In}_{0.49}\text{Ga}_{0.51}\text{P}$ (hereafter referred to as InGaP) has attracted considerable attention as an alternative to the AlGaAs emitter of GaAs-based HBTs. One of the main reasons for this interest is the much more favourable band line-up (i.e., smaller conduction and larger valence band offsets) of InGaP/GaAs heterojunctions compared to the AlGaAs/GaAs material system. Other advantages of InGaP over AlGaAs (which are discussed in detail in section 3.1 of this thesis) include high etch selectivity between InGaP and GaAs, much better reliability, lower impact ionisation rates, lower 1/f noise, absence of the so-called DX centres, and lower surface and interface recombination velocities. These advantages have resulted in superior DC [12-13] and RF [14-16] performances observed for InGaP/GaAs HBTs.

HBTs having a wide bandgap material in the collector as well as the emitter are called double HBTs (DHBTs). A wide bandgap collector offers numerous benefits for high power and high
temperature applications. These include larger breakdown voltage, lower leakage current due to the suppression of hole injection from base to collector, lower offset voltage owing to the symmetrical energy band structure between base-emitter and base-collector junctions, and shorter switching time transients. There is, however, an important restriction in the use of wide-gap collectors: the existence of an abrupt conduction band spike at the B-C heterojunction may impede the free collection of electrons. This may result in reduction of current gain, dramatic increase of the saturation (or knee) voltage, and degradation of the high frequency characteristics of DHBTs. Design of the B-C heterojunction of DHBTs is an issue thoroughly addressed in the present work.

In this thesis, design and optimisation of DHBTs for microwave power applications is presented with an emphasis on the InGaP/GaAs material system. Numerical simulation of basic semiconductor equations and small-signal modelling of HBTs, both completely and originally developed in the present work, are utilised as modelling tools for explaining device behaviour and optimising its structure. DC and RF characteristics of the InGaP/GaAs DHBTs, designed based on the above studies and fabricated in our research laboratory, are measured, and the capabilities of these devices for high power and high frequency applications are demonstrated. The original findings of this thesis are published in the form of 5 prestigious journal and 9 international conference proceeding articles, and 2 additional articles are currently under preparation. This work should play a major role in understanding and further improvement of HBT device behaviour for ever-growing areas of HBT applications.

This thesis consists of three parts, each incorporating a few chapters. Background theory of HBTs and a survey of published results in the literature are covered in Part I, which includes Chapters 2 and 3. The basic concept of HBTs, formulation of DC current gain and various components of base current, advantages of using InGaP both as the emitter and collector of (D)HBTs, as well as alternative configurations, material systems, and structures of HBTs are clearly and thoroughly addressed in this part.

Second part of this thesis deals with the numerical device modelling. Basic semiconductor equations, in the form of partial differential equations, and their boundary conditions are discussed in Chapter 4. These equations consist of various semiconductor parameters such as band parameters, carrier mobilities, and recombination lifetimes, which must be accurately known prior to the quantitative analysis of semiconductor devices. This is dealt with in Chapter 5, where a large number of published theoretical and experimental reports are reviewed to include the physical parameters for various III-V compounds in the computer program developed in this work. The III-V materials included in this chapter cover a very wide range: GaAs, InP, Al_{x}Ga_{1-x}As, In_{1-x}Ga_{x}P, In_{1-x}Ga_{x}As, In_{1-x}Al_{x}As, and even the quaternary material In_{1-x}Ga_{x}As_{y}P_{1-y} all for the entire range of composition. This not only gives the designer a good degree of freedom as to the choice of material, but also allows a direct comparison between various types of HBTs such as AlGaAs/GaAs, InGaP/GaAs,
Chapter 1: Introduction

InP/InGaAs, and InAlAs/InGaAs. The database of material properties developed in this chapter should also be extremely beneficial to the commercial device simulation packages such as ATLAS [17] and MEDICI [18], which are currently lacking such accurate information.

Once the semiconductor equations and their parameters are addressed, Chapter 6 discusses the numerical techniques to discretise, linearise, and solve those differential equations, as well as the methods to find the device terminal characteristics from the solutions. This is followed by demonstration of some of the results obtained from this numerical simulation program in Chapter 7. Studies of thermionic-field emission of carriers across an abrupt heterojunction interface, temperature dependence of current gain in various types of HBTs, optimising the performance of graded base (using both In$_{1-x}$Ga$_x$As and Al$_x$Ga$_{1-x}$As) HBTs, calculation of forward- and reverse-biased junction capacitances, and base push-out at high current densities, are among the examples investigated in this chapter. These are meant not only to verify the capability of the developed numerical code, but also to actually use it for the HBT structure optimisation, as when the optimum grading condition for graded base HBTs are found.

Part III of this thesis (experimental results) starts with the familiarisation with device fabrication technology in Chapter 8, where fabrication processes used in the present work for both large- and small-geometry HBTs are briefly explained. A large portion of this chapter belongs to the description of a new improved mask set designed in this work for fabrication of small-geometry HBTs.

Chapter 9 demonstrates some of the DC characteristics of HBTs with a main theme to highlight the advantages of InGaP/GaAs DHBTs over both InGaP/ and AlGaAs/GaAs single HBTs. Dependence of the offset voltage of HBTs on device structural and geometrical parameters as well as base current is investigated in this chapter. Then, after explaining the reason behind a high current phenomenon in HBTs, known as the kink effect, its occurrence is utilised in a novel technique to extract the collector and base series resistances of the device using only DC measurement results. Also the interrelation between the forward and reverse Gummel plots of HBTs is studied in great depth. Additionally, variation of HBT DC parameters with temperature are included in this chapter, where the capability of the numerical model to predict device behaviour at various temperatures is again confirmed.

Chapter 10, discussing the high frequency characteristics of HBTs, includes a major part of the work carried out in this research project. After a detailed description of HBT small-signal equivalent circuit models, an improved fully analytical technique is introduced which is capable of extracting all the elements of the equivalent circuit. Very reasonable approximations are used in the course of parameter extraction, which are shown to be valid for a wide range of bias and frequency, making the method applicable to various types of HBTs. Additionally, novel techniques are presented which extract the total delay time, forward transit time, and their components in HBTs using the measured s-parameter data only at
relatively low frequencies (typically < 5 GHz). The proposed methods are demonstrated to be extremely accurate, resulting in unique and physically meaningful values of parameters. This allows a direct observation of the bias dependence of small-signal equivalent circuit elements under the most stringent conditions such as when Kirk effect, self-heating, or current crowding has occurred.

Chapter 11 uses the understanding of the HBT device physics, gained throughout the earlier chapters, together with the numerical device model and small-signal parameter extraction technique developed in the present work, to optimise the DC and high frequency performance of InGaP/GaAs DHBTs. In particular, the design issues of the base-collector heterojunction of DHBTs is addressed. It is shown that with careful design of this heterojunction, the high frequency performance of DHBTs can be improved to the same level as in SHBTs, while the former still benefits from additional advantages such as a larger breakdown voltage and a lower offset voltage. Finally, this thesis is concluded in Chapter 12 with a summary of the main achievements in the present work and suggestions for future improvements.
PART I - BACKGROUND THEORY AND LITERATURE SURVEY

CHAPTER 2. THEORY OF HETEROJUNCTION BIPOLAR TRANSISTORS

This chapter discusses the background theory of HBTs and reviews the relevant work in the literature. The basic HBT theory and the advantages over the conventional homojunction bipolar transistor are first discussed in section 2.1. Then in section 2.2 components of the base current and their influence on limiting the current gain will be studied in more details. Section 2.3 deals with the comparison between the abrupt and graded base-emitter (B-E) heterojunction. Effects of different types of grading, ballistic transport in the base region of abrupt HBTs, and the thermionic-field emission across the abrupt heterointerfaces will be addressed in this section. This chapter concludes with a brief discussion on the choice of base dopant and its reliability issues in section 2.4.

2.1. The Concepts of Heterojunction Structures

In discussion of p-n junctions, it is usually assumed that the semiconductor material is homogeneous throughout the entire structure. This type of junction is called a homojunction. When two different semiconductor materials are used to form a junction, the junction is called a semiconductor heterojunction. Since the two materials used to form a heterojunction have different energy bandgaps, the energy bands will have discontinuities at the junction interface. We may have an abrupt junction in which the semiconductor changes abruptly from a narrow bandgap material to a wide bandgap material. On the other hand, in Al$_x$Ga$_{1-x}$As/GaAs system for instance, the value of $x$ may continuously vary over a distance of several nanometers to form a graded heterojunction. Changing the value of $x$ in Al$_x$Ga$_{1-x}$As system allows us to engineer or design the bandgap energy.

In order to have a good quality heterojunction, the lattice constants of the two materials must be well-matched, otherwise mechanical strain and dislocations are introduced which degrade the material quality. For example, Al$_x$Ga$_{1-x}$As is lattice-matched to GaAs for any mole fraction of aluminium. However, Al mole fractions above about 0.35 should be avoided since the material starts to become indirect bandgap at $x \approx 0.4$ causing a severe reduction in electron mobility. In$_{0.49}$Ga$_{0.51}$P/GaAs and In$_{0.55}$Ga$_{0.45}$As/InP are other examples of lattice-matched systems.

Fig. 2.1 shows a general ideal N-p heterojunction in thermal equilibrium. It can represent the band diagram for B-E junction of an abrupt AlGaAs/GaAs Npn HBT. The two materials have electron affinities $\chi_1$ and $\chi_2$, and energy bandgaps $E_{g1}$ and $E_{g2}$. When these two materials come together to form a junction, the conduction and valence bands are aligned so that the Fermi level is flat along the device. Therefore, there will be discontinuities in the conduction
Chapter 2 Theory of Heterojunction Bipolar Transistors

Figure 2.1- Energy band diagram for an abrupt N-p heterojunction at thermal equilibrium.

and valence bands:

$$\Delta E_C = \chi_2 - \chi_1$$  \hspace{1cm} (2.1)

$$\Delta E_V = \Delta E_g - \Delta E_C$$  \hspace{1cm} (2.2)

The relative band alignment of common heterojunction systems is schematically shown in Fig. 5.15 of chapter 5. The conduction band discontinuity is undesirable for Npn HBTs, since it limits the injection of electrons from the emitter to the base and consequently reduces the current gain. On the other hand, the valence band discontinuity should be as large as possible in order to provide a large barrier to prevent hole injection from the base into the emitter. For Al$_x$Ga$_{1-x}$As/GaAs system, majority of the bandgap difference appears as the conduction band offset: $\Delta E_C \approx (0.60-0.65)\Delta E_g$ [19] (see chapter 5 for details). This is one of the major drawbacks of this material system for use in Npn HBTs.

We will now look at the theory of the bipolar transistors to examine the effects and advantages of using a heterojunction rather than a homojunction. Figure 2.2 shows a schematic diagram of an n-p bipolar transistor in common emitter configuration operating in forward active mode (i.e., forward-biased B-E and reverse-biased B-C junctions). The forward-biased B-E junction causes an electron current $I_{bE}$ to be injected into the base, and a hole current $I_{bP}$ to be injected into the emitter. Some of the electrons injected from the emitter will recombine inside the B-E space-charge region (SCR), giving rise to a recombination current, $I_{b,SCR}$. Further electrons recombine with holes in the bulk of the base region ($I_{b,bulk}$), and some others at the extrinsic base surface ($I_{b,BS}$). The remaining electrons reached to the edge of the B-C depletion region, are swept to the collector under the influence of the high
electric field at the reverse-biased junction. Furthermore, there is an additional current component, \( I_{\text{CBO}} \), that arises from the generation of carriers in the B-C SCR. Therefore, the base, collector, and emitter currents can be written in terms of their components as:

\[
I_B = I_{\text{B,p}} + I_{\text{B,bulk}} + I_{\text{B,BS}} + I_{\text{B,SCR}} - I_{\text{CBO}}
\]

\[
I_C = I_{nE} - I_{\text{B,bulk}} - I_{\text{B,BS}} + I_{\text{CBO}}
\]

\[
I_E = I_B + I_C = I_{nE} + I_{\text{B,p}} + I_{\text{B,SCR}}
\]

The common-emitter DC current gain of a transistor is defined as the ratio \( I_C / I_B \). Usually, bipolar transistors are designed to have a gain of at least 10. In this case, the recombination and generation components of current are negligible compared to the electron current injected from emitter to base, \( I_{nE} \), and one can write:

\[
\frac{1}{\beta} = \frac{I_B}{I_C} \approx \frac{I_{\text{B,bulk}} + I_{\text{B,p}} + I_{\text{B,BS}} + I_{\text{B,SCR}} - I_{\text{CBO}}}{I_{nE}}
\]

(2.4)

If one assumes that the dominant component of the base current is the hole current injected from base to emitter, then the current gain can be expressed as \( \beta \approx I_{nE} / I_{\text{B,p}} \). This ratio can be expressed in terms of semiconductor parameters as [20]:

\[
\beta \approx \frac{D_{\text{nB}} I_{\text{pE}} N_{\text{DE}}}{D_{\text{pE}} W_B N_{\text{AB}}} \exp \left[ \frac{q(V_p - V_n)}{kT} \right]
\]

(2.5)

where \( D_{\text{nB}}, D_{\text{pE}} \) are the minority carrier diffusion coefficients in the base and emitter, \( N_{\text{AB}}, N_{\text{DE}} \) the doping concentrations in the base and emitter, and \( V_n, V_p \) the potential barriers in the conduction and valence bands, respectively; \( I_{\text{pE}} \) is the minority carrier diffusion length in the emitter, and \( W_B \) is the neutral base width. For homojunctions \( V_n \) and \( V_p \) can be expressed simply as:

\[
V_n = V_p = V_{\text{bi}} - V_{\text{BE}} \quad \text{for homojunction bipolar transistor}
\]

(2.6)
where $V_{bi}$ is the built-in potential of the junction, as defined in Fig. 2.1. To obtain a high gain it is therefore necessary that $N_{DE} >> N_{AB}$. However, due to the bandgap narrowing, increasing the emitter doping beyond $10^{19}$ cm$^{-3}$ improves $\beta$ only very slowly [20] and reducing the base doping decreases the Early voltage, increases the base series resistance which is undesirable for high frequency applications, and also enhances the current crowding.

On the other hand, for an abrupt heterojunction, as shown in Fig. 2.3(a), $V_n$ and $V_p$ are no longer equal:

$$\begin{align*}
V_n &= (V_{bi} - \Delta E_B) - V_{BE} \\
V_p &= (V_{bi} + \Delta E_V) - V_{BE}
\end{align*}$$

for abrupt HBT \hspace{1cm} (2.7)

where $\Delta E_B$ is the amount of band bending in the base. As will be shown later, due to the heavy doping of the base compared to the emitter in conventional HBTs, $\Delta E_B$ is generally very small and can be neglected. Therefore:

$$\beta_{\text{abrupt}} = \beta_{\text{hom}} \cdot \exp(\Delta E_V / kT)$$

\hspace{1cm} (2.8)

For an Al$_{0.3}$Ga$_{0.7}$As/GaAs abrupt HBT, $\Delta E_V = 0.157$ eV and therefore the exponential term in (2.8) will be in the order of 440 at room temperature. If the emitter material is properly graded in the last few hundred angstroms close to the base to remove the spike and notch in the conduction band (Fig. 2.3(b)), then $V_n$ and $V_p$ will be derived as:

$$\begin{align*}
V_n &= (V_{bi} - \Delta E_C) - V_{BE} \\
V_p &= (V_{bi} + \Delta E_V) - V_{BE}
\end{align*}$$

for graded HBT \hspace{1cm} (2.9)

Therefore, 

\hspace{1cm} 17
\( \beta_{\text{graded}} = \beta_{\text{homom}} \cdot \exp \left[ \frac{\Delta E_C + \Delta E_V}{kT} \right] = \beta_{\text{homom}} \cdot \exp(\Delta E_g / kT) \) (2.10)

Using the same example of Al\(_{0.3}\)Ga\(_{0.7}\)As/GaAs HBT, \( \Delta E_g \approx 0.414 \text{ eV} \) and one obtains for the exponential term in (2.10) a large value of \(-9 \times 10^6\). However, this does not mean that arbitrary high current gains can be achieved. It simply means that the hole back injection current \( I_{B,p} \) becomes a negligible part of the base current compared to the recombination currents which are the topic for the next section. High current gain values, above say 100, are of limited interest by themselves, except perhaps in phototransistors. The principal benefit of a wide bandgap emitter is therefore not the ability to achieve high current gains, but the freedom to change doping levels in emitter and base without significant constraints by injection efficiency consideration, and thereby to re-optimise the transistor at a higher performance level.

We start this discussion with the choice of emitter doping. A wide-gap emitter permits a drop in emitter doping by several orders of magnitude without deterioration of \( \beta \). It is well known that the junction capacitance of a highly asymmetrically doped pn junction depends only on the doping level of the less heavily doped side. Since the emitter doping of HBTs is practically dropped below the typical base doping of a homojunction bipolar transistor, the B-E junction capacitance can be decreased in HBTs (\( C_{\text{JBG}} \propto \sqrt{N_{DE}} \)). The result is only a slight improvement in speed, since the emitter charging time is generally small compared to other components of delay time (see chapter 10 for more details).

But the most important single change made possible by a wide-gap emitter is a significant increase in base doping, limited only by technological constraints and by the need to keep the minority lifetime inside the base above some reasonable limit. The advantages of a highly doped base region are as follows:

- Reduction of the base series resistance will result in improvement of some high current effects such as emitter current crowding.
- Base-width modulation will be minimised, thus increasing the Early voltage and the output resistance of the transistor in common-emitter configuration.
- Due to the minimised base-width modulation and low base series resistance, the base region can be made very thin, thus reducing the base transit time and increasing the cutoff frequency, \( f_T \).
- Maximum frequency of oscillation, \( f_{\text{max}} \), is a figure of merit for microwave transistors. It may be written as:

\[
 f_{\text{max}} \approx \sqrt{ \frac{f_T}{8\pi R_{bb} C_{bc}} } 
\] (2.11)

Since \( f_T \) is increased and \( R_{bb} \) is reduced, \( f_{\text{max}} \) will be improved significantly by using a thin highly doped base region. As a result, HBTs with impressive high frequency performance (\( f_{\text{max}} \) approaching 1THz [2]) have already been demonstrated.
Chapter 2

Theory of Heterojunction Bipolar Transistors

2.2. Components of the Base Current

In the previous section the basic theory of heterojunctions was discussed and compared with homojunctions. Comparing Eqs. (2.8) and (2.10), it is clearly seen that an HBT with a graded B-E heterojunction should theoretically provide higher current gain than the abrupt HBT due to the exponential dependence on $\Delta E_g$ in the former. However, this simple theory makes several assumptions in that all other generation-recombination currents are neglected. A true comparison of the current gain in abrupt and graded HBTs requires accurate analysis of various base current components such as $I_{B,p}$, $I_{B,bulk}$, $I_{B,SCR}$, and $I_{B,BS}$.

In addition to the above components, there are some other contributions to the base current which can be important in special layout configurations of the device or in very small B-E bias (i.e., leakage currents). The current component due to the generation of carriers inside the B-C SCR, $I_{CBO}$, is almost insensitive to $V_{BE}$, and hence it is important only in the very low bias region and causes saturation in the collector and base currents observed in the Gummel plots. The situation for the B-C peripheral leakage [21] is the same as $I_{CBO}$. This current component originates from the high electric field existed in the dielectric near the base and collector periphery, giving rise to a large probability for the electrons in the p-type base and holes in the n-type collector to tunnel through the dielectric and reach the collector and base regions, respectively [21].

Base contact recombination, $I_{B,cont}$, is not thought of as a major base recombination problem for most process technologies. This base current component arises from the electrons injected from emitter to the base, where they not only diffuse downward toward the collector, but also laterally toward the base contact. Since the carrier concentration decreases rapidly from the edge of the intrinsic base region toward the base contacts, $I_{B,cont}$ is small when the B-E spacing is sufficiently long. However, HBTs designed for high speed operation employ self-aligned B-E structures to reduce the extrinsic base resistance and the B-C junction capacitance. Self-alignment techniques result in a very short B-E spacing (0.1-0.3 $\mu$m), potentially leading to a large $I_{B,cont}$. However, it has been shown in [22] that ~0.1 $\mu$m spacing between base metal and emitter mesa is sufficient to eliminate this current component, and hence base contact recombination should not be a major concern even for self-aligned HBTs.

The other components of the base current are related to the B-E applied voltage through the following exponential relation:

$$I = I_0 \cdot \exp \left( \frac{qV_{BE}}{\eta kT} \right)$$  \hspace{1cm} (2.12)

where $I_0$ is the saturation current and $\eta$ is the ideality factor of the corresponding component of current. In the following we discuss the four most important components of the base current qualitatively.
2.2.1. Base Bulk Recombination

The base bulk recombination is attributed to the electrons which are injected from the emitter and recombine in the neutral base region. Since these are minority carriers, the thicker the base and the higher the doping concentration in the base, the more chance there is for recombination. As in the design of homojunction transistors, the base thickness of the transistor must be much smaller than the diffusion length of electrons in the base. Since HBTs have a high doping concentration in the base (> $1 \times 10^{19}$ cm$^{-3}$), electron diffusion length tends to be very small. Therefore, HBTs require a much thinner base to avoid base bulk recombination. Typically, base thicknesses are 1000Å or less with a doping concentration of $1 \times 10^{19}$ cm$^{-3}$ or greater.

Base bulk recombination gives a base ideality factor identical to that of collector, which is close to unity [23-24]. It has been shown that the base current of modern abrupt InP/InGaAs, InAlAs/InGaAs, and InGaP/GaAs HBTs is dominated by this component in the moderate to high $V_{BE}$ range [12, 25-26]. Increasing the gain of such devices is usually accompanied by a corresponding increase in base sheet resistance, $R_{sh,B}$. It is observed that $\beta \propto (R_{sh,B})^m$, where $m$ is in the range (1-2), depending on the base bulk recombination being dominated by radiative or Auger recombination mechanisms [27-28]. Hence, the ratio of $\beta / R_{sh,B}$ can be a good indicator of the base material quality [28]. Finally, base bulk recombination may be reduced by compositionally grading the base region, while the existence of a conduction band barrier at the B-C heterojunction of double HBTs significantly enlarges this component of base current [29].

2.2.2. Back Injection of Holes from Base to Emitter

Similar to the base bulk recombination, the current component attributed to the back injected holes from base to emitter has an ideality factor of unity [30-31]. Although this component of base current is reduced to some extent by the discontinuity in the valence band of the B-E heterojunction, it is shown to be the dominant component for the medium to high $V_{BE}$ range in abrupt AlGaAs/GaAs HBTs [26,30]. In fact, in [30] the temperature dependence of the current gain in abrupt AlGaAs/GaAs HBTs is attributed to the small valence band discontinuity at the B-E junction of these transistors, which makes the injection of holes to the emitter the dominant component of the base current. This is in agreement with the numerical calculation results in the present work. In section 7.4 it is shown that at high temperatures hole back injection can even dominate the base current of InGaP/GaAs HBTs, which offer much larger valence band discontinuities than their AlGaAs counterparts, causing the reduction of gain with increasing temperature in such devices.

2.2.3. Space-Charge Recombination

SCR recombination is found to be the dominant component of the base current for small to medium $V_{BE}$ range in most of the HBTs studied so far [12,26]. It gives a base ideality factor
close to 2, and is therefore responsible for the variation of current gain with collector current [23]. Grading the B-E heterojunction in AlGaAs/GaAs HBTs extends the dominance range of the SCR recombination component to all operating conditions [23,32]. This higher SCR recombination current in graded HBTs results from the fact that the depletion region adjacent to the base layer is made of narrower bandgap materials. Therefore, the larger thermal carrier concentration associated with narrow bandgap material leads to a higher recombination current. This is one of the reasons why abrupt AlGaAs/GaAs HBTs sometimes show larger current gains than their graded counterparts, despite the fact that Eqs. (2.8) and (2.10) predict a much larger gain in the latter device.

In [33] and [34], it is shown that the SCR recombination is dominated by the recombination in the emitter side, which has an ideality factor close to 2. The base side of SCR recombination, which has an ideality factor close to unity, appears to be always much smaller than the base bulk recombination current. Therefore, the base side SCR recombination can be ignored for almost all imaginable device conditions of Npn HBTs [33], although there are suggestions of this component being the dominant contribution to the base current for Pnp HBTs [35]. Also in [26] it is found that in abrupt HBTs the SCR recombination current depends strongly on the density of trapping states at the B-E heterointerface. This indicates that care should be taken during device fabrication so that lattice mismatch between emitter and base, which causes the interface recombination traps, is minimised.

### 2.2.4. Extrinsic Base Surface Recombination

One of the major contributions to the undesirable base current is surface recombination, which is due to the relatively large surface recombination velocity (~1×10⁶ cm/s) of free GaAs surface. Consequently, some minority carriers injected from the emitter recombine with the base majority carriers at the surface, giving rise to the extrinsic base surface recombination current, I_{B,BS}. This base current is proportional to the emitter periphery rather than the emitter area, unlike the collector current. For small devices whose device perimeter-to-area ratio is large, I_{B,BS} is a major component of the overall base current, and the current gain is substantially reduced from that of a large device. This is known as the emitter-size effect [36]. Unlike the GaAs surface, the free InGaAs surface exhibits more ideal characteristics, with a dramatically lower surface recombination velocity of about 1×10³ cm/s. Therefore, in InP/InGaAs HBTs, I_{B,BS} is inconsequential, even when the device size is small. Similarly, in silicon BJTs, I_{B,BS} is hardly ever mentioned, because the extrinsic base surfaces are well passivated by SiO₂.

The ideality factor associated with I_{B,BS} has been a matter of controversy. In one view, the ideality factor is thought to be ~2 because the surface recombination takes place in a depletion region [37], in the same way that the B-E SCR recombination is characterised by an ideality factor of 2. While this view is conceptually appealing for a first order approximation, it neglects the possibility that the availability of the recombination carriers are intrinsically
different in these two types of recombination [38]. When the limited availability of carriers is taken into consideration, a theoretical ideality factor close to 2 for low $V_{BE}$ but approaching unity at higher $V_{BE}$ is possible [38]. This theoretical finding is later supported by experimental verification for both InGaP/GaAs HBTs [39] and HEBTs [40]. More recent publications are consistent with ideality factors closer to unity rather than 2 for the extrinsic base surface recombination [39,41]. According to the discussions of the ideality factors for various components of base current, it appears that as a rule of thumb one can say: ‘recombination currents originating from the base region have ideality factors close to 1, while those from the emitter side have ideality factors close to 2.’ Table 2.1 summarises the important characteristics of various base current components discussed in this section.

<table>
<thead>
<tr>
<th>Current Component</th>
<th>Ideality Factor</th>
<th>Proportional to</th>
<th>Most Important in Devices with</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I_{B,BS}$</td>
<td>&lt;1.3</td>
<td>Perimeter</td>
<td>small geometry, unpassivated base</td>
</tr>
<tr>
<td>$I_{B,cont}$</td>
<td>&lt;1.3</td>
<td>Perimeter</td>
<td>self-aligned B-E structures</td>
</tr>
<tr>
<td>$I_{B,bulk}$</td>
<td>$\eta_B = \eta_C \approx 1$</td>
<td>Area</td>
<td>heavy base doping</td>
</tr>
<tr>
<td>$I_{B,SCR}$</td>
<td>~2</td>
<td>Area</td>
<td>large number of traps, graded HBTs</td>
</tr>
<tr>
<td>$I_{B,p}$</td>
<td>1</td>
<td>Area</td>
<td>high temperature, abrupt HBTs (with small $\Delta E_Y$)</td>
</tr>
</tbody>
</table>

### 2.2.4.a. Emitter-Size Effect

Since the extrinsic base surface recombination is proportional to the perimeter of the emitter mesa, $P_E$, while other components of the base current are proportional to the emitter area, $A_E$, one can write:

$$I_B = A_E (J_{B,bulk} + J_{B,SCR} + J_{B,p}) + P_E K_{B,BS}$$  \hspace{1cm} (2.13)

where $K_{B,BS}$ is the extrinsic base surface recombination current per unit length of the emitter periphery, and the $J$ values are the current densities corresponding to other components of the base current. Dividing both sides of (2.13) by $I_C = J_C A_E$ gives:

$$\frac{1}{\beta} = \frac{J_{B,bulk} + J_{B,SCR} + J_{B,p}}{J_C} + \frac{P_E K_{B,BS}}{A_E J_C}$$ \hspace{1cm} (2.14)

Therefore, if one plots ($J_C / \beta$) versus ($P_E / A_E$) for devices with various geometries, a linear graph should be obtained with the gradient equal to $K_{B,BS}$ (see Fig. 2.4). Usually the perimeter-to-area ratio decreases for devices with smaller geometries, therefore one expects to observe lower gains in such devices as compared to the large geometry HBTs (i.e., the emitter-size effect). For instance, if a circular emitter geometry with radius $r$ is used, then ($P_E / A_E$) = $2 / r$, and the current gain degrades for smaller emitter diameters.

In [42] a detailed analysis based on a two-dimensional model is carried out which takes into account the surface Fermi level pinning and surface recombination of the carriers in the
emitter and extrinsic base surfaces. The results show that surface recombination is significant only at the boundary region between the intrinsic base and the extrinsic base. The degradation of the current gain due to the reduction of the emitter area is shown from calculation to be suppressed for a very thin base or a graded base structure. On the other hand, the effect is enhanced for a double HBT due to the carrier blocking effect of the conduction band offset at the B-C heterojunction, which enhances the lateral diffusion of electrons inside the base.

2.2.4.b. Emitter-Edge Thinning Design

Because free GaAs surface is characterised by a high surface recombination velocity, extrinsic base surface recombination current often dominates the total base current, especially when the emitter perimeter-to-area ratio is high. However, when a thin depleted wide bandgap ledge is left on the extrinsic base to passivate this surface (Fig. 2.5), this base current component can be greatly reduced [43]. This method is called the emitter-edge thinning (EET) design.

The wide bandgap ledge is depleted due to a combination of surface Fermi level pinning on the top and being the more lightly doped layer in the B-E junction below [44: chap. 3]. The passivationledge needs to be thin enough so that it is fully depleted. If the passivated ledge is partially undepleted, sizeable emitter current can still flow through the ledge and then inject into the B-E junction [45]. Hence, the active device area would be much larger than the intrinsic emitter mesa area defined by lithography. On the other hand, the passivation ledge cannot be too thin such that it does not passivate the surface effectively. Typical thicknesses of the passivation ledge are in the range 500-1000Å [45]. Also Liu et al. [13] observed that a passivation ledge width as narrow as 1 μm is sufficient to eliminate the extrinsic base surface recombination.

Lin and Lee [43] showed that by using the EET technique a current gain of 12,500 can be obtained which was the highest reported for HBTs at that time. The base doping of this HBT, however, was very low at $5 \times 10^{17}$ cm$^{-3}$.

The fabrication of EET AlGaAs/GaAs HBTs requires an extra processing step and the AlGaAs ledge needs to be carefully monitored in order not to etch all the AlGaAs, unless a selective etchant for etching GaAs on top of AlGaAs layers is available (see chapter 8 for more details). This problem is substantially relaxed by using InGaP as the emitter material of
HBTs, since highly selective etching solutions exist for etching of InGaP on GaAs and vice versa. In [46] a very attractive dual InGaP etch stop layer design is used to fabricate a self-aligned, emitter-edge thinned HBT.

2.3. Abrupt versus Graded Base-Emitter Junctions

Both abrupt and graded B-E junction HBTs have been discussed briefly in section 2.1. The abrupt junction HBT is the simplest structure that consists of a wide bandgap emitter and a narrow bandgap base with an abrupt interface. As discussed previously, with an abrupt AlGaAs/GaAs junction a large conduction band spike is generated which limits the emitter injection efficiency. To reduce this spike, grading of the emitter is used (Fig. 2.3) with a graded region typically about 200-300Å thick. This smoothes out the conduction band spike, reduces the B-E turn-on voltage [47] (see Fig. 2.6), and improves the emitter injection efficiency. Reduction of the turn-on voltage in graded AlGaAs/GaAs HBTs has some important consequences. For a given collector current, the power dissipation of abrupt HBTs would be larger, due to the increased turn-on voltage. This raises the junction temperature at high current densities (i.e., device self-heating), and degrades both the current gain and cutoff frequency of HBTs [48]. Device self-heating also decreases the electron saturation velocity (see section 10.5.1), which consequently reduces the onset of the Kirk effect [48]. The above deleterious effects lead to an earlier current gain and $f_T$ fall-off in abrupt HBTs as compared to their graded counterparts [48]. Additionally, a lower B-E turn-on voltage enhances the symmetry between the B-E and B-C junctions, leading to a significantly lower offset voltage ($V_{CE,offset}$) in graded HBTs.

In spite of the above advantages of a graded B-E heterojunction, Enquist et al. [49] observed that for all their HBT structures, the abrupt junctions always gave the higher current gain, a result that is completely contradicting the early theoretical predictions (e.g., in [50-51]). This experimental observation can be attributed to various physical mechanisms which are detailed below.

A graded B-E heterojunction is experimentally demonstrated [23], as well as theoretically proved [47], to increase the recombination current in the emitter SCR. This is due to the fact that the $\text{Al}_x\text{Ga}_{1-x}\text{As}$ emitter immediately next to the base layer has a narrower bandgap in graded HBTs as compared to abrupt devices. Therefore, the intrinsic carrier concentration inside the emitter of the former device is significantly larger, leading to a higher SCR recombination rate, larger base current, and a reduced current gain especially at low to medium $V_{BE}$. Additionally, it is shown in [52] that an abrupt B-E heterojunction significantly reduces the sensitivity of the current gain to interfacial traps located at the B-E
heterointerface, which is one of the main limiting factors in achieving reproducible current gains in graded AlGaAs/GaAs HBTs (for more details, see section 7.3).

Another reason for underestimation of abrupt HBT performance in early studies is the neglect of the thermionic-field emission (TFE) of electrons across the conduction band spike. It is well known that the drift-diffusion transport model is not valid when there are discontinuities in the energy bands larger than $kT$ in a length shorter than the mean free path of carriers. This is the case for abrupt HBTs. In such cases, the transport of carriers is limited by the thermionic-emission and tunnelling of carriers across the heterojunction interface [53]. Quasi-Fermi level splitting at an abrupt heterointerface is a characteristic of the condition when carrier transport is limited by TFE (see Fig. 2.3(a)). When TFE was included in theoretical studies of HBTs [47], it was shown that abrupt HBTs have comparable current drive capability and maximum current gain to those of graded devices. Additionally, an abrupt HBT is characterised by a collector ideality factor larger than unity. The collector ideality factor of an abrupt HBT can be written as [44: p.156]:

$$\eta_C = \left(1 + \frac{\varepsilon_E N_{DE}}{\varepsilon_B N_{AB}}\right) \cdot \frac{E_{00}}{kT} \cdot \coth\left(\frac{E_{00}}{kT}\right)$$

(2.15)

where $\varepsilon_E$, $\varepsilon_B$ are the dielectric constants, and $N_{DE}$, $N_{AB}$ are the doping concentrations of emitter and base, respectively, and $E_{00}$ is defined as:

$$E_{00} = \frac{qh}{4\pi} \sqrt{\frac{N_{DE}}{m^* \varepsilon_E}}$$

(2.16)

The first term in (2.15) is due to the partitioning of the voltage drop across the base and emitter regions in the formulation of thermionic-emission [47]. For the typical doping
concentration of the emitter and base regions, this factor can be in the range (1.01-1.02). The second term in (2.15) arises from the bias dependence of the tunnelling factor (see section 7.2 and Fig. 7.3), and may increase the ideality factor to the range (1.1-1.2) for abrupt HBTs.

A final reason for the superior DC current gain and cutoff frequency of abrupt HBTs as compared to graded HBTs is the possibility of the so-called velocity overshoot and ballistic transport inside the base region of the former devices. The previously mentioned drawbacks of the conduction band spike accompanying an abrupt B-E junction is partially compensated by the fact that such a barrier would inject the electrons into the base region with a substantial kinetic energy. When energetic electrons are injected into the base or collector regions, they initially have the high mobility and high velocity associated with the Γ-valley. When they are energetic enough, they eventually transfer into the low mobility X- or L-valleys, resulting in saturation of velocity to $<1 \times 10^7$ cm/s. But before this transfer occurs, they travel a distance of a few hundred angstroms with a fairly high velocity exceeding $5 \times 10^7$ cm/s [54]. Such a fast, near-ballistic, electron transport through the base is shown to reduce the base transit time of abrupt HBTs significantly [55]. Moreover, considering the velocity overshoot in the collector region of HBTs for energy transport simulations is shown to enhance the calculated intrinsic $f_T$ of graded base HBTs by 160 and 50 percent for 0.1 and 0.5 μm collectors, respectively, as compared to the case when drift-diffusion approach is adopted [54].

Designing the thickness of graded region in graded HBTs requires particular attention. When the B-E forward bias is large, a conduction band spike and dip forms at the point where the composition grading starts [50-51] (see Fig. 2.3(b)). When the graded region is too thin, the spike formed at large $V_{BE}$ limits the transport of electrons from emitter to base, effectively eliminating the advantages of a graded heterojunction. On the other hand, when this grading region is too thick, the large density of electrons inside the conduction band dip increases the SCR recombination rate, which degrades the current gain of the device [50]. A graded region of 200-300Å seems to produce an optimum compromise. Additionally, when a parabolic grading profile is implemented in the emitter, the resulting performance is not significantly different from linearly graded devices to warrant complicated process steps [51].

Due to the lattice match between Al$_x$Ga$_{1-x}$As and GaAs for the entire range of composition, grading the B-E heterojunction of AlGaAs/GaAs HBTs is relatively easy. However, other material systems, such as InGaP/GaAs and InP/InGaAs, do not benefit from this advantage. In such material systems one or a combination of the following techniques may be used to reduce the potential spike at the B-E heterojunction, thereby improving the injection efficiency and reducing the B-E turn-on voltage and $V_{CE,offset}$: (i) inserting a narrow bandgap spacer layer between emitter and base [56]; (ii) using a doping spike inside the emitter [57-59]; (iii) step-grading the heterojunction using quaternary materials [60]; (iv) employing a superlattice to grade the heterojunction [61]. When a spacer layer is inserted at the B-E heterojunction, it is shown that the turn-on voltage may approach that of the graded HBTs,
with the expense of a slight reduction in maximum current gain and cutoff frequency [47]. The spacer layer also helps to achieve better collector current reliability and uniformity across the wafer when designed to accommodate the expected base dopant outdiffusion (see next section).

2.4. The Choice of Base Dopant

The choice of the base dopant is an important issue both for the performance and reliability of HBTs. Beryllium (Be) and carbon (C) are the two most widely used dopants for p-GaAs. Beryllium is known to have a large diffusion coefficient in GaAs. Consequently, the Be diffusion from the base into emitter during the HBT operation is the most serious concern in using this impurity as the base dopant. Theoretical studies [56] as well as experimental results [62] have shown that Be outdiffusion drastically degrades both the DC and RF performance of HBTs. First, the B-E turn-on voltage increases, since the junction is now formed inside the wide bandgap material. Second, the current gain is significantly degraded, because the B-E is now a homojunction, and the main advantage of a heterojunction in improving the injection efficiency is lost. Other effects of Be outdiffusion include increase of the offset voltage, and degradation of both $f_I$ and $f_{\text{max}}$ [63].

The deleterious effects of base dopant diffusion during the high temperature growth and/or fabrication cycles can be partially compensated by insertion of a thin narrow bandgap spacer layer between emitter and base. But, the movement of these impurities during device operation, due to the so-called recombination-enhanced impurity diffusion (REID) [64], is a reliability concern. In a simplistic view, a REID mechanism can be described as the migration of impurity atoms with the aid of the energy released during a recombination process.

Carbon as a p-type dopant for III-V materials has some advantages over Be. First, in GaAs crystal, for instance, the C acceptors occupy the arsenic sites, while Be acceptors occupy the gallium sublattice. This leads to a very low diffusion coefficient for C [63], which avoids the difficulties due to the Be outdiffusion. Second, very high p-type doping levels are possible for C-doped GaAs, primarily due to the high solid solubility of carbon. Very good quality HBTs with base hole concentrations in excess of $1 \times 10^{20}$ cm$^{-3}$ using carbon are reported [65-67]. Thirdly, C-doped GaAs layers are shown to have 20-30% higher hole mobilities as compared to Be- or Zn-doped samples with the same amount of hole concentration (see subsection 5.4.3.a for more details). However, there are also disadvantages associated with C doping. First, due to the amphoteric nature of C in III-V materials, special growth conditions or post-growth treatments are required before an effective p-doping is obtained. Nevertheless, by adjusting growth parameters, high p-doping using carbon acceptors is achieved not only for GaAs, but also for In$_{0.49}$Ga$_{0.51}$P [68] and In$_{0.55}$Ga$_{0.47}$As [69]. Additionally, because carbon dopant has a significantly smaller atomic size than that of Ga or As, stress is developed in HBTs whose GaAs base layer is heavily doped with C (exceeding $5 \times 10^{19}$ cm$^{-3}$) [44: section 7.7]. This stress can create defects and dislocations which act as catalysts for solid-state
reactions such as REID. Indium atoms, which are larger than either Ga or As and do not form an acceptor or donor levels, are sometimes intentionally added to offset the size difference of carbon and thereby reduce the stress [70]. Experimental results establish that such indium co-doping improves transistor reliability [70].

But the most important issue in using carbon as the p-type dopant of III-V compounds is the formation of C-H complexes. Since C-H bonds form readily and there are many hydrogen sources during the MOCVD growth, where hydrogen is used as the carrier gas, high concentrations of hydrogen incorporate in the base and passivate carbon acceptors. Hydrogen can also be adsorbed during plasma depositions (e.g. in PECVD of SiN) [71: section 8.3], or introduced by a proton implant isolation [72-73]. In the presence of hydrogen, the C atoms tend to be passivated, giving rise to formation of electrically inactive C-H complexes. In the GaAs crystal, the hydrogen can also act as donor, compensating the ionised C acceptors. Experimental data suggest that the hydrogen atoms are incorporated as C passivators for concentrations up to $3\times10^{18}$ cm$^{-3}$ and as donors beyond. During electrical stress, the C-H complexes are broken by the injected minority carriers and the activated base dopant concentration increases [73-75]. Alternatively, a high temperature anneal can also lead to the release of hydrogen and reactivation of carbon dopants [62,76]. Consequently, the B-E turn-on voltage increases, and the current gain and the base sheet resistance decrease [77]. The contamination by hydrogen is believed even to be responsible for the early increase of the DC current gain and reduction of turn-on voltage (i.e., the burn-in effect) frequently observed during the electrical stress of HBTs with a carbon doped base layer [73,78].

The hydrogen contamination can be minimised by various methods. These include (i) C-doping of the base layer by adjusting the V/III ratio without the usage of any dopant precursors [79]; (ii) using nitrogen as the carrier gas in MOCVD growth [67]; (iii) post-growth annealing of the wafer at temperatures around 650°C [76]; (iv) using GSMBE growth method which involves much less hydrogen than the MOCVD technique [77]; and (v) using He$^+$ rather than H$^+$ for implant isolation of small-geometry devices [72,80].

Despite all the problems associated with C-H complexes, it is shown in [62] that C-doped AlGaAs/GaAs HBTs are much more stable than Be-doped HBTs. Highly reliable InGaP/GaAs HBTs are reported in the literature [74-75,79]. In a recent investigation, no significant device degradation was observed after 10'000 hours of stress under moderate current densities of $J_C = 25$ kA/cm$^2$ and high junction temperature of 264°C for an InGaP/GaAs HBT with a carbon doped base [79]. Consequently, C-doped base layers are used for majority of GaAs HBTs studied in this thesis.
CHAPTER 3. HBT STRUCTURES

The basic concept of HBTs was discussed in the previous chapter. This chapter deals with various material systems and structural alterations used in heterostructure bipolar devices. Section 3.1 summarises the advantages of In_{0.49}Ga_{0.51}P over AlGaAs as the emitter material of GaAs-based HBTs. Then in section 3.2, alternative configurations of HBTs, including double HBTs, Pnp HBTs, collector-up HBTs, and graded base HBTs, are discussed. Section 3.3 briefly reviews the material systems other than GaAs-based for making HBTs. Finally, structural variants to that of the conventional HBT conclude this chapter.

3.1. InGaP/GaAs HBTs

Early works on HBTs were mainly concentrated on the AlGaAs/GaAs material system. One possible reason for this is the easy lattice matching of \(\text{Al}_x\text{Ga}_{1-x}\text{As}\) to GaAs with any Al mole fraction. Although early devices showed promising results, there are inherent disadvantages with this material system. The main problem with this structure is that the majority of the bandgap discontinuity \(\Delta E_g\) is in the conduction band \((\Delta E_C \approx 0.62\Delta E_g)\), reducing the emitter injection efficiency. This discontinuity can be reduced by grading the Al content near the junction. However, as mentioned in the last chapter, grading the bandgap significantly increases the B-E SCR recombination rate. Furthermore, AlGaAs/GaAs HBTs are difficult to fabricate using wet or dry etching techniques, which require carefully controlled etching to reach the base. This is obviously not suitable for mass production of HBT devices. Also the Al in AlGaAs/GaAs HBTs generates DX centres, increasing the recombination currents, hence reducing the current gain.

The Aluminium-free In_{0.49}Ga_{0.51}P/GaAs HBTs were introduced by Mondry and Kroemer [81] in 1985 as a replacement for the conventional AlGaAs emitter devices. Although the early results were not as impressive as those of the more developed AlGaAs/GaAs technologies, InGaP/GaAs HBTs gradually showed clear advantages such that today they have completely replaced the AlGaAs HBTs in many research laboratories and industrial centres worldwide [4,6,11,65]. One of the main reasons for this is the larger valence band discontinuity \((\Delta E_V > 0.30\text{eV})\) and the smaller conduction band discontinuity \((\Delta E_C < 0.15\text{eV})\) of the InGaP/GaAs material system (see subsection 5.3.2). This is very desirable to reduce the back injection of holes from the base to the emitter, therefore increasing the emitter injection efficiency without the need for grading. In [82] a large number of InGaP/GaAs and (abrupt) AlGaAs/GaAs HBTs were analysed and it was shown that the dominant carrier transport mechanism in InGaP HBTs is the carrier diffusion through the base layer, while thermionic-field emission across the B-E heterointerface dominates the transport in AlGaAs devices. This finding suggests that the conduction band barrier across the InGaP/GaAs B-E heterojunction is so small that the barrier spike does not affect the
carrier transport.

The superior band alignment of InGaP/GaAs HBTs has some very important consequences. As shown in [52], the large $\Delta E_V$ of InGaP/GaAs material system significantly reduces the sensitivity of current gain to the interfacial traps and defects at the B-E heterojunction. A reduced back injection of holes in InGaP/GaAs HBTs was also found to be the main reason for the less sensitivity of the current gain of these devices to the variation of temperature [30]. These are more thoroughly discussed in chapter 7 with the aid of numerical simulation results.

The second major advantage of InGaP over AlGaAs as the emitter material for HBTs is related to the high etching selectivity between InGaP and GaAs. While HCl and its mixtures with H$_2$O and/or H$_3$PO$_4$ etches InGaP very fast, it has almost no effect on GaAs layers, which need an oxidising agent such as H$_2$O$_2$ for effective etching (see chapter 8 for more details). This simplifies the processing of InGaP/GaAs HBTs significantly and also enhances the yield and uniformity of the devices across the wafer. The etch selectivity is specially important between the emitter and base materials during the emitter mesa etch process to reach the base. Since the base layer of HBTs is usually very thin (< 1000Å), a non-selective etch can easily overetch the base, or even completely miss it if the process is not carefully monitored. Even a slight overetch of the base may increase the base series resistance, degrading the power gain of the transistor [83], and also may lead to a significant emitter-size effect [84]. Highly selective etching solutions may also be used to undercut the emitter metal [46] or the collector [85] of HBTs in self-aligned fabrication processes.

Using InGaP as the emitter of the carbon-doped base HBTs has another advantage. Any outdiffusion of the carbon dopants (perhaps due to the recombination-enhanced impurity diffusion- REID) will not shift the p-n junction from the heterointerface, because carbon does not easily form an acceptor state in InGaP [68,86]. Recent reports suggest that C-doped base InGaP/GaAs HBTs are much more reliable than their AlGaAs/GaAs counterparts [11,74]. It is speculated in [86] that the superior reliability characteristics of InGaP HBTs is due to the fact that Carbon does not form an acceptor inside the InGaP emitter.

InGaP has also shown a lower impact ionisation rate than AlGaAs with a similar bandgap energy [87]. This leads to a larger breakdown voltage when a wide bandgap material is also used as the collector of (double) HBTs. Among other advantages of InGaP over AlGaAs the lower 1/f noise [88] and the absence of the so-called DX centres [89-90] in InGaP/GaAs HBTs has to be mentioned.

A final advantage of using InGaP rather than AlGaAs is its lower surface and interface recombination velocity [91]. InGaP/GaAs has the lowest reported interface recombination velocity of 1.5 cm/s among the III-V materials [92]. This reduces the recombination rate
inside the B-E SCR and at the emitter mesa sidewall, and also makes InGaP a more
effective passivating ledge in the EET design.

The above advantages of InGaP/GaAs HBTs over their AlGaAs/GaAs counterparts have
resulted in superior DC and RF performance observed in the former devices. Liu and Fan
[12] reported a near ideal I-V characteristics of InGaP/GaAs HBTs in which the measured
current gain of the device remained constant over five decades of collector current and was
greater than unity at ultrasmall current densities on the order of $1 \times 10^{-6}$ A/cm$^2$. A very high
current gain of 2690 is also reported in [13] for an InGaP HBT with a thick base doped to
$5 \times 10^{18}$ cm$^{-3}$. Ahmari et al. [14,93] have demonstrated an InGaP/GaAs HBT with a graded
$\text{In}_{1-x}\text{Ga}_x\text{As}$ base layer which showed large $\beta_0$, $f_T$, and $f_{\text{max}}$ of 162, 126GHz, and 197GHz,
respectively. A simple collector undercut technique is employed in [94] to reduce the B-C
capacitance, which resulted in $f_T$ and $f_{\text{max}}$ values of 80 and 171GHz, respectively. Finally,
Mochizuki and co-workers from Hitachi have used variety of techniques to reduce B-C
capacitance and base series resistance of InGaP/GaAs HBTs, and as a result demonstrated
very high $f_T$ and $f_{\text{max}}$ values of 170 and 275GHz, respectively [15-16]. These techniques
include heavy carbon doping of the base [65], buried polycrystalline GaAs selectively
grown under the base electrode [16,66,95], Ti/WSi base contact, and buried SiO$_2$ in the
extrinsic B-C region [15].

3.2. Various HBT Configurations

3.2.1. Double HBTs

HBTs that have a wide bandgap material in the collector as well as the emitter are called
double HBTs (DHBTs). There are some advantages in using wide bandgap collectors
particularly for high power and high temperature applications. Firstly, this structure has a
suppressed hole injection from base to collector under saturation regime of operation. In
many digital logic families, the B-C junction of the transistors are forward-biased during
part of the logic cycle. If the base region is more heavily doped than the collector, as would
normally be desirable, a copious injection of holes from the base into the collector takes
place, which increases dissipation and slows down the switching speed. This highly
destructive phenomenon can be avoided when a wide bandgap collector is utilised. In [96]
it is shown that InGaP/GaAs DHBTs require significantly shorter times to remove the
stored saturation charge than their SHBT counterparts. This has been attributed to the hole
blocking effect of the B-C valence band discontinuity.

Secondly, the smaller intrinsic carrier density ($n_i$) of a wide bandgap collector leads to a
decreased generation rate and hence a lower B-C leakage current. This allows the operation
of DHBTs at higher junction temperatures than can be achieved by SHBTs (see Figs. 9.19
and 9.23). Therefore, DHBTs are more suitable for high temperature applications.

Thirdly, higher B-C breakdown voltages can be achieved using wide bandgap collectors.
HBT Structures

Since the breakdown voltage is almost proportional to $E_B^{3/2}$ [97], the breakdown voltage of the InGaP collector should be 50% larger than that of the GaAs collector. Figure 3.1 shows the output characteristics of typical InGaP/GaAs single and double HBTs fabricated in our research laboratory. Both devices have identical cap layer, emitter, base, and sub-collector regions, and even their collector thickness and doping is the same (5000Å thick doped to $2 \times 10^{16}$ cm$^{-3}$). The only difference is in the material for the collector region, which is simply GaAs in the case of SHBT, but is composed of a 200Å GaAs spacer and 4800Å InGaP collector in the case of DHBT. The DHBT clearly shows an almost 50% higher B-C and C-E breakdown voltages. This is obviously more desirable for power applications.

![Figure 3.1- Output characteristics of typical InGaP/GaAs single (blue) and double (red) HBTs, demonstrating the higher breakdown voltage and lower offset voltage for the DHBT.](image)

In addition, the advantage of symmetrical energy band discontinuities between B-E and B-C will result in smaller offset voltages in DHBTs [98] (also see Fig. 3.1). A final benefit of a wide gap collector lies in the possibility of designing transistors in which the role of the emitter and collector can be interchanged by simply changing the biasing conditions, while retaining the advantages of a wide gap emitter. An example for the application of this design in the emitter-coupled logic (ECL) ICs is presented in [20].

There is, however, an important restriction in the use of wide-gap collectors, which must not be overlooked. It is essential that the free collection of electrons by the reverse-biased collector not be impeded by any heterobarrier due to a conduction band discontinuity. Early DHBT output characteristics exhibited rather high saturation voltages (i.e., the C-E voltage below which the transistor operates in the saturation mode) [99] due to the B-C conduction band spike. This discontinuity also has some other deleterious effects; it
increases the base bulk recombination, thus reducing the base transport factor and the device current gain. Additionally, a significant proportion of electrons arriving from the base to collector may be trapped behind this potential barrier leading to a notable enlargement of intrinsic delay times and degradation of high frequency performance of the device [60]. Furthermore, it has been shown that any barrier at the B-C junction will reduce the onset of Kirk effect, limiting the maximum operating collector current density of the device [100].

Various techniques have been suggested in the literature in order to reduce the effective conduction band barrier to minority electrons transiting from base to collector, including:

(i) **Continuous compositional grading between the base and collector materials:** This technique can be easily adopted for AlGaAs/GaAs HBTs. The quaternary material $\text{In}_{0.52}(\text{Al}_{x}\text{Ga}_{1-x})_{0.48}\text{As}$ may also be used to grade InAlAs/InGaAs heterojunctions [101]. However, for material systems such as InP/InGaAs or InGaP/GaAs, continuous grading of the composition (using InGaAsP quaternary material) requires simultaneous control of both group III and group V sublattice elements during the growth, which is practically difficult to achieve.

(ii) **Inserting a thin spacer (or set-back) layer between base and collector:** For example, an undoped GaAs spacer layer may be inserted between the InGaP collector and GaAs base of an InGaP/GaAs DHBT [60,102]. The thickness of this GaAs layer should be designed such that the collector conduction band barrier is below the base conduction band edge at zero B-C bias. Since this collector design lowers the conduction band barrier across the B-C heterojunction, current-voltage characteristics with small values of saturation voltage (<1V) and offset voltage are measured (see Fig. 3.1). The current gain of the device will also be significantly improved, because majority of electrons will now be able to jump over this potential barrier. However, inclusion of the undoped GaAs layer creates a triangular potential well at the B-C heterojunction, as in the bipolar quantum-well resonant tunnelling transistor (BiQuaRTT) [103], causing negative differential resistance (NDR) in the form of oscillations in the I-V characteristics. These oscillations are more apparent at low temperatures, but may also be observed at room temperature [102]. Fig. 3.2 shows an example of the oscillating output characteristics for an InGaP/GaAs DHBT with 200Å GaAs spacer measured in the present work at 78K.

Obviously, the above oscillations are potentially harmful to the linearity of the transistor. But the main problem associated with the spacer design is related to the high frequency performance of the device. As mentioned earlier, the carrier trapping behind the conduction band spike at the B-C heterojunction may significantly degrade the $f_T$ and $f_{\text{max}}$ of the DHBT [60]. This problem is addressed in more details in chapter 11.

33
(iii) Inserting a doping spike between the base and collector: Heavily doping a thin portion of the wide bandgap collector layer adjacent to the base has two effects: first, it reduces the effective potential barrier width, thus increasing the electron tunnelling probability; second, this layer together with the p'-base region form a dipole which creates an additional electric field shifting the conduction band spike to lower energies. In this case, a precise control of the doping is required to obtain the desired collector barrier lowering/narrowing. If the total number of ionised impurity atoms is too large, Zener breakdown may occur in the B-C junction. Figure 3.3 shows the output characteristics of an InGaP/GaAs DHBT in which the collector layer was composed of a 50Å GaAs spacer layer (n = 3×10^{16} cm\(^{-3}\)), 150Å InGaP doping spike (n = 1×10^{18} cm\(^{-3}\)), and 4800Å InGaP collector (n = 3×10^{16} cm\(^{-3}\)). In this structure the doping spike was relatively thick and too close to the base. Additionally, capacitance-voltage measurements of the B-C junction revealed that the high doping of the collector is not completely limited to the thin doping spike region. As a result, the device showed a small Zener breakdown voltage due to the direct band-to-band tunnelling of electrons. More careful design and growth of the doping spike collector structure has resulted in significantly improved DC and RF performance of InGaP/GaAs DHBTs in the present work (see chapter 11) as well as in other published works [4,104-105]. This method is shown to slightly reduce the breakdown voltage and Early voltage of the device as compared to the spacer design [104,106].

(iv) Step-grading of the B-C heterojunction: In material systems such as InGaP/GaAs and InP/InGaAs, where a continuous compositional grading is not practical, a step-graded structure using the quaternary material InGaAsP may be employed to smooth out the variation of conduction band at the heterojunction. InGaAsP lattice-matched to InP has widely been used to step-grade the B-C heterojunction of InP/InGaAs HBTs [107-109] and impressive DC and RF characteristics have been observed. However, the use of InGaAsP
Chapter 3

lattice-matched to GaAs for step-grading the B-C heterojunction of InGaP/GaAs HBTs has been suggested for the first time in the present work [60,110]. It will be shown in chapter 11 that introducing two thin layers of InGaAsP lattice-matched to GaAs with intermediate bandgap values between those of GaAs and In_{0.49}Ga_{0.51}P can significantly improve the high frequency performance of InGaP/GaAs HBTs without compromising the DC characteristics.

(v) Using a material system with zero conduction band offset: One example of such techniques is the use of Al_{0.11}Ga_{0.89}As near the collector end of the base region of InGaP/GaAs DHBTs [111]. It has been shown that the band alignment of Al_{x}Ga_{1-x}As/In_{0.49}Ga_{0.51}P heterojunctions changes from a type-I (for x < 0.11) to type-II alignment (for x > 0.11) [112-113]. Therefore, Al_{0.11}Ga_{0.89}As/In_{0.49}Ga_{0.51}P heterojunction is expected to have no conduction band discontinuity. This finding has been utilised in [111] to design an InGaP/AlGaAs/InGaP DHBT in which the Al mole fraction of the base region is graded from 0.21 near the emitter end to 0.11 near the collector end. Subsequently, very sharp saturation characteristics is observed in the output characteristics, demonstrating the absence of conduction band spike at the B-C heterojunction. However, this technique suffers from three major drawbacks. First, the transport properties of AlGaAs are poorer than GaAs. In particular, the electron and hole mobilities and minority carrier lifetimes in AlGaAs reduce sharply with increasing the Al mole fraction (see chapter 5 for more details). This may degrade the current gain and high frequency performance of the AlGaAs base HBTs as compared to GaAs base devices. Second, a large Al content near the emitter reduces the valence band discontinuity at the B-E heterojunction, thus decreasing the emitter injection efficiency and current gain of the device. The above two issues in using AlGaAs graded base HBTs are addressed in further details in sections 3.2.4 and 7.5. A third drawback of this technique relates to the fact that the conduction and valence band line-up of In_{0.49}Ga_{0.51}P is very much sensitive to growth conditions as well as substrate orientation and level of ordering (this is discussed in subsection 5.3.2). Therefore, although one growth method may result in negligible conduction band discontinuity between Al_{0.11}Ga_{0.89}As and In_{0.49}Ga_{0.51}P, wafers grown under different conditions may not perform as well.

(vi) Using short-period linearly graded chirped superlattices (CSLs) to grade the B-C heterojunction: A graded heterojunction can be approximated by a short-period superlattice, in which alternating layers of wide- and narrow-bandgap materials with varying thicknesses are periodically repeated. While the period of the superlattice is kept constant, the relative thicknesses of the wide- and narrow-bandgap materials can be designed such that the resultant average composition changes approximately linearly from the narrow-bandgap base material towards the wide-bandgap collector material. Both InP/InGaAs [114] and InAlAs/InGaAs [115-116] CSLs have been used for compositional grading of the InGaAs/InP B-C heterojunction of DHBs, and resulted in excellent DC and
high frequency performance. The all-arsenide InAlAs/InGaAs superlattice has the advantage that it does not involve a group V inter-mixing problem associated with growth switching between InP and InGaAs. But, the large conduction band offset of the InAlAs/InGaAs heterojunction necessitates a very small CSL period (on the order of 15Å), otherwise oscillations may occur in the I-V characteristics of the device [116].

In chapter 11 design of B-C heterostructures of InGaP/GaAs DHBTs will be considered in greater depth in order to optimise their high frequency performance.

3.2.2. Pnp HBTs

Most of the HBT works in the literature have concentrated on Npn devices mainly due to the superior transport properties of electrons in III-V compounds as compared to holes. However, Pnp HBTs have also attracted interest in recent years for inclusion with Npn HBTs in complementary HBT (CHBT) based circuits where the Pnp can be used as an active load or as a pull-up transistor in a push-pull amplifier [117-118]. Small hole mobility and diffusion coefficient not only increases the base transit time of Pnp HBTs compared to the Npn devices, but also increases the emitter and collector series resistances of the transistor. As a result, the microwave performance of Pnp HBTs are inferior to that of Npn HBTs.

On the positive side, the large electron mobility inside the base of Pnp HBTs allows much lower base doping and thickness as compared to Npn's, while keeping the base series resistance reasonably low [119]. A thin base (on the order of 300-500Å) reduces the base transit time, and a lower base doping increases both the hole diffusion coefficient and lifetime inside the base, all contributing to improved DC and high frequency performance of Pnp HBTs [120]. Compositional grading of the base material reduces the base transit time even further [121]. Although the hole mobility inside the collector of Pnp HBTs is much smaller than the electron mobility of Npn devices, the collector SCR delay times of the two HBTs are more or less comparable, since the saturation velocities of both types of carriers are almost similar (see chapter 5). Moreover, there are reports that suggest a larger hole minority carrier lifetime in some III-V compounds comparing to that of electrons [120]. Pnp HBTs also offer the added advantage of negligible emitter current crowding (due to the low base series resistance), which may make the design of multi-finger microwave power HBTs much more flexible [122]. All of these have resulted in reports of current gain for Pnp HBTs which are comparable to those of Npn devices [123], but the best reported peak $f_T$ and $f_{max}$ of Pnp's are limited to $< 70$GHz [124], which are much lower than the peak Npn performance.

As a final comment, it is useful to mention that material systems offering a large $\Delta E_C$, such as InAlAs/InGaAs and AlGaAs/GaAs, are more suitable for making Pnp HBTs. This is due to the fact that performance of Pnp HBTs is determined by transport of holes, not electrons.
3.2.3. Collector-Up HBTs

It will be shown in chapter 10 that the high frequency figures-of-merit of HBTs (i.e., $f_T$ and $f_{max}$) are much more sensitive to the B-C junction capacitance $C_{bc}$ than the B-E capacitance $C_{be}$. This is due to the fact that $C_{be}$ is almost shorted by the small resistance of the forward biased B-E junction when the transistor is operated under forward active mode. One way to reduce $C_{bc}$ is to design collector of the HBT on top, which then would require smaller area and improves the high frequency performance [20]. Consequently, collector-up (C-UP) HBTs have attracted considerable attention in the literature for high frequency applications. Additionally, the possibility to operate (D)HBTs in both C-UP and E-UP configuration simultaneously is quite desirable from the circuit designers' point of view [20,125]. However, the injection of electrons from extrinsic emitter into the extrinsic base in the C-UP configuration may significantly degrade the current gain and cutoff frequency of these devices, since all these electrons will be directly collected by the base ohmic contact and none of them would have the chance to reach the collector. Ion implantation of the extrinsic emitter region, either to damage this area or to turn it into p-type, may partly or completely cure this unwanted phenomenon [126-127]. Two-dimensional simulation of C-UP HBTs [128-129] has shown that in order to have a reasonable current gain and cutoff frequency in these devices, the effective area of active n-type emitter must be even slightly smaller than the collector mesa area. Other techniques to eliminate the electron injection into the extrinsic base region include a selective area regrowth technique [130], and a buried SiO$_2$ layer underneath WSi base electrode [125].

3.2.4. Graded Base HBTs

The current gain of bipolar transistors may be improved by grading the bandgap of the base region from a wide bandgap at the emitter end to a narrow bandgap at the collector end. The electric field produced inside the base adds a drift component to the electron current, which is otherwise purely dominated by the diffusion of minority carriers across the base. By sweeping electrons out of the base, the additional drift component improves the base transport factor, decreases the base transit time, and increases the current gain and the device speed. Base grading is particularly essential for III-V Pnp [121] and SiGe Npn HBTs [131], where a small diffusion constant for minority carriers inside the base may result in a large base transit time for uniform base devices. For GaAs-based HBT's, graded base structures using both Al$_x$Ga$_{1-x}$As [111,132] and strained In$_{1-x}$Ga$_x$As [14,93,133-134] as the base material have been investigated and resulted in improvement of the device performance, although the improvement in the latter case is shown to be more significant. The compositional grading of the base is discussed in more details in section 7.5, where it is shown that many design issues have to be carefully taken into account for an optimised performance of graded base HBTs. These issues include saturation of electron velocity at high electric fields, formation of a conduction band barrier at the B-C heterojunction (when...
the bandgap near the collector end is modified), modification of the emitter injection efficiency (when the bandgap near the emitter end is changed), carbon incorporation inside the base, strain due to lattice mismatch (in the case of In$_{1-x}$Ga$_x$As base), and degradation of base material properties (in the case of Al$_x$Ga$_{1-x}$As base).

3.3. Alternative Materials for HBTs

3.3.1. InP/In$_{0.53}$Ga$_{0.47}$As HBTs

InGaAs-based HBTs lattice-matched to InP substrates have emerged as an alternative technology to GaAs-based HBTs for high-speed low-power analogue, digital, and optoelectronic applications. InGaAs/InP devices offer numerous advantages due to the superior material properties: 1) a higher Γ-L valley separation which leads to pronounced velocity overshoot and shorter transit times; 2) a sufficiently large valence band discontinuity (0.34-0.42 eV) to suppress hole injection which results in higher current gain; 3) a lower surface recombination velocity which allows lateral scaling of the emitter area without compromising gain; 4) a smaller base bandgap which reduces the turn-on voltage and minimises the power consumption; 5) higher carrier mobilities which translate into superior minority carrier transport and lower bulk access resistances; 6) excellent specific contact resistance for non-alloyed ohmic contacts to both n- and p-type InGaAs; 7) better thermal conductivity of InP substrate; and 8) compatibility with 1.3-1.55 µm lightwave communication systems. As a result, excellent high frequency performance has been observed for these HBTs which surpasses those of the GaAs-based devices. One of the problems associated with the use of InP/InGaAs SHBTs is the extremely large leakage current associated with the narrow bandgap InGaAs collector material [135] which results in a small breakdown voltage [136]. Therefore, the usage of these devices is only limited to low- to medium-power applications [100]. As mentioned earlier, the utilisation of a wide bandgap InP collector with appropriate design of the B-C heterojunction may cure this problem.

3.3.2. In$_{0.52}$Al$_{0.48}$As/In$_{0.53}$Ga$_{0.47}$As HBTs

In$_{0.52}$Al$_{0.48}$As lattice-matched to InP with a room temperature bandgap of 1.45 eV may replace InP as the emitter material for InGaAs-based HBTs. InAlAs/InGaAs heterojunction offers $\Delta E_C = 0.52$ eV and $\Delta E_V = 0.18$ eV (see chapter 5), which seems much less favourable for Npn HBTs than its InP/InGaAs counterpart. Indeed, when the velocity overshoot and ballistic transport of electrons across the base material are ignored, it is shown in [56] that an InP emitter is quite advantageous to an InAlAs emitter. Nevertheless, InAlAs/InGaAs HBTs have become quite popular for high-speed applications. In fact the best high frequency performance of HBTs to date belongs to transferred-substrate InAlAs/InGaAs HBTs with an impressive $f_{\text{max}} > 800$GHz [2]. This is partly due to the fact that the large conduction band offset at the B-E heterojunction launches energetic electrons.
into the InGaAs base which can lead to a near-ballistic transport of electrons across the thin base layer. Another advantage of the InAlAs/InGaAs heterojunction system is that compositional grading can be achieved easily using the all-arsenide quaternary material In$_{0.53}$(Al$_x$Ga$_{1-x}$)$_{0.47}$As lattice-matched to InP; this does not involve group V intermixing. Also needless to say is that the large $\Delta E_C$ of InAlAs/InGaAs heterojunctions makes them most suitable for blocking back injection of electrons in Pnp HBTs [119].

### 3.3.3. InP/In$_{0.53}$Al$_{0.22}$Ga$_{0.25}$As Continuous-Conduction-Band HBTs

The relative alignment of conduction and valence bands for the most important III-V binary and ternary compounds is shown in Fig. 5.15. It can be seen in this figure that InP and In$_{0.52}$Al$_{0.48}$As form a type-II (or staggered) heterojunction system (i.e., with both the conduction and valence bands of In$_{0.52}$Al$_{0.48}$As above those of InP), while InP/In$_{0.53}$Ga$_{0.47}$As forms a type-I (or straddling) line-up. Therefore, one expects the conduction band offset of the InP/In$_{0.53}$(Al$_x$Ga$_{1-x}$)$_{0.47}$As heterojunction system to vanish for an intermediate composition between those of In$_{0.53}$Ga$_{0.47}$As and In$_{0.52}$Al$_{0.48}$As. A similar situation exists for the In$_{0.49}$Ga$_{0.51}$P/Al$_{0.11}$Ga$_{0.89}$As heterojunction system [111].

Liu et al. [137] demonstrated a continuous-conduction-band (CCB) HBT with InP emitter and In$_{0.53}$Al$_{0.22}$Ga$_{0.25}$As base, which exhibited a low offset voltage of 50 meV and better breakdown and output conductance characteristics compared with InGaAs base HBTs. However, the results in [137] slightly undermined by the fact that the authors still use spacer and doping spike layers between the base and emitter.

### 3.3.4. SiGe HBTs

The persisting tendency of microelectronics industry to use the firmly established silicon technology has led to the development of SiGe HBTs and MOSFETs. With the high level of integration and the low cost of development of Si technology as the main driving force, SiGe HBTs have emerged since late 1980's, and ever since have produced excellent device performance [138-139]. A compositionally graded, strained Si$_{1-x}$Ge$_x$ base layer grown on Si substrate can significantly reduce the base transit time, and hence improve both the device current gain and cutoff frequency [140]. Almost the entire bandgap discontinuity between Si and SiGe appears inside the valence band [138: subsection 2.1.3], which makes this material system suitable for high emitter injection efficiency in Npn HBTs. Addition of germanium, which has a high hole mobility, and the strain-induced energy shifts and distortions in the band diagram enhances the hole mobility of SiGe base layers, resulting in a reduced base series resistance [138: subsection 2.1.7]. On the negative side, growth of a high quality pseudomorphic SiGe on Si substrate is relatively complicated. Due to the narrow bandgap of both Si and Ge, SiGe HBTs are not the best choices for high power or high temperature applications. Additionally, since both Si and Ge are indirect bandgap materials, devices based on these materials can not emit or detect photons efficiently, and
therefore, they are not suitable for optical detectors and receivers.

3.3.5. InP/GaAs$_{0.51}$Sb$_{0.49}$/InP DHBTs

InP/GaAs$_{0.51}$Sb$_{0.49}$ displays a staggered band line-up with both the conduction and valence bands of GaAs$_{0.51}$Sb$_{0.49}$ above those of InP ($\Delta E_C = 0.18$ eV, $\Delta E_V = 0.81$ eV) [141]. This band line-up eliminates any possibility of the formation of current blocking at the B-C heterojunction. Bolognesi et al. [142] demonstrated a nearly ideal InP/GaAs$_{0.51}$Sb$_{0.49}$/InP DHBT, in which an abrupt B-C heterojunction did not show any current blocking effect. Although the emitter injection efficiency of an HBT with staggered band line-up is modified by a factor $\exp[(\Delta E_V - \Delta E_C)/kT]$ compared to a homojunction bipolar transistor, $(\Delta E_V - \Delta E_C)$ in this material system is still quite large such that back injection of holes is not an important issue. The early DC results for the DHBTs fabricated based on this new material system have been impressive: a very small offset voltage of 14meV, a turn-on voltage of 0.41V which is smaller than that of InP/InGaAs HBTs, an almost constant current gain of 50 for almost 7 decades of collector current, and a good breakdown voltage owing to the InP collector [141-142]. Additionally, the 0.18 eV conduction band offset at the B-C heterojunction may act as a ballistic electron launcher into the collector.

3.3.6. Nitride-Based HBTs

Recently, group III nitrides and SiC have attracted considerable attention for variety of high power, high temperature, and high frequency applications, due to their excellent transport as well as mechanical properties [143]. Among these materials, GaN has a direct bandgap of 3.4 eV, a reasonable electron mobility of 1200 cm$^2$/Vs, a breakdown field of $>5 \times 10^6$ V/cm, an electron saturation velocity of $2.5 \times 10^7$ cm/s, and a thermal conductance of 1.5 W/cmK which is three times larger than that of GaAs [143]. All of these has led to impressive early results, especially for AlGaN/GaN FETs, and a very bright prospect for future advancements. However, until recently, application of this material system for bipolar devices was hampered by the lack of an efficient shallow acceptor for p-type doping, an appropriate ohmic contact to p-type materials, and a good quality substrate [144]. McCarthy et al. [145] have recently reported the first AlGaN/GaN HBT with a Mg-doped base ($\text{Mg} = 1 \times 10^{19}$ cm$^{-3}$, $p = 5 \times 10^{17}$ cm$^{-3}$). This first device had a very low gain of only 3, a large offset voltage of ~5 V, and an unacceptable base sheet resistance and base ohmic contact resistance. Not so long after this first report, Shelton et al. [146] demonstrated an AlGaN/GaN HBT with improved base quality which showed a room temperature current gain as high as 100. Once the problems associated with the GaN p-type doping and ohmic contact are resolved, much better performance is expected due to the excellent intrinsic material properties. Alternatively, good quality p-SiC can be used as the base material in GaN/SiC HBTs. With GaN as the emitter and SiC as the base and collector, super high current gain ($> 10^6$) and operation up to 500°C has recently been
3.4. Alternative Heterojunction Bipolar Device Structures

3.4.1. Tunnelling Emitter Bipolar Transistors (TEBTs)

An alternative heterojunction bipolar device design has been proposed by Xu and Shur [148] in which the solid wide-gap emitter of a conventional HBT is replaced by only a thin layer of wide-gap material, typically less than 200Å thick, between emitter and base. The thin wide bandgap layer creates potential barriers both in the conduction and in the valence bands. While the electrons can easily tunnel through the barrier in the conduction band, the holes, due to their large effective mass, are repelled by the barrier in the valence band. Thus, hole injection from the base to the emitter is suppressed. Such transistor structures have been originally called tunnelling emitter bipolar transistors (TEBTs). Some other authors refer to this structure as hole barrier bipolar transistors (HBBTs) [149].

Bachem et al. [150] have studied the current gains of TEBTs with different InGaP barrier thicknesses of 20Å and 50Å, and compared them with a standard InGaP/GaAs HBT. Their results showed that the standard HBT exhibited a current gain of up to 65, whereas the TEBTs with 20Å and 50Å thick InGaP layers gave values of 10 and 115, respectively. This experiment shows that holes from the base are only prevented from being injected into the emitter if the InGaP barrier layer is thick enough, at about 50Å. This is still thin enough to allow electrons to tunnel through and improve the gain further compared with a conventional HBT. On the other hand, Liu et al. [31] have performed a systematic study of the emitter thickness effect on the DC characteristics of InGaP/GaAs HBTs. They found that the current gain of HBTs, especially at low current levels, continuously increases when the InGaP emitter thickness increases from 50Å to 1000Å. They explained this based on the fact that when the InGaP emitter is thin, it would be fully depleted and then the B-E depletion region extends into the narrow bandgap GaAs cap layer, where a significant SCR recombination occurs, degrading the current gain.

Lauterbach et al. [151] also observed a large base leakage current at low biases for their TEBT with 50Å InGaP hole barrier, but other DC features of their device were quite acceptable. Finally, Leier et al. [149] used a slightly thicker (200Å) InGaP barrier layer, and demonstrated good DC and high frequency performance for their self-aligned InGaP/GaAs TEBT. Their result included a small signal current gain of $\sim 30, f_T = 95$ GHz, and $f_{max} = 110$ GHz, which was the best high frequency performance for InGaP/GaAs HBTs at that time.

3.4.2. Heterostructure-Emitter Bipolar Transistors (HEBTs)

Although conventional HBTs offer many advantages over homojunction transistors, the implementation of an HBT is far more complicated. For instance, it requires a precise
control of the heterojunction interface. A perfect alignment of the compositional junction to the doping junction is essential for good device performance. Such an alignment is sometimes difficult to achieve because of the high diffusivity of the dopants usually employed inside the base (e.g., Be or Zn). Another important disadvantage of conventional HBTs is the noticeable collector-emitter offset voltage, $V_{CE,offset}$, due to the turn-on voltage difference between the B-E heterojunction and the B-C homojunction. This collector offset voltage is undesirable, since it increases the power consumption in saturation logic circuits.

In 1988, a new structure of HBT was proposed in which the base-injected carriers are confined by band offset inside the emitter so that a high injection efficiency is maintained [152]. Such a transistor has the merits of both an HBT and a homojunction bipolar transistor. Fig. 3.4 shows the energy band diagram of the new device. Compared with a conventional HBT, the heterointerface is moved away from the B-E junction. As illustrated by the figure, the carrier injection mechanism is the same as in a homojunction transistor. The influence of the conduction band discontinuity is minimised, since it is pushed into the quasi-neutral region of the emitter, thereby reducing the B-E turn-on voltage and $V_{CE,offset}$. However, the hole injection barrier still exists, blocking the base-injected carriers into the emitter.

The first transistor of this kind, which was called a heterostructure-emitter bipolar transistor (HEBT), was successfully fabricated by Wu et al. [153] using AlGaAs/GaAs material system. A current gain of 28 were achieved for a device with the base doping and thickness of $1 \times 10^{19}$ cm$^{-3}$ and 1500Å, respectively. The GaAs emitter layer of this structure was 400Å thick and doped to $1 \times 10^{18}$ cm$^{-3}$, followed by a 1000Å n-Al$_{0.3}$Ga$_{0.7}$As confinement layer. Thei et al. [154] have fabricated and discussed the performances of InGaP/GaAs and AlGaAs/GaAs HEBTs. From their calculation, an emitter thickness not smaller than 230Å and 320Å was found necessary to eliminate the B-E potential spike for InGaP and AlGaAs HEBTs, respectively. More recently, Yang et al. [155-156] discovered

![Figure 3.4- Schematic band diagram of a HEBT under normal operation mode.](image)
that the current gain of InGaP and AlGaAs HEBTs are much less sensitive to the variation of temperature than their HBT counterparts. This last point is further elaborated in section 7.4.

Unfortunately, the large SCR recombination rate inside the narrow bandgap emitter layer of HEBTs degrades their current gain, especially at low to medium current levels. A 700Å GaAs emitter was found the main reason for the gain degradation in an InGaP/GaAs HEBT [154]. As a result, recent HEBTs are being fabricated with a narrow bandgap emitter thickness of ≤300Å [40,155-157], which makes them no more than a conventional HBT with a B-E spacer design.

3.4.3. Other Heterojunction Bipolar Device Structures

In recent years some variants of the HEBT structure have been suggested. Tsai et al. [158-159] reported a heterostructure-emitter and heterostructure-base transistor (HEHBT) which has an emitter structure exactly similar to a HEBT, but the base is composed of a 100Å In$_{0.2}$Ga$_{0.8}$As layer on top of a 1000Å GaAs region. The advantages of this new structure are claimed to be a larger B-E valence band discontinuity for hole confinement, and a smaller surface recombination velocity of the exposed InGaAs extrinsic base. However, the growth of the lattice mismatched InGaAs layer may result in a degraded quality of base and emitter regions. Additionally, the InGaAs quantum-well inside the base may significantly increase the base bulk recombination, while the main advantages of an InGaAs base are not exploited in this structure. No RF results has been reported for this new device, and it is quite likely that the carrier trapping inside the InGaAs quantum-well seriously degrades the high frequency performance.

Another new structure, an n-Al$_{0.3}$Ga$_{0.7}$As/n-Al$_{0.2}$Ga$_{0.8}$As/p-GaAs dual-emitter HBT, has been proposed by Chor and Peng [160], which is very similar to the HEBT structure, but with the GaAs emitter layer replaced by an Al$_{0.2}$Ga$_{0.8}$As emitter. Numerical simulations suggest that this new structure is advantageous over Al$_{0.3}$Ga$_{0.7}$As/GaAs and Al$_{0.2}$Ga$_{0.8}$As/GaAs HBTs as well as Al$_{0.3}$Ga$_{0.7}$As/GaAs HEBTs, in terms of a higher current gain and cutoff frequency, and a less sensitive current gain to the variations of both collector current and temperature [160]. Experimental verification of this new structure has not been published yet.
A semiconductor device is characterised by a set of physical parameters, e.g., mobility, lifetime, as well as technological parameters like geometry and impurity profile. The aim of device modelling is then to derive from this set a field of electrostatic potential and carrier concentration for electrons and holes in space (and time). These three quantities yield, in turn, vector parameters like electric field and current density. Finally, integration of the latter vector over respective contact areas results in the terminal characteristics of a device.

In trying to achieve closed form analytical solutions of the non-linear basic equations, approximations are necessary, e.g., with respect to doping profiles, space-charge density, recombination models, etc. Although the resulting analytical models give good insight into the physical device behaviour, the underlying assumptions often yield an oversimplified picture. This disadvantage can be avoided if the device equations are solved by numerical means in which the complete system of basic equations are solved without simplifications.

In this thesis, a physics-based numerical model is developed to study the physical mechanisms responsible for carrier transport in HBTs and to design various layers of the transistor for an optimised performance. This model takes into account many of the important physical phenomena such as bandgap narrowing (BGN), carrier degeneracy, thermionic-field emission (TFE) across an abrupt heterojunction interface, and spatial variation of band parameters. This part of the report presents an overview of the methods used to numerically simulate HBTs in one spatial dimension. Also the calculated results are compared against the measured data. This provides some confidence on the model for later use in device design and optimisation.

Chapter 4 discusses the derivation of the basic semiconductor equations, taking into account Fermi-Dirac statistics and TFE of the carriers across an abrupt heterojunction. Also the boundary conditions for the basic equations are explained in this chapter. In chapter 5 we review the empirically or theoretically founded models of the underlying physical parameters such as carrier mobilities, carrier lifetimes, and the amount of bandgap narrowing. One of the strong factors in the present numerical model is its capability to consider various III-V semiconductor compounds in the design of HBT layers. A considerable number of published theoretical and experimental reports are reviewed to include the physical parameters for GaAs, AlGaAs, InGaP, InP, InGaAs, InAlAs, and InGaAsP in the model. This not only gives a good degree of freedom to the designer as to the choice of material, but also allows a direct comparison between various types of HBTs such as AlGaAs/GaAs, InGaP/GaAs, InP/InGaAs, and InAlAs/InGaAs. Then, chapter 6 deals with the numerical techniques to discretise, linearise, and solve the system of partial differential equations describing the charge transport in semiconductor devices. Finally, in chapter 7 the results of the numerical model are compared against measured data. A number of physical phenomena are also studied to demonstrate the capability of the simulation program.
CHAPTER 4. BASIC SEMICONDUCTOR EQUATIONS

4.1. Basic Equations

The basic semiconductor equations include the Poisson’s equation and the current continuity equations for electrons and holes. The Poisson’s equation can be derived from the differential form of the Gauss’s theorem:

\[
\text{div}(\mathbf{D}) = \rho
\]  

(4.1)

where \( \mathbf{D} \) and \( \rho \) are the displacement vector and charge density, respectively, and can be expressed as:

\[
\mathbf{D} = \varepsilon \cdot \mathbf{E} = -\varepsilon \cdot \text{grad}(\psi)
\]  

(4.2)

\[
\rho = q \left( p - n + N_D^+ - N_A^- \right)
\]  

(4.3)

where \( \psi \) is the electrostatic potential, \( \varepsilon \) is the dielectric constant, \( n \) and \( p \) are the electron and hole concentrations, \( \mathbf{E} \) is the electric field vector, and \( N_D^+ \) and \( N_A^- \) are the ionised donor and acceptor densities, respectively. The continuity equations are written as [161]:

\[
\frac{\partial n}{\partial t} + R = \frac{\text{div}(\mathbf{J}_n)}{q}
\]  

(4.4)

\[
\frac{\partial p}{\partial t} + R = -\frac{\text{div}(\mathbf{J}_p)}{q}
\]  

(4.5)

where \( \mathbf{J}_n \) and \( \mathbf{J}_p \) are the electron and hole current densities, respectively, and \( R \) is the recombination rate. Assuming steady state condition and only one spatial dimension and using (4.2) and (4.3) we can rewrite the equations (4.1), (4.4), and (4.5) as:

\[
\frac{d}{dz} \left( \varepsilon \frac{d\psi}{dz} \right) = q \cdot \left( n - p + N_A^- - N_D^+ \right) \quad \text{Poisson’s Equation (4.6)}
\]

\[
\frac{d\mathbf{J}_n}{dz} = q \cdot R \quad \text{Electron Continuity Equation (4.7)}
\]

\[
\frac{d\mathbf{J}_p}{dz} = -q \cdot R \quad \text{Hole Continuity Equation (4.8)}
\]

In order to have a complete set of equations we need some relationships between current densities and the dependent variables (\( \psi, n, \) and \( p \)). It has been shown [162] that for an isotropic material in those regions where current density is dominated by drift and diffusion, \( \mathbf{J}_n \) and \( \mathbf{J}_p \) can be written as:

\[
\mathbf{J}_n = n \mu_n \frac{dE_{fn}}{dz}
\]  

(4.9)
Chapter 4

Basic Semiconductor Equations

\[ J_p = p \mu_p \frac{dE_{fp}}{dz} \] (4.10)

Here \( E_{fn} \) and \( E_{fp} \) represent the quasi-Fermi energies for electrons and holes, respectively. Equations (4.9) and (4.10) are quite general and can be used for degenerate materials as well as materials with non-uniform band structure. However, we still need some equations relating the quasi-Fermi energies to the carrier densities. Using Fermi-Dirac statistics one can express the carrier densities with the following equations [161]:

\[ n = N_C \cdot F_{1/2} \left( \frac{E_{fn} - E_C}{kT} \right) \] (4.11)

\[ p = N_V \cdot F_{1/2} \left( \frac{E_V - E_{fp}}{kT} \right) \] (4.12)

\( F_{1/2} \) is the Fermi-Dirac integral of order \((1/2)\) defined by:

\[ F_{1/2}(\eta) = \frac{2}{\sqrt{\pi}} \int_0^\infty \frac{\sqrt{x}}{1 + \exp(x - \eta)} \, dx \] (4.13)

If we define:

\[ \gamma_{n,p} = \frac{F_{1/2}(\eta_{n,p})}{\exp(\eta_{n,p})} \] (4.14)

where \( \eta_{n,p} \) are the arguments of \( F_{1/2} \) in (4.11) and (4.12), then one can easily derive:

\[ J_n = q \cdot V_T \cdot \mu_n \cdot \frac{dn}{dz} - q \cdot n \cdot \mu_n \cdot \frac{d\psi_n}{dz} \] (4.15)

\[ J_p = -q \cdot V_T \cdot \mu_p \cdot \frac{dp}{dz} - q \cdot p \cdot \mu_p \cdot \frac{d\psi_p}{dz} \] (4.16)

\( V_T (= kT/q) \) is the thermal voltage, and \( \psi_n \) and \( \psi_p \) are defined by:

\[ \psi_n = \psi + \frac{\chi - \chi_r}{q} + kT \cdot \frac{\ln \left( \frac{N_C}{N_{Cr}} \right)}{q} \] (4.17)

\[ \psi_p = \psi + \frac{\chi - \chi_r}{q} + \frac{E_g - E_{gr}}{q} - kT \cdot \frac{\ln \left( \frac{N_V}{N_{Vr}} \right)}{q} \] (4.18)

In the above two equations, \( \chi \) is the electron affinity of the material and all the parameters with subscript 'r' represent the values of that parameter at a reference point. Also because only gradients of \( \psi_n \) and \( \psi_p \) appear in the current equations, we have added the values at the reference point (which are constant). There is no need for the reference point to be a real point in the device. It can be defined as an intrinsic point \( \gamma_{nr} = \gamma_{pr} = 1 \) in one semiconductor.
material (GaAs in the present simulation program). The first term on the right-hand-side of 
(4.15) and (4.16) is due to the gradient of carrier densities and is called the diffusion current, 
while the second term is originated from the effective electric field and so is called the drift 
term. It is worth mentioning that although the forms of (4.15) and (4.16) suggest that the 
Einstein relation \( D / \mu = kT / q = V_T \) is used for modelling the diffusion constant (which is not 
holding for the highly-doped degenerate materials), a further refinement of (4.15)-(4.18) will 
prove that, in fact, a "modified-Einstein relationship" [161] holds between the diffusion 
constant and mobility:

\[
\begin{align*}
\frac{D_n}{\mu_n} &= V_T \cdot \frac{F_{1/2}}{F_{-1/2}} \left( \frac{E_{fn} - E_C}{kT} \right) \\
\frac{D_p}{\mu_p} &= V_T \cdot \frac{F_{1/2}}{F_{-1/2}} \left( \frac{E_V - E_{fp}}{kT} \right)
\end{align*}
\] (4.19)

where \( F_{1/2} \) is the Fermi-Dirac integral of order (-1/2) (see Appendix C).

By definition, we have:

\[
\begin{align*}
\text{grad}(E_{fn}) &= -q \cdot \text{grad}(\phi_n) \\
\text{grad}(E_{fp}) &= -q \cdot \text{grad}(\phi_p)
\end{align*}
\]

Here \( \phi_n \) and \( \phi_p \) are the quasi-Fermi potentials for electrons and holes, respectively. Thus:

\[
\begin{align*}
E_{fn} &= C_1 - q \cdot \phi_n \\
E_{fp} &= C_2 - q \cdot \phi_p
\end{align*}
\]

Also the vacuum energy level can be expressed as:

\[
E_{\text{vac}} = E_0 - q \cdot \Psi
\] (4.20)

\( C_1, C_2, \) and \( E_0 \) are some arbitrary constant energies. Therefore, from the above equations and 
(4.11), (4.12), and (4.14) one can write:

\[
\begin{align*}
n &= n_i \cdot \omega \cdot \gamma_n \cdot \exp \left[ \frac{C_1 - E_0 + q \cdot (\Psi - \phi_n)}{kT} \right] \\
p &= \frac{n_i \cdot \gamma_p}{\omega} \cdot \exp \left[ \frac{E_0 - C_2 + q \cdot (\phi_p - \Psi)}{kT} \right]
\end{align*}
\]

where:
In order to avoid numerical problems due to calculation of exponential terms for large values of \( \chi + E_g / 2 \), we define \( C_1, C_2, \) and \( E_0 \) as follows:

\[
C_1 = C_2 = 0
\]

\[
E_0 = kT \cdot \text{Ln}(\omega_r) = \frac{kT}{2} \cdot \text{Ln}\left(\frac{N_{Cr}}{N_v}\right) + \chi_r + \frac{E_{gr}}{2}
\]

Here \( \omega_r \) is the value of \( \omega \) at the reference point. Finally, we have:

\[
\begin{align*}
E_{fn} &= -q \cdot \phi_n \\
E_{fp} &= -q \cdot \phi_p
\end{align*}
\]

\[
\begin{align*}
n &= n_i \cdot \gamma_n \cdot \left( \frac{\omega}{\omega_r} \right) \cdot \exp\left( \frac{\psi - \phi_n}{V_T} \right) \\
p &= n_i \cdot \gamma_p \cdot \left( \frac{\omega_r}{\omega} \right) \cdot \exp\left( \frac{\phi_p - \psi}{V_T} \right)
\end{align*}
\]

It is well known that the drift-diffusion transport model as described above is not valid when there are discontinuities in the energy bands larger than \( kT \) in a length shorter than the mean free path [163]. This is the case for abrupt HBTs. Transport through the base-emitter interface, where discontinuities in the energy bands occur, is controlled by thermionic emission and tunnelling transmission through the spike [164-165]. Therefore, the formulations of (4.9) and (4.10) are no longer valid for abrupt heterointerfaces. The thermionic emission and tunnelling mechanisms across an abrupt heterojunction interface are taken into account by using an expression of TFE current formulated based on the WKB approximation [166: p.34] as a boundary condition at the abrupt heterointerface which eventually limits the current transport over the barrier while maintaining the current continuity [53].

The band diagram of an abrupt heterojunction is shown in figure 4.1. The TFE boundary condition for electrons at the interface with a general shape is expressed as [56]:

\[
\omega = \sqrt{\frac{N_C}{N_V}} \cdot \exp\left( \frac{\chi + E_g / 2}{kT} \right)
\]

\[
n_i = \sqrt{N_C \cdot N_V} \cdot \exp\left( -\frac{E_g}{2kT} \right)
\]
Figure 4.1- Schematic conduction band diagram of an abrupt heterojunction. $E_{f_{1,2}}$ represent the electron quasi-Fermi levels in each semiconductor region.

$$n(z_j) = \exp\left(\frac{E_C(z_j^-) - E_{C,\text{max}}}{kT}\right)$$

$$n(z_j^+) = \exp\left(\frac{E_C(z_j^+) - E_{C,\text{max}}}{kT}\right)$$

Here $A^*$ is the effective Richardson constant for electrons:

$$A^* = \frac{4\pi \cdot m_n^* \cdot q \cdot k^2}{h^3}$$

$z_j$ is the position of the junction (heterointerface), and $E_{C,\text{max}}$ is defined by:

$$E_{C,\text{max}} = \text{Max}\left\{E_C(z_j^-), E_C(z_j^+)\right\}$$

The contribution of tunnelling is formulated through a parameter $\delta$ being evaluated based on the conduction band profile. As to the value of the effective mass used in equation (4.28), it is a common practice to use the smaller value of $m_n^*$ between the two sides of the heterointerface [167]. Recently a more detailed study of the thermionic emission current across an abrupt heterointerface [168] revealed that the above common practice is not always true. However, in the same paper [168] it has been shown that the modifications in the magnitude of current introduced by the more accurate formulation of effective mass is not significant. Therefore, in the present work, the minimum of $m_n^*$ of the two sides will be used in the calculation of the Richardson constant.

In the present thesis, there are two major modifications to the formulation of TFE current as compared to the original work of Yang, et al., [169]. First, is the inclusion of $\gamma_n$ terms in the denominator of the terms in bracket in (4.27). These terms only reflect an approximate treatment of Fermi-Dirac statistics (not an accurate treatment) for the formulation of TFE current. The $\gamma_n$ terms have a minimal effect on the current transport across a p-n heterojunction at low to moderate bias range, since the heterojunction is depleted from electrons. At high forward bias regime, where the number of electrons injected to the base
side of the junction increases and the Fermi-Dirac statistics starts to come to term, the terminal currents of the device are limited by the series resistances rather than the current transport across the heterojunction. Therefore, the inclusion of $\gamma_h$ terms has a minimal contribution on the TFE current across a p-n heterojunction. On the other hand, the Fermi-Dirac statistics has a major effect on the current transport across an N-n heterojunction. The omission of the $\gamma_h$ terms in the calculation of terminal currents of an HBT with n$^+$ cap layer will result in a huge underestimation of collector current, which would be limited by current transport across n$^+$-N cap layer-emitter heterojunction rather than the N-p emitter-base heterojunction!

The second modification to the original work of [169] is the generalisation of the calculated tunnelling parameter, $\delta$, for use in all possible shapes of conduction band. This has been further elaborated in figure 4.2. First the shape of the conduction band near the interface is examined to see if the tunnelling is possible. If the tunnelling factor is non-zero, minimum of the conduction band on both sides of the interface is found ($z_L$ and $z_R$). This can be the point on the interface, a local minimum of the conduction band, the ending point of the layer, or the point 400Å away from the interface\(^\dagger\), whichever reaches first (see Appendix B). Finally, $\delta$ is evaluated using the following relation which is an improved version of that used in [47,169]:

$$\delta = \frac{1}{kT} \left[ \frac{E_{C_{\text{max}}}}{E_{\min}} - \frac{E_z}{kT} \right] \exp \left\{ \frac{4\pi}{\hbar} \int_{2m^*_n} \left( E_C(z) - E_z \right) dz \right\} \cdot dE_z \quad (4.30)$$

where $E_z$ is the energy of the conduction band at $z = z_E (z_L \leq z_E \leq z_j \text{ or } z_j \leq z_E \leq z_R)$,

\(^\dagger\) 400Å is considered as the maximum barrier thickness for which the tunnelling probability is non-zero.
and \( E_{\text{min}} = \max\{E_C(z_L), E_C(z_R)\} \). In fact the program developed during this work is written even more generally to include the contribution of tunnelling through any local maximum in the conduction band (not necessarily in an abrupt heterojunction). For example, a conduction band spike may occur where the grading of Al in graded AlGaAs/GaAs HBTs starts [51], and the simulation program is capable of taking the effect of tunnelling current through this spike into account. More details regarding the programming techniques for calculation of tunnelling factor are given in Appendix B.

Due to the large effective mass of holes in III-V compounds, the contribution of tunnelling for hole current density is ignored. Therefore, the thermionic emission boundary condition is applied at the interface:

\[
J_{p,\text{interface}} = A_h^* T^2 \left[ \frac{p(z^-_j)}{N_V(z^-_j) \gamma_p(z^-_j)} \cdot \exp \left( \frac{E_{V,\text{min}} - E_V(z^-_j)}{kT} \right) + \frac{p(z^+_j)}{N_V(z^+_j) \gamma_p(z^+_j)} \cdot \exp \left( \frac{E_{V,\text{min}} - E_V(z^+_j)}{kT} \right) \right] \tag{4.31}
\]

Again in (4.31) \( A_h^* \) is the effective Richardson constant for holes (using the smaller effective mass of holes between the two sides), and:

\[
E_{V,\text{min}} = \min \{E_V(z^-_j), E_V(z^+_j)\} \tag{4.32}
\]

### 4.2. Boundary Conditions

In order to uniquely determine the solution to a set of differential equations, proper boundary conditions must be known. In the case of our one-dimensional modelling, the boundaries include the emitter (top) and collector (bottom) contacts. However, because we need to apply a constant voltage or current to the (imaginary) base contact, we must define some boundary conditions for this contact too.

#### 4.2.1. Emitter and Collector Ohmic Contacts

Usually the assumptions of thermal equilibrium \( E_{fn} = E_{fp} = E_f \) and charge neutrality \( \rho = 0 \) are used to find a boundary condition for each of the continuity equations. Therefore:

\[
E_{fn} = E_{fp} \rightarrow n_0 p_0 = N_C \cdot N_V \cdot \gamma_n \cdot \gamma_p \cdot \exp \left( \frac{E_V - E_C}{kT} \right) \rightarrow n_0 p_0 = n_i^2 \cdot \gamma_n \cdot \gamma_p \tag{4.33}
\]

\[
\rho = 0 \rightarrow n_0 - p_0 = N_D^+ - N_A^- \tag{4.34}
\]

(4.33) and (4.34) can be used to find \( n_0 \) and \( p_0 \). Sometimes we assume that the surface recombination velocity at the ohmic contact is infinite and then:
\[
\begin{align*}
\begin{cases}
    n &= n_0 \\
    p &= p_0
\end{cases}
\end{align*}
\]

(4.35)

However, in the case of finite surface recombination velocity for the minority carriers \(S_p\), the minority carrier density will be modified as [170]:

\[
\begin{align*}
\begin{cases}
    n &= n_0 \\
    p &= p_0 - \frac{J_p}{q \cdot S_p}
\end{cases}
\end{align*}
\]

(4.36)

The minus sign in front of \(J_p\) in the above boundary condition results from the definition of the current density direction, which is assumed to be inward for the emitter contact and outward for the collector (i.e., the same direction as \(z\)-axis), while the real directions of the emitter and collector current in the normal active mode of the transistor operation is just the opposite.

For the boundary condition of the Poisson equation, two different cases are considered:

(i) **Zero series resistance**: In the equations relating the carrier densities to the quasi-Fermi potentials (i.e., (4.25) and (4.26)), we simply set the quasi-Fermi potential for the majority carriers equal to the applied external voltage \(V_g\) or \(V_c\) [170]. Then:

\[
\phi_n = V_{E,C} \rightarrow \psi = V_{E,C} + V_T \cdot \ln \left( \frac{n_0}{(\omega/\omega_T) \cdot \gamma_n \cdot n_i} \right)
\]

(4.37)

(ii) **Non-zero series resistance**: In the case of non-zero series resistance we can not assume that the quasi-Fermi potential for the majority carriers is equal to the externally applied voltage. Rather, we have:

\[
\begin{align*}
\begin{cases}
    \frac{V_C - \phi_n}{R_C} &= -J_C \cdot A_E \\
    \frac{V_E - \phi_n}{R_E} &= +J_E \cdot A_E
\end{cases}
\end{align*}
\]

(4.38)

In the above equations \(V_{C,E}\) represent the externally applied voltage to collector/emitter terminals, \(A_E\) is the active area (emitter area) of the transistor, and again +/- signs are coming from conventional definition of current direction in emitter and collector. \(R_{C,E}\) are the collector/emitter series resistances and can include such terms like ohmic contact resistance, externally connected series resistance, and the sub-collector horizontal series resistance.

Finally, the boundary condition for electrostatic potential reads:

\[
\begin{align*}
\begin{cases}
    \psi_C &= (V_C + J_C \cdot A_E \cdot R_C) + V_T \cdot \ln \left( \frac{n_0}{(\omega/\omega_T) \cdot \gamma_n \cdot n_i} \right) \\
    \psi_E &= (V_E - J_E \cdot A_E \cdot R_E) + V_T \cdot \ln \left( \frac{n_0}{(\omega/\omega_T) \cdot \gamma_n \cdot n_i} \right)
\end{cases}
\end{align*}
\]

(4.39)


4.2.2. Base Contact

Since in one dimensional modelling the base contact (usually defined as the midpoint in the base layer) is an imaginary contact, we can not use the thermal equilibrium assumption, and therefore, we have just two boundary conditions: one for the electrostatic potential, which replaces the Poisson equation, and the other for the charge neutrality, which is used in place of hole current continuity. The electron continuity equation is solved directly in the base contact, and then the hole concentration is obtained from:

$$p_B - n_B = N_A^- - N_D^+$$  \hspace{1cm} (4.40)

In order to plot the output characteristics of a bipolar transistor, one needs to set the base current to a constant value. For this reason, the ability to have a current-controlled base contact is included in the present model. Consequently, we have two types of boundary conditions for the electrostatic potential at the base contact:

(i) **Voltage-controlled ohmic contact:** Following the same trend as used for collector and emitter ohmic contacts yields:

$$\frac{V_B - \phi_p}{R_B} = -J_B \cdot A_E$$  \hspace{1cm} (4.41)

and therefore:

$$\psi = (V_B + J_B \cdot A_E \cdot R_B) - V_T \cdot \ln \left( \frac{p_B \cdot (\omega/\omega_r)}{n_i \cdot \gamma_p} \right)$$  \hspace{1cm} (4.42)

(ii) **Current-controlled ohmic contact:** In this case, the base series resistance is not important. If we define 'M' as the index of the base ohmic contact mesh point, then the following equation can be used as a boundary condition at the base contact [171]:

$$J_B = \frac{-I_B}{A_E} = J_E - J_C = J_n(z_{M-1/2}) + J_p(z_{M-1/2}) - J_n(z_{M+1/2}) - J_p(z_{M+1/2})$$  \hspace{1cm} (4.43)

In order to use the above equation as a boundary condition, one should replace the current densities on the RHS of (4.43) with their corresponding relations at the mid-interval points (see Eqs. (6.25) and (6.27)).

53
CHAPTER 5. PHYSICAL PARAMETERS IN III-V SEMICONDUCTORS

The basic semiconductor equations, the derivation of which previously discussed in chapter 4, just determine the structure of the set of equations which must be solved in order to simulate the internal behaviour of a device. As we have already noticed, some additional physical parameters are inherently associated with the basic semiconductor equations. Any quantitative, or even qualitative, simulation of a device relies heavily on applicable models for these parameters. Therefore, this chapter deals with the appropriate models for the physical parameters like energy bandgap, carrier mobility, recombination rate, dielectric constant, etc. Also some important phenomena such as bandgap narrowing due to heavy doping and incomplete ionisation of dopant atoms will be discussed here.

One of the strong factors in the present numerical model is its capability to consider various III-V semiconductor compounds in the design of HBT layers. A large number of published theoretical and experimental reports are reviewed to include the physical parameters for GaAs, AlGaAs, InGaP, InP, InGaAs, InAlAs, and InGaAsP in the computer program developed in this work. This not only gives the designer a good degree of freedom as to the choice of material, but also allows a direct comparison between various types of HBTs such as AlGaAs/GaAs, InGaP/GaAs, InP/InGaAs, and InAlAs/InGaAs. An attempt has been made to allow the user choose arbitrary mole fractions for the above materials, although the majority of the simulations performed in this work are for compounds lattice-matched to GaAs or InP substrates. However, due to the very limited experimental data on some compound materials (like InGaP, InAlAs, and InGaAsP), one has to consider interpolation schemes as the only available option to model the variation of some parameters in a continuous range of mole fraction. In these cases, many variations of interpolation schemes are studied to find the best fit to the sometimes limited reported data.

5.1. Interpolation Schemes

Material parameters of III-V binary compounds such as GaAs, InP, GaP, and InAs (and ternary compounds such as In_{1-x}Ga_{x}As and Al_{x}Ga_{1-x}As, to some extent) have received considerable attention in the past and many experimental data and theoretical studies for these parameters can be found in the literature. Bandgap energy and transport related parameters of quaternary and other ternary compounds, on the other hand, has been the topic of few or no experimental/theoretical publications. Furthermore, it is almost impractical to study and measure the parameters of quaternary materials, such as In_{1-x}Ga_{x}As_{y}P_{1-y}, for all possible combinations of x and y. These facts necessitate the use of an interpolation scheme. Such a scheme is essentially based on known values of the physical parameters for the related binary and ternary alloys. In this section various interpolation methods used for physical parameters of III-V ternary and quaternary compounds are summarised.
For many parameters of ternary compounds (generally shown as $A_{1-x}B_xC$), such as lattice constant, a linear interpolation between the corresponding values of the binaries $AC$ and $BC$ (known as Vegard’s law) would give an accurate estimate of the ternary parameter, $T$ [172]:

$$T_{ABC}(x) = (1-x) \cdot B_{AC} + x \cdot B_{BC}$$  \hspace{1cm} (5.1)

where $B_{AC}$ and $B_{BC}$ are the corresponding parameters for the binary materials. However, some material parameters, such as energy bandgaps and thermal conductivity, exhibit strong non-linearity with respect to the alloy composition [173-174]. In such cases, a quadratic interpolation of the form:

$$T_{abc}(^+) + ^ = \frac{abc}{x(l-x)} (5.2)$$

would give a much better fit to the experimental data. In the above equation, $c_{ABC}$ is the so-called bowing parameter for the ternary material $A_{1-x}B_xC$. As to the quaternary material $A_{1-x}B_xC_yD_{1-y}$, some parameters again obey the linear interpolation in terms of the parameters of the ending binaries:

$$Q_{ABCD}(x, y) = (1-x)y \cdot B_{AC} + (1-x)(1-y) \cdot B_{AD} + xy \cdot B_{BC} + x(1-y) \cdot B_{BD}$$  \hspace{1cm} (5.3)

If relationships for the ternary material parameters are available, the quaternary parameter can be expressed as [172]:

$$Q_{ABCD}(x, y) = \frac{x(1-x)[(1-y) \cdot T_{ABD} + y \cdot T_{ABC}]}{x(l-x)} + y(l-y)\frac{[l-x) \cdot T_{ACD} + x \cdot T_{BCD}]}{y(l-y)}$$  \hspace{1cm} (5.4)

If the ternary parameters are expressed quadratically as in (5.2), then (5.4) can be rewritten as:

$$Q_{ABCD}(x, y) = (1-x)y \cdot B_{AC} + (1-x)(1-y) \cdot B_{AD} + xy \cdot B_{BC} + x(1-y) \cdot B_{BD} -$$

$$c_{ABC} \cdot (1-x)xy - c_{ABD} \cdot (1-x)x(1-y) - c_{ACD} \cdot (1-y)y(l-y) - c_{BCD} \cdot xy(l-y) + m_Q \cdot \Delta_Q$$  \hspace{1cm} (5.5)

with $m_Q = 1$ and $\Delta_Q$ given by:

$$\Delta_Q = \frac{xy(1-x)(1-y)}{x(l-x) + y(l-y)}[c_{ABC} \cdot y + c_{ABD} \cdot (1-y) + c_{ACD} \cdot (1-x) + c_{BCD} \cdot x]$$  \hspace{1cm} (5.6)

For all the binary and ternary compounds composing $A_{1-x}B_xC_yD_{1-y}$, $\Delta_Q$ approaches zero. Some theoretical works in the literature (e.g., for bandgap of InGaAsP in [175]) have used a simpler form of (5.5) in which the last term has been ignored by setting $m_Q = 0$. In this work, we have used $0 \leq m_Q \leq 1$ as a fitting parameter to fit the interpolation equation (5.5) to the experimental data given for InGaAsP in the literature.

There may be circumstances where one can assume $c_{ACD} = c_{BCD} = 0$. For instance, it is generally believed that the bandgap bowing parameters of InAsP and GaAsP are much smaller than those of InGaP and InGaAs [172,175-176]. Under these circumstances where $c_{ACD} = c_{BCD} = 0$, (5.5) can be further simplified to:
The interpolation scheme of (5.7) will be employed when modelling the low-field, low-doping mobility in InGaAsP. In the following sections, the choice of interpolation formula will largely depend on factors such as required accuracy, the physical nature of the parameter, and available experimental data. For example, the bandgap energy is the most critical parameter in device modelling and a slight variation in this parameter can significantly affect the terminal characteristics of the semiconductor device. Therefore, (5.5) has been used to interpolate the bandgap energy in InGaAsP, and \( m_Q \) has been additionally used as a fitting parameter. On the other hand, if the material parameter can be given by a specific expression owing to some physical basis, it is natural to consider that interpolation scheme may also obey the same expression. The dielectric constant \( \varepsilon \) is one such case that follows the Clausius-Mosotti relation [172]. Then, the interpolation scheme has a form given by:

\[
\varepsilon(x,y) = \frac{(1-x)y \cdot \varepsilon(AC) + (1-x)(1-y) \cdot \varepsilon(AD) + xy \cdot \varepsilon(BC) + x(1-y) \cdot \varepsilon(BD)}{(1-x)y \cdot \varepsilon(AC) + (1-x)(1-y) \cdot \varepsilon(AD) + xy \cdot \varepsilon(BC) + x(1-y) \cdot \varepsilon(BD)}
\]

(5.8)

### 5.2. Lattice Parameters and Dielectric Constant

#### 5.2.1. Lattice Constant

The length of each side of a cubic unit cell in crystalline structures is called lattice constant, \( a \). As to the values of lattice constant in III-V binary semiconductors, there are little or no discrepancy in the data reported in many text/data books (e.g., [172,177-179]) and published articles (e.g., [173-174]). The values used in this work for the five binary materials of interest, namely AlAs, GaAs, InAs, InP, and GaP, are listed in Table 5.1 along with some other parameters for these materials which are discussed in the forthcoming sections.

The lattice constant of the III-V ternary and quaternary materials is known to obey Vegard’s law [172,180]. Using (5.2) it can be shown that \( \text{In}_{0.52}\text{Ga}_{0.48}\text{As} \) and \( \text{In}_{0.52}\text{Al}_{0.48}\text{As} \) are lattice-matched to InP and \( \text{In}_{0.48}\text{Ga}_{0.51}\text{P} \) is lattice matched to GaAs. Also using (5.3) for \( \text{In}_{1-x}\text{Ga}_{x}\text{As}_{y}\text{P}_{1-y} \):

\[
a_{\text{InGaAsP}}(x,y) = (1-x)y \cdot a_{\text{InAs}} + (1-x)(1-y) \cdot a_{\text{InP}} + xy \cdot a_{\text{GaAs}} + x(1-y) \cdot a_{\text{GaP}}
\]

(5.9)

the conditions of lattice-matching to InP and GaAs will be derived as:

\[
x = \frac{0.1896y}{0.4175 - 0.0125y} \quad 0 \leq y \leq 1 \quad \text{for lattice-matching to InP}
\]

(5.10)

\[
x = \frac{0.2154 + 0.1896y}{0.4175 - 0.0125y} \quad 0 \leq y \leq 1 \quad \text{for lattice-matching to GaAs}
\]

(5.11)
5.2.2. Dielectric Constant

The reported data of low-frequency (or static) and high-frequency (or optical) relative dielectric constants, $\varepsilon_1$ and $\varepsilon_h$, show a slightly more discrepancy than those of lattice constant [172,174,177-179,181-184]. The values of these two parameters used in this work for the binary compounds are also given in Table 5.1. As to the choice among the reported values, the most consistent or most recent published data have been adopted.

Table 5.1- Material parameters of the III-V binary compounds of interest.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Unit</th>
<th>AlAs</th>
<th>GaAs</th>
<th>InAs</th>
<th>InP</th>
<th>GaP</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>Å</td>
<td>5.6611</td>
<td>5.6533</td>
<td>6.0583</td>
<td>5.8687</td>
<td>5.4512</td>
</tr>
<tr>
<td>$\varepsilon_1$</td>
<td>$\varepsilon_0$</td>
<td>10.06</td>
<td>12.90</td>
<td>15.15</td>
<td>12.61</td>
<td>11.10</td>
</tr>
<tr>
<td>$\varepsilon_h$</td>
<td>$\varepsilon_0$</td>
<td>8.16</td>
<td>10.52</td>
<td>12.25</td>
<td>9.61</td>
<td>9.08</td>
</tr>
<tr>
<td>$E_{\text{el}}(0)$</td>
<td>eV</td>
<td>2.924</td>
<td>1.519</td>
<td>0.415</td>
<td>1.421</td>
<td>2.88</td>
</tr>
<tr>
<td>$E_{\text{ex}}(0)$</td>
<td>eV</td>
<td>2.239</td>
<td>1.981</td>
<td>2.278</td>
<td>2.32</td>
<td>2.338</td>
</tr>
<tr>
<td>$E_{\text{et}}(0)$</td>
<td>eV</td>
<td>2.590</td>
<td>1.815</td>
<td>1.529</td>
<td>2.07</td>
<td>2.74</td>
</tr>
<tr>
<td>$\alpha_F$</td>
<td>$10^{-4}$ eV/K</td>
<td>8.8</td>
<td>5.405</td>
<td>2.76</td>
<td>4.906</td>
<td>8.0</td>
</tr>
<tr>
<td>$\alpha_X$</td>
<td>$10^{-4}$ eV/K</td>
<td>6.0</td>
<td>4.6</td>
<td>5.87</td>
<td>7.66</td>
<td>5.771</td>
</tr>
<tr>
<td>$\alpha_{\text{lo}}$</td>
<td>$10^{-4}$ eV/K</td>
<td>8.5</td>
<td>6.05</td>
<td>3.36</td>
<td>4.88</td>
<td>8.21</td>
</tr>
<tr>
<td>$\beta_F$</td>
<td>K</td>
<td>334</td>
<td>204</td>
<td>83</td>
<td>327</td>
<td>300</td>
</tr>
<tr>
<td>$\beta_X$</td>
<td>K</td>
<td>408</td>
<td>204</td>
<td>83</td>
<td>327</td>
<td>372</td>
</tr>
<tr>
<td>$\beta_{\text{lo}}$</td>
<td>K</td>
<td>330</td>
<td>204</td>
<td>83</td>
<td>327</td>
<td>372</td>
</tr>
<tr>
<td>$\chi$</td>
<td>eV</td>
<td>3.839</td>
<td>4.07</td>
<td>4.970</td>
<td>4.315</td>
<td>3.785</td>
</tr>
<tr>
<td>$m_{nF}$</td>
<td>$m_0$</td>
<td>0.15</td>
<td>0.065</td>
<td>0.023</td>
<td>0.079</td>
<td>0.126</td>
</tr>
<tr>
<td>$m_{nX}$</td>
<td>$m_0$</td>
<td>0.71</td>
<td>0.85</td>
<td>0.64</td>
<td>0.676</td>
<td>0.82</td>
</tr>
<tr>
<td>$m_{nL}$</td>
<td>$m_0$</td>
<td>0.66</td>
<td>0.56</td>
<td>0.29</td>
<td>0.655</td>
<td>0.756</td>
</tr>
<tr>
<td>$m_{h}$</td>
<td>$m_0$</td>
<td>0.16</td>
<td>0.080</td>
<td>0.026</td>
<td>0.120</td>
<td>0.16</td>
</tr>
<tr>
<td>$m_{hh}$</td>
<td>$m_0$</td>
<td>0.76</td>
<td>0.48</td>
<td>0.45</td>
<td>0.56</td>
<td>0.54</td>
</tr>
</tbody>
</table>

The dielectric constants have been reported to slightly vary with temperature [185-186]. However, since the change in $\varepsilon_1$ and $\varepsilon_h$ with temperature is very small and also these parameters are not as critical in the device terminal characteristics as others like energy bandgap and mobility, the temperature variation of them has been ignored in this work.

The dielectric constants in the ternary and quaternary compounds may be calculated from (5.8). Using this interpolation scheme, $\varepsilon_1$ and $\varepsilon_h$ in In$_{0.49}$Ga$_{0.51}$P are found to be 11.81 and 9.34, in excellent agreement to the values of 11.75 and 9.34, respectively, reported in [187]. Similarly, a good agreement can be observed between values of $\varepsilon_1$ and $\varepsilon_h$ in In$_{0.53}$Ga$_{0.47}$As using the binary data in Table 5.1 (14.01 and 11.59, respectively) and those given in [172] (13.94 and 11.61, respectively).
5.3. Band Parameters

5.3.1. Energy Bandgaps

The forbidden bandgap energy without doubt is the most important parameter of a semiconductor which has significant effects on almost all the physical phenomena and the terminal characteristics in a semiconductor device. Therefore, in this work particular attention is paid to the accurate determination of this parameter for various compounds. Since the III-V compounds of interest may change from a direct bandgap to indirect bandgap (and vice versa) as the composition of materials (x and y) varies, one has to consider a multi-valley conduction band model to thoroughly describe the behaviour of band parameters.

We define three different bandgaps corresponding to the energy difference between the valence band edge \( E_V \) and the three conduction band minima, namely: \( \Gamma \)-band, \( L \)-band, and \( X \)-band. At low doping levels, where the conduction and valence bands are not disturbed by the presence of a large amount of impurities, and in the absence of strain, each of the above three bandgaps in binary compounds are considered to be uniquely determined by the lattice temperature. Various models for the temperature dependence of bandgap are proposed in the literature \[188\]. In this work, we employ the three-parameter Varshni’s equation \[189\] which is widely used in the semiconductor literature. Other types of temperature dependence formulations will be referenced in the study of particular materials wherever necessary.

According to Varshni’s equation, the direct and indirect bandgaps of binary compounds can be written as:

\[
E_{GV}(T) = E_{GV}(0) - \frac{\alpha_v \cdot T^2}{T + \beta_v}
\]

(5.12)

where \( v \) stands for \( \Gamma \), \( X \), or \( L \), \( E_{GV}(0) \) is the corresponding bandgap at \( T = 0K \), and \( \alpha_v \), \( \beta_v \) are some fitting parameters. The values of the above parameters used in this work for the five binary compounds of interest are summarised in Table 5.1, and they are individually studied in the following subsections.

For the ternary compound \( A_xB_{1-x}C \), a parabolic variation of bandgap with \( x \) is assumed:

\[
E_{GV}(x, T) = x \cdot E_{GV}(AC, T) + (1-x) \cdot E_{GV}(BC, T) - c_v(ABC) \cdot x(1-x)
\]

(5.13)

where the bandgap bowing parameter, \( c_v \), is assumed to be temperature independent \[190\] (also supported by the data in \[176\]). For the quaternary material \( In_{1-x}Ga_xAs_yP_{1-y} \), an interpolation scheme based on (5.5) with an appropriate choice of \( m_Q \) will be used.

5.3.1.a. GaAs

GaAs is the most widely studied of the III-V semiconductors. The reported values of GaAs direct bandgap shows very slight discrepancy varying in the range 1.420-1.432eV at room
temperature, and 1.508-1.532eV at temperatures close to 0K. (For compilations see [176,191-192].) Table 5.2 summarises the reported values of \(E_{g\nu}(0)\), \(\alpha_{\nu}\), and \(\beta_{\nu}\) for GaAs. The parameters of Aspens [193] and Thurmond [194] are the most widely used parameter set in the literature, and are also adopted in the present work. The results of Grilli, et al. [195], which have been obtained from a fit to photoluminescence data measured in the range 2K to 280K, largely underestimate the available experimental data above \(T = 300K\) [196]. The values of \(\alpha_{T}\) and \(\beta_{T}\) given in [197-199] are basically minor adjustments of those given by Aspens and Thurmond.

Table 5.2- The Varshni parameters of Eq. (5.12) for direct and indirect bandgaps in GaAs.

<table>
<thead>
<tr>
<th>Minima</th>
<th>(E_d(0)) (eV)</th>
<th>(\alpha \times 10^{-4}) eV/K</th>
<th>(\beta) (K)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\Gamma)</td>
<td>1.519</td>
<td>5.405</td>
<td>204</td>
<td>Aspens &amp; Thurmond [193-194]</td>
</tr>
<tr>
<td>(\Gamma)</td>
<td>1.517</td>
<td>5.5</td>
<td>225</td>
<td>Lautenschlager, et al. [197]</td>
</tr>
<tr>
<td>(\Gamma)</td>
<td>1.515</td>
<td>5.1</td>
<td>190</td>
<td>Shen, et al. [198]</td>
</tr>
<tr>
<td>(\Gamma)</td>
<td>1.5194</td>
<td>10.6</td>
<td>671</td>
<td>Grilli, et al. [195]</td>
</tr>
<tr>
<td>(\Gamma)</td>
<td>1.532</td>
<td>5.2</td>
<td>239</td>
<td>Chan, et al. [199]</td>
</tr>
<tr>
<td>(X)</td>
<td>1.981</td>
<td>5.2</td>
<td>204</td>
<td>Aspens [193]</td>
</tr>
<tr>
<td>(X)</td>
<td>1.975</td>
<td>6.2</td>
<td>550</td>
<td>Lee, et al. [200]</td>
</tr>
<tr>
<td>(L)</td>
<td>1.815</td>
<td>6.05</td>
<td>204</td>
<td>Aspens [193]</td>
</tr>
<tr>
<td>(L)</td>
<td>1.820</td>
<td>7.8</td>
<td>490</td>
<td>Lee, et al. [200]</td>
</tr>
<tr>
<td>(L)</td>
<td>—</td>
<td>7.2</td>
<td>205</td>
<td>Lautenschlager, et al. [197]</td>
</tr>
</tbody>
</table>

A slightly more accurate, 4-parameter formulation of GaAs direct bandgap versus temperature has recently been suggested by Passler and Oelgart [201]. The formula is given as:

\[
E_g(T) = E_g(0) - \frac{\alpha \Theta}{2} \left[ \frac{1}{1 + \left( \frac{2T}{\Theta} \right)^p} - 1 \right] 
\]

where the parameters \(\alpha = 0.475\) meV/K, \(\Theta = 222.4K\), \(p = 2.667\), \(E_d(0) = 1.5192eV\) gives an excellent fit to the data measured in the range 4K to 340K. The variation of GaAs direct bandgap with temperature modelled using equations (5.12) and (5.14) with the parameter sets of Aspens and Thurmond, Shen et al., and Passler and Oelgart are compared in Fig. 5.1 with the experimental data by Passler and Oelgart (for 4K ≤ T ≤ 340K), and Panish and Casey [196] (for T ≥ 473K). As can be seen in Fig. 5.1, the Varshni’s equation may produce errors of ~5meV in the temperature range (20K-150K). However, this error is considered negligible, and the Varshni’s equation is preferred in the present work to other more complicated formulations of bandgap versus temperature [188,201-202]. A semi-empirical Bose-Einstein relation has also been proposed [197] for the variation of bandgap with temperature in semiconductor materials:

\[
E_g(T) = E_g(0) - \frac{2a_B}{\exp(\Theta/T) - 1} 
\]

where \(a_B\) is an electron-phonon coupling constant and \(\Theta\) is a temperature corresponding to the
average phonon frequency. As far as fitting to the available experimental data is concerned, the situation for Eq. (5.15) is similar to that of (5.14).

Compared to the data available for direct bandgap in GaAs, those for indirect bandgaps, X and L, are scarce and show much more discrepancy. The reported values of $E_{gX}$ vary in the range 1.90-1.92eV at room temperature, and 1.971-2.18eV at temperatures close to 0K. As to $E_{gL}$, the values in the range 1.705-1.75eV at $T = 300K$, and 1.798-1.85eV at $T = 0K$ are given in the literature (see, for instance, [176,180,182,193,203]). The Varshni parameters reported for X and L energy gaps of GaAs are also listed in Table 5.2. Again, in the present work the widely used values given by Aspens are adopted.

5.3.1.b. AlAs and Al$_x$Ga$_{1-x}$As

AlAs is an indirect bandgap semiconductor in which the minimum of the conduction band lies in the X-valley. Compared to material properties of GaAs, those of AlAs are much less studied, mainly due to the fact that the poor transport properties of AlAs makes it much less technologically important.

The room temperature bandgap of AlAs ($E_{gX}$) reported in the literature varies in the range (2.14-2.168)eV, while those measured or calculated near 0K almost consistently show values in the range (2.23-2.25)eV (see for example [92,176,179,181,204]). However, the only Varshni parameters reported in the literature for $E_{gX}$(AlAs) are those of Casey and Panish [181] which are adopted in this work and are listed along with the parameters fitted to other bandgaps of AlAs and Al$_x$Ga$_{1-x}$As in Table 5.3.
Table 5.3 - The Varshni parameters for direct and indirect bandgap of Al_{x}Ga_{1-x}As given for different values of aluminium mole fraction.

<table>
<thead>
<tr>
<th>x</th>
<th>Minima</th>
<th>E_{g}(0) (eV)</th>
<th>$\alpha \times 10^4$ eV/K</th>
<th>$\beta$ (K)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>X</td>
<td>2.239</td>
<td>6.0</td>
<td>408</td>
<td>Casey &amp; Panish [181]</td>
</tr>
<tr>
<td>1.0</td>
<td>$\Gamma$</td>
<td>3.135</td>
<td>8.41</td>
<td>408</td>
<td>Jones et al. [205]</td>
</tr>
<tr>
<td>1.0</td>
<td>$\Gamma$</td>
<td>2.891</td>
<td>8.78</td>
<td>322</td>
<td>Simlinger et al. [206]</td>
</tr>
<tr>
<td>0.69</td>
<td>$\Gamma$</td>
<td>2.485</td>
<td>7.88</td>
<td>302</td>
<td>Logothetidis et al. [207]</td>
</tr>
<tr>
<td>0.53</td>
<td>$\Gamma$</td>
<td>2.251</td>
<td>7.04</td>
<td>261</td>
<td>Logothetidis et al. [207]</td>
</tr>
<tr>
<td>0.27</td>
<td>$\Gamma$</td>
<td>1.932</td>
<td>6.58</td>
<td>248</td>
<td>Logothetidis et al. [207]</td>
</tr>
<tr>
<td>0.18</td>
<td>$\Gamma$</td>
<td>1.771</td>
<td>6.3</td>
<td>236</td>
<td>Shen et al. [198]</td>
</tr>
<tr>
<td>0.0</td>
<td>$\Gamma$</td>
<td>1.519</td>
<td>5.405</td>
<td>204</td>
<td>Aspens &amp; Thurmond [193-194]</td>
</tr>
<tr>
<td>0.38</td>
<td>$\Gamma$</td>
<td>2.053</td>
<td>7.82</td>
<td>383</td>
<td>El Allali et al. [208]</td>
</tr>
<tr>
<td>0.27</td>
<td>$\Gamma$</td>
<td>1.895</td>
<td>9.00</td>
<td>507</td>
<td>El Allali et al. [208]</td>
</tr>
<tr>
<td>0.18</td>
<td>$\Gamma$</td>
<td>1.773</td>
<td>8.06</td>
<td>451</td>
<td>El Allali et al. [208]</td>
</tr>
<tr>
<td>0.10</td>
<td>$\Gamma$</td>
<td>1.658</td>
<td>6.44</td>
<td>304</td>
<td>El Allali et al. [208]</td>
</tr>
<tr>
<td>0.0</td>
<td>$\Gamma$</td>
<td>1.519</td>
<td>8.95</td>
<td>538</td>
<td>El Allali et al. [208]</td>
</tr>
<tr>
<td>0.69</td>
<td>L</td>
<td>—</td>
<td>7.8</td>
<td>270</td>
<td>Logothetidis et al. [207]</td>
</tr>
<tr>
<td>0.53</td>
<td>L</td>
<td>—</td>
<td>7.55</td>
<td>241</td>
<td>Logothetidis et al. [207]</td>
</tr>
<tr>
<td>0.27</td>
<td>L</td>
<td>—</td>
<td>6.62</td>
<td>171</td>
<td>Logothetidis et al. [207]</td>
</tr>
</tbody>
</table>

The X-bandgap bowing parameter for Al_{x}Ga_{1-x}As is reported in the range (0.055-0.245)eV [176,204]. Since the Varshni parameters for AlAs are chosen from [181], we use the bandgap bowing parameter given in the same reference as $c_x$ (AlGaAs) = 0.143eV.

As to the direct bandgap of AlAs, there is a large discrepancy between the reported data. For example, if one considers the Varshni parameters for AlAs given in [205] and [206], the calculated bandgaps given by the two equations are almost 0.25eV away from each other for the entire range of temperature (see Table 5.3 and Fig. 5.2). Therefore, in this work none of the above two extremes are used. Instead, the experimental data measured for Al_{x}Ga_{1-x}As as a function of temperature with different values of x in the range (0-0.69) are analysed to extrapolate the temperature variation of AlAs direct energy gap. The Varshni parameters fitted to experimental data given in [194,198,207] for x = 0.0, 0.18, 0.27, 0.53, 0.69, and in [208] for x = 0.0, 0.0, 0.10, 0.18, 0.27, 0.38 are summarised in Table 5.3. $E_{g \Gamma}(0)$ can best be fitted by a linear equation of the form:

$$E_{g \Gamma}(Al_{x}Ga_{1-x}As, 0K) = 1.519(1-x) + 2.924x$$  \hspace{1cm} (5.16)

In other words, the bowing parameter of the direct bandgap of AlGaAs will be zero, in consistence with theoretical calculations which all give small or non-existing $c_{\Gamma}(AlGaAs)$ (see [209] and references therein). Similar extrapolation of data for $\alpha_{\Gamma}$ and $\beta_{\Gamma}$ results in (The $\alpha_{\Gamma}$ and $\beta_{\Gamma}$ of El Allali et al. [208] are not used for this extrapolation.):

$$\alpha_{\Gamma}(AlAs) = 8.8 \times 10^{-4} \text{ eV/K} \hspace{1cm} \beta_{\Gamma}(AlAs) = 334 \text{ K}$$

which are fairly close to similar parameters given by Simlinger et al. [206]. Figure 5.2
Chapter 5 Physical Parameters in III-V Semiconductors

3.6 — 3.2 —

Calc. for AlAs: Chen & Sher [176]

Calc. for AlAs; Chen & Sher [210]

Jones et al. [205]

Simlinger et al. [206]

Logothetidis et al. [207]

El Allali et al. [208]

Shen et al. [198]

fitting used here

\[ X = 1.0 \]

\[ X = 0.69 \]

\[ X = 0.53 \]

\[ X = 0.38 \]

\[ X = 0.27 \]

\[ X = 0.10 \]

200 400 600 800

Temperature (K)

800

F igu re 5.2- Direct energy bandgap of Al\(_x\)Ga\(_{1-x}\)As versus temperature with aluminium mole fraction as a parameter. The calculated values in this work are compared against the Varshni equations fitted to measured data by Shen et al. [198] for \( x = 0.18 \), by Logothetidis et al. [207] for \( x = 0.27, 0.53, 0.69 \), and by El Allali et al. [208] for \( x = 0.10, 0.18, 0.27, 0.38 \). Also shown are the Varshni equations for AlAs given in [205] and [206], and the calculated values of \( E_{\text{gr}}(\text{AlAs}, 0\text{K}) \) in [176,210].

compares the direct bandgap of Al\(_x\)Ga\(_{1-x}\)As versus temperature calculated using equations (5.12) and (5.13) and the data obtained in this work (summarised in Table 5.1) with those reported in the literature. Very good agreement can be observed between the results of this work and the Varshni equations fitted to experimental data by Shen et al. [198] for \( x = 0.18 \), Logothetidis et al. [207] for \( x = 0.27, 0.53, 0.69 \) and El Allali et al. [208] for \( x = 0.10, 0.18, 0.27, 0.38 \). The calculated values of \( E_{\text{gr}}(\text{AlAs}) \) at \( T = 0\text{K} \) in [176,210] are also included in Fig. 5.2, and are in excellent agreement to the value predicted in this thesis. However, it has to be mentioned that some recent experimental data [211] suggest a significant upward deviation of \( E_{\text{gr}}(\text{Al}_x\text{Ga}_{1-x}\text{As}) \) for \( x > 0.8 \). The third order polynomial fitted to these data shows a notable disagreement with the fitting in the present thesis for \( x > 0.8 \) (see Fig. 5.3).

The L-bandgap in Al\(_x\)Ga\(_{1-x}\)As is the least important and thus the least studied among the three bandgaps. It contributes to the transport and other properties of AlGaAs mostly near the direct-indirect cross-over composition (\( x \sim 0.4 \)), and to some extent for \( 0 < x < 0.3 \) due to the importance of satellite minima separation in some physical phenomena such as velocity overshoot. Consequently, the data for this conduction band minima are scarce and, where existing, show a large discrepancy. The only near 0K L-bandgap data of AlAs available in the literature are those calculated in [176] and [210] and vary in the range (2.48-2.59)eV. In this thesis the value of 2.59eV is adopted which results in a direct-indirect cross-over behaviour at room temperature similar to those obtained by other authors [176] (see Fig. 5.3).
No temperature variation of L-bandgap in AlGaAs has been reported in the literature. The Varshni parameters reported in Table 5.3 from [207] are measured for E_l optical transition. In this work, it has been assumed that the variation of L-bandgap with temperature is similar to that of E_l optical energy, as suggested in [92]. If one follows the trend of $\alpha_L$ and $\beta_L$ variation with $x$, the Varshni parameters in AlAs can be extrapolated as:

\[ \alpha_L(\text{AlAs}) = 8.5 \times 10^{-4} \text{eV/K} \quad \beta_L(\text{AlAs}) = 330\text{K} \]

The L-bandgap bowing factor for AlGaAs is generally accepted to be very small in the range (0.0-0.2)eV [176]. A value of $c_L(\text{AlGaAs}) = 0.15\text{eV}$ [210] is chosen in this work.

All the Varshni parameters for GaAs and AlAs used in this thesis are summarised in Table 5.1. Using these values together with the bandgap bowing parameters for AlGaAs, as given above, the variation of the three conduction band minima in AlGaAs with aluminium mole fraction at room temperature is calculated and shown in Fig. 5.3. A direct-indirect bandgap transition composition of 0.40 can be observed in this figure.

### 5.3.1.c. InAs and In$_{1-x}$Ga$_x$As

The ternary compound In$_{1-x}$Ga$_x$As is technologically very important for microwave and optoelectronic applications. This material is direct bandgap for the entire range of composition. In$_{0.53}$Ga$_{0.47}$As lattice-matched to InP has been widely investigated in InP-based HBTs and has resulted in excellent DC and high frequency performance [212]. Strained layers of In$_{1-x}$Ga$_x$As have also been used to improve the emitter ohmic contact specific resistance of GaAs-based HBTs [213]. In addition, graded layers of strained In$_{1-x}$Ga$_x$As with small In mole fractions have been used to grade the base bandgap of Al-free InGaP/GaAs HBTs and resulted in a significant reduction of base transit time [93].

Due to the importance of In$_{1-x}$Ga$_x$As, the direct energy bandgap of this material has been the subject of many experimental and theoretical investigations. Table 5.4 gives a summary of reported Varshni parameters for this material system. The bandgap of InAs is consistently reported in the literature to be in the range (0.41-0.43)eV for temperatures close to 0K, and in the range (0.35-0.36)eV near room temperature [182,192]. In this thesis the more recent Varshni parameters of Fang...
Figure 5.4- Temperature variation of In$_{1-x}$Ga$_x$As direct bandgap using equation (5.13) with $\alpha$(InGaAs) = 0.42eV and Varshni parameters for InAs and GaAs summarised in Table 5.1 (solid lines). This is compared against other formulations of direct bandgap versus temperature reported in the literature for $x = 0$ [181,189-190]; $x = 0.47$ [182,217-218]; $x = 0.85$ and $x = 0.94$ [215,216] (symbols). See also Table 5.4.

The direct bandgap bowing parameter of InGaAs is reported to be in the range (0.28-0.56)eV [176,182]. In this work a value of $\alpha$(InGaAs) is chosen to give room temperature and 0K bandgaps of the widely studied In$_{0.53}$Ga$_{0.47}$As consistent with the reported ranges of (0.75-0.76)eV and (0.81-0.837)eV [176,182], respectively. In that sense, $\alpha$(InGaAs) = 0.42eV is found to be a good choice, giving rise to $E_{gT}$(InGaAs,0K) = 0.829eV and $E_{gT}$(InGaAs,300K) = 0.750eV. According to the Varshni parameters used for InAs and GaAs in this work and Eq. (5.13) with $\alpha$(InGaAs) = 0.42eV, the direct bandgap of In$_{1-x}$Ga$_x$As is calculated for $x = 0$, 0.47, 0.85, 0.94 and plotted versus temperature (solid lines).
in Fig. 5.4. Other temperature variations of bandgap summarised in Table 5.4 for the above Ga compositions are also shown in Fig. 5.4 (symbols). It can be observed that Eq. (5.13) gives excellent fits to all the data reported in the literature.

The indirect bandgaps in InAs are much larger than the direct one and their importance for the transport properties of InAs is insignificant. Consequently, they are much less studied than the Γ-bandgap. The only data available for $E_{gX}(\text{InAs}, 0K)$ and $E_{gL}(\text{InAs}, 0K)$ are those calculated by Chen and Sher [176], which are 2.278eV and 1.529eV, respectively. However, a large scatter exists among the room temperature values of indirect bandgap for InAs in the literature [178,182,183,203]:

$$E_{gX}(\text{InAs,300K}) = (1.37 - 2.26)eV$$
$$E_{gL}(\text{InAs,300K}) = (1.06 - 1.76)eV$$

Bearing in mind that the indirect valleys in InAs are located at least 0.72eV above the Γ-valley, the actual magnitude of their bandgap becomes less significant. Therefore, we adopt the following room temperature values from [178] which are rather consistent with the 0K data given above, and in excellent agreement with the recently measured L-Γ and X-Γ separations in [203]:

$$E_{gX}(\text{InAs,300K}) = 2.14eV$$
$$E_{gL}(\text{InAs,300K}) = 1.45eV$$

If one further assumes $\beta_X(\text{InAs}) = \beta_L(\text{InAs}) = \beta_\Gamma(\text{InAs}) = 83K$, it follows:

$$E_{gX}(\text{InAs},T) = 2.278 - \frac{5.87 \times 10^{-4} T^2}{T + 83}$$  \hspace{1cm} (5.17)$$
$$E_{gL}(\text{InAs},T) = 1.529 - \frac{3.36 \times 10^{-4} T^2}{T + 83}$$  \hspace{1cm} (5.18)$$

Finally, the indirect bandgap bowing parameters may be found by fitting the Γ-X and Γ-L valley separations of $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ to the following widely accepted values [182]:

$$E_{\Gamma-X}(\text{In}_{0.53}\text{Ga}_{0.47}\text{As}) = 1.15eV$$
$$E_{\Gamma-L}(\text{In}_{0.53}\text{Ga}_{0.47}\text{As}) = 0.55eV$$

This results in $c_X(\text{InGaAs}) = 0.509eV$ and $c_L(\text{InGaAs}) = 1.087eV$. The calculated composition dependence of the room temperature bandgaps of $\text{In}_{1-x}\text{Ga}_x\text{As}$ using the parameters given in this section is shown in Fig. 5.5.

5.3.1.d. $\text{In}_{1-x}\text{Al}_x\text{As}$

With the temperature variation of the three bandgaps in InAs and AlAs already known, the problem of formulating the bandgap variation in $\text{In}_{1-x}\text{Al}_x\text{As}$ reduces to finding only three suitable bandgap bowing parameters. The data available for the technologically important $\text{In}_{0.52}\text{Al}_{0.48}\text{As}$ lattice-matched to InP will be used here to find appropriate bowing parameters.

Widely accepted $E_{g\Gamma}(\text{In}_{0.52}\text{Al}_{0.48}\text{As}) = 1.450eV$ measured by Wakefield et al. [219] at room temperature is used here to define $c_\Gamma(\text{InAlAs}) = 0.304eV$. This value is smaller than
Chapter 5  Physical Parameters in III-V Semiconductors

c_{\Gamma}(\text{InAlAs}) = 0.740\text{eV} suggested by Wakefield et al., since the direct bandgap of AlAs assumed in their work is larger than what is used here. Also using $E_{g\chi}(\text{In}_{0.52}\text{Al}_{0.48}\text{As}) = 2.081\text{eV}$ and $E_{gL}(\text{In}_{0.52}\text{Al}_{0.48}\text{As}) = 1.884\text{eV}$, both calculated at $T = 0\text{K}$ [176], one can obtain:

c_{\chi}(\text{InAlAs}) = 0.713\text{eV} \quad c_{L}(\text{InAlAs}) = 0.617\text{eV}

Figure 5.6 shows the calculated room temperature variation of $\Gamma$-, $X$-, and $L$-bandgaps in In$_{1-x}$Al$_x$As versus the Al mole fraction. A direct-to-indirect bandgap transition is observed at around $x = 0.70$.

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{figures}
\caption{Variation of the $\Gamma$-, $X$-, and $L$-bandgaps of In$_{1-x}$Ga$_x$As with gallium mole fraction at room temperature.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{figures}
\caption{Variation of the $\Gamma$-, $X$-, and $L$-bandgaps of In$_{1-x}$Al$_x$As with aluminium mole fraction at room temperature.}
\end{figure}

5.3.1.e. InP

The available data for direct bandgap and other optical transitions in InP are summarised in [220]. Table 5.5 summarises the available reports on the temperature dependence of the direct energy bandgap in InP based on the Varshni equation (Eq. (5.12)) and Bose-Einstein relation (Eq. (5.15)). These functional dependencies together with some other data summarised in [220] are plotted in Fig. 5.7. It can be seen that the formulations given by Varshni [189], Hang et al. [221], and Temkin et al. [222] are almost identical in the temperature range (100-700)K, and also fit the average of other data available in the literature (shown by symbols). However, the Bose-Einstein relation proposed in [220] overestimates the bandgap of InP in the temperature range (100-400)K. In this work the Varshni parameters originally proposed by Varshni are used, since the value of $E_g(0\text{K})$ used by Hang et al. is slightly larger than the range (1.41-1.42)\text{eV} most frequently reported in the literature.

As usual, the indirect bandgap of InP has received comparably less attention. The ranges of room temperature and 0K indirect bandgaps of InP reported in the literature are as follows [172,176,180,183,186,210]:

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{figures}
\caption{Variation of the indirect bandgaps of In$_{1-x}$Ga$_x$As and In$_{1-x}$Al$_x$As with aluminium mole fraction at room temperature.}
\end{figure}
Table 5.5 - Parameters for the temperature dependence of the direct bandgap in InP based on the Varshni equation (5.12) and Bose-Einstein relation (5.15) reported by various authors.

<table>
<thead>
<tr>
<th>$E_g(0)$ (K)</th>
<th>$\alpha_T$ ($10^{-4}$ eV/K)</th>
<th>$\beta_T$ (K)</th>
<th>$E_{gT}(0)$ (K)</th>
<th>$\Theta_T$ (K)</th>
<th>Temperature Range (K)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.432</td>
<td>4.1</td>
<td>136</td>
<td>1.423</td>
<td>0.051</td>
<td>(77-873)</td>
<td>Hang et al. [221]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.412</td>
<td>0.217</td>
<td>697</td>
<td>(30-420)</td>
<td>Lautenschlager et al. [220]</td>
</tr>
<tr>
<td>1.421</td>
<td>3.63</td>
<td>162</td>
<td>(0-300)</td>
<td></td>
<td>Varshni [189]</td>
<td></td>
</tr>
<tr>
<td>1.4206</td>
<td>4.906</td>
<td>327</td>
<td>(0-300)</td>
<td></td>
<td>Temkin et al. [222]</td>
<td></td>
</tr>
<tr>
<td>1.4206</td>
<td>4.9</td>
<td>327</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$E_{gX}$(InP,0K) = (2.08-2.44)eV with values most consistently reported in the range (2.32-2.38)eV; 2.32 eV used here

$E_{gX}$(InP,300K) = (2.01-2.30)eV; 2.21 eV used here

$E_{gL}$(InP,0K) = (2.03-2.28)eV; 2.07 eV used here

$E_{gL}$(InP,300K) = (1.95-2.21)eV; 2.00 eV used here

In this work, the 0K values calculated by Chen and Sher [210] are chosen together with room temperature data that are rather consistent with them. Assuming:

$\beta_T$(InP) = $\beta_X$(InP) = $\beta_L$(InP) = 327K

one can obtain appropriate $\alpha$’s as follows:

$\alpha_X$(InP) = 7.66 x 10^{-4} eV / K

$\alpha_L$(InP) = 4.88 x 10^{-4} eV / K

5.3.1.f. GaP and In$_{1-x}$Ga$_x$P

GaP is an indirect bandgap material with the minimum of the conduction band lying in the X-valley. However, we start the discussion of this subsection by formulating the direct energy gap of GaP and In$_{1-x}$Ga$_x$P due to the technological importance of In$_{0.49}$Ga$_{0.51}$P lattice-matched to GaAs, which is a direct gap material.

In$_{0.49}$Ga$_{0.51}$P has a direct bandgap of ~1.89 eV at room temperature. The exact magnitude of the bandgap, however, can depend on whether the grown layer is ordered or disordered. The crystalline structure in an ordered In$_{0.49}$Ga$_{0.51}$P layer is such that sheets of pure Ga, P, In, and P atoms alternate on the (001) planes of the basic unit-cell [31], without intermixing of the Ga and In atoms on the same lattice.
plane. MOCVD grown In$_{0.49}$Ga$_{0.51}$P layers at relatively high growth temperatures of $> 680^\circ$C are usually ordered. By lowering the growth temperature to $550^\circ$C, however, we repeatedly obtain semi-ordered In$_{0.49}$Ga$_{0.51}$P layers. The In$_{0.49}$Ga$_{0.51}$P layer grown by MOMBE at even lower temperature of $450^\circ$C is most possibly disordered [31]. Nevertheless, the above is not a general rule; ordered In$_{0.49}$Ga$_{0.51}$P layers have also been obtained at $620^\circ$C, and increasing the growth temperature above $620^\circ$C has resulted in reduction of the level of orderliness [223]. Moreover, there are suggestions that a significant decrease in ordering (or increase in bandgap) can be caused by introducing a high concentration of dopants [224].

It is found that the ordered material has slightly less energy bandgap ($\sim 80$ meV) than the disordered material [223]. In addition, it is suspected that the degree of orderliness can influence the exact bandgap alignment of the In$_{0.49}$Ga$_{0.51}$P/GaAs heterojunction. Values of $\Delta E_c$ at this heterojunction are reported in the range (0.03-0.39)eV! (For a compilation of various results, see [31].) The large discrepancies suggest that there could be a fundamental structural difference in the In$_{0.49}$Ga$_{0.51}$P layers being studied. One possibility is to relate this $\Delta E_c$ difference to the degree of ordering of the In$_{0.49}$Ga$_{0.51}$P layer. The ordered In$_{0.49}$Ga$_{0.51}$P material usually gives rise to negligible $\Delta E_c$, whereas the semi-ordered material has a significantly larger $\Delta E_c$ [31]. It is also possible that an interfacial layer of a few monolayers thick with unknown composition exists in-between the In$_{0.49}$Ga$_{0.51}$P/GaAs heterojunction. The interfacial layer, likely InGaAsP [225], could affect the apparent band discontinuity measured by different techniques by acting as a grading layer between In$_{0.49}$Ga$_{0.51}$P and GaAs. The existence of such an interfacial layer is especially probable in a MOCVD growth due to the transient gas flow during gas switching [44]. More details on the band alignment of In$_{0.49}$Ga$_{0.51}$P/GaAs heterojunctions can be found in subsection 5.3.2 when discussing the electron affinity of InGaP.

As to the formulation of bandgap and electron affinity in In$_{1-x}$Ga$_x$P, disordered material is assumed throughout this section. However, the user of the simulation program is given the choice to manually adjust both the bandgap and band alignment (within reasonable range) according to the level of orderliness of the grown InGaP material to fit the experimental data.

The following information in the literature is used as the basis for finding appropriate Varshni coefficients for direct bandgap of GaP and direct bandgap bowing factor of InGaP:

1. The reported room temperature and 0K direct gap of GaP are in the ranges (2.73-2.85)eV and (2.866-2.895)eV, respectively [183-184,225-226].
2. The direct bandgap of disordered In$_{0.49}$Ga$_{0.51}$P is ranging from 1.883 to 1.918eV at 300K, and from 1.968 to 2.065eV at temperatures close to 0K [176,210,227-229].
3. $c_f$(InGaP) is given by various authors in the range (0.40-0.88)eV with majority of the values between 0.70 to 0.80eV [175-176,179,230].
4. The following formulae for the temperature dependence of direct bandgap in GaP is given in [184] and [231]:

68
\[ E_{gr}(GaP) = 2.866 - 0.108 \left[ \coth \left( \frac{164}{T} \right) - 1 \right] \quad \text{for} \quad 100K < T < 300K \quad (5.19a) \]

\[ E_{gr}(GaP) = 2.877 - \frac{9.6 \times 10^{-4} T^2}{T + 460} \quad \text{for} \quad 10K < T < 360K \quad (5.19b) \]

5. Ishitani et al. [223] have measured the temperature dependence of direct bandgap in \( In_{0.5}Ga_{0.5}P \) and \( In_{0.3}Ga_{0.7}P \) in the temperature range \( 10K \leq T \leq 300K \). They analysed the bandgap shrinkage with temperature in completely disordered form of these two materials and suggested the following functional dependence:

\[ E_{gr}(In_{0.5}Ga_{0.5}P) = 1.984 - 1.733 \times 10^{-5} T - 2.2 \times 10^{-7} T^2 - \frac{0.134}{\exp(525/T)-1} - \frac{0.0086}{\exp(55/T)-1} \quad (5.20) \]

\[ E_{gr}(In_{0.3}Ga_{0.7}P) = 2.292 - 6.57 \times 10^{-6} T - 2.67 \times 10^{-7} T^2 - \frac{0.155}{\exp(500/T)-1} - \frac{0.0079}{\exp(55/T)-1} \quad (5.21) \]

In the present work, it is shown that \( c_T(InGaP) = 0.67eV \) together with the following Varshni coefficients for direct bandgap of GaP perfectly suits all the above data:

\[ E_{gr}(GaP) = 2.880 - \frac{8.0 \times 10^{-4} T^2}{T + 300} \quad (5.22) \]

In Fig. 5.8 the temperature dependence of \( E_{gr}(In_{1-x}Ga_xP) \) (\( x = 0.5, 0.62, 0.7, 1.0 \)) obtained using the above bowing factor and Varshni coefficients for GaP and InP is compared with those formulated in [184,223,231-232]. Excellent agreement is found between the two sets of results. Therefore, it is proved again that the unified formulation of Eqs. (5.12) and (5.13) used in the present work is sufficient for temperature dependence of energy bandgap in III-V binary and ternary compounds.

The \( X \)-gap of GaP has been formulated as a function of temperature by Thurmond [194] and Panish and Casey [196]:

\[ E_{gx}(GaP) = 2.338 - \frac{5.771 \times 10^{-4} T^2}{T + 372} \quad [194] \quad (5.23) \]

\[ E_{gx}(GaP) = 2.338 - \frac{6.2 \times 10^{-4} T^2}{T + 460} \quad [196] \quad (5.24) \]

Both equations give very similar energy gaps and excellently represent the measured data tabulated in [196] from 0K up to 1273K. Therefore, the more recent result of Thurmond is adopted in this work. The bowing factor for the \( X \)-gap of \( In_{1-x}Ga_xP \) is reported in the range \( (0.0-0.22) eV \) [172,176,210,230,233]. Here \( c_X(InGaP) = 0.17eV \) [210] is adopted. \( E_{gx}(In_{0.5}Ga_{0.5}P) \) thus calculated would be 2.287eV at 0K and 2.188eV at 300K, in excellent agreement to the \( X \)-bandgaps mentioned in [210,228,234].

As to the \( L \)-bandgap, the calculated 0K data in [210] and the room temperature data in [234] are used. If one further assumes \( \beta_L(GaP) = \beta_X(GaP) = 372K \), then it follows:
§ •

2.2

x = 0.62

A

300 200 0.8 1.0

0.6 0.4

0.2 0.0

Temperature (K)

Figure 5.8- Temperature dependence of direct energy bandgap in In$_{1-x}$Ga$_x$P for x = 0.5, 0.62, 0.7, 1.0. The solid lines represent the data in this work, which are compared against the experimental curves in [223] for x = 0.5, 0.7, in [184] and [231] for GaP, and in [232] for x = 0.62.

Figure 5.9- Variation of the Γ-, X-, and L-bandgaps of In$_{1-x}$Ga$_x$P with gallium mole fraction at room temperature.

\[ E_{gL}(\text{GaP}) = 2.74 - \frac{8.21 \times 10^{-4} T^2}{T + 372} \quad (5.25) \]

and \( c_L(\text{InGaP}) = 0.41 \text{eV} \). Using the above mentioned data, the direct and indirect bandgap of In$_{1-x}$Ga$_x$P is calculated at room temperature in the entire range of composition and shown in Fig. 5.9. A direct-indirect crossover at x ~ 0.71 can be observed in Fig. 5.9.

5.3.1.g. In$_{1-x}$Ga$_x$As$_y$P$_{1-y}$

The special properties of the InGaAsP quaternary alloy make it a useful material for the fabrication of optoelectronic devices such as heterojunction lasers, light-emitting diodes, and avalanche photodiodes [175]. The room temperature bandgap of this material can be adjusted to vary in the range ~ 0.36eV (e.g., InAs) to 2.26eV (indirect bandgap GaP). This material can be grown lattice-matched to InP or GaAs substrates. (For conditions of lattice-matching to InP and GaAs, see Eqs. (5.10)-(5.11).) Recently, there has been an increasing interest in the use of InGaAsP lattice-matched to InP and GaAs substrates as a grading interlayer in InP/InGaAs [107] and InGaP/GaAs HBTs [60]. This necessitates the accurate formulation of bandgap and other transport related parameters of this quaternary material for any realistic device analysis.

In this work, the bandgap of In$_{1-x}$Ga$_x$As$_y$P$_{1-y}$ is formulated using (5.5) and (5.6). Since the temperature dependence of bandgap for the constituting binary compounds are already known and appropriate bandgap bowing parameters are chosen for InGaP and InGaAs, one just faces the suitable choice of bandgap bowing factors for InAsP and GaAsP ternaries for a complete formulation of energy bandgap in InGaAsP. In this process, \( 0 \leq m_Q \leq 1 \) may be used as a fitting parameter as suggested in section (5.1).
Chapter 5
Physical Parameters in III-V Semiconductors

1.4 -1

\[ T = 300\, \text{K} \]

InGaAsP Lattice-Matched to InP

\[ \text{Direct Energy Bandgap (eV)} \]

\[ \begin{align*}
\text{As. Mole Fraction, } y & \quad \text{Bandgap} (\text{eV}) \\
0.0 & \quad 1.8 \\
0.2 & \quad 1.4 \\
0.4 & \quad 1.0 \\
0.6 & \quad 0.6 \\
0.8 & \quad 0.2 \\
1.0 & \quad 0.0 \\
\end{align*} \]

\[ \text{Figure 5.10- Variation of bandgap in InGaAsP lattice-matched to InP with arsenic mole}
\]

fraction, \( y \). \( y = 0 \) and \( y = 1 \) correspond to InP and \( \text{In}_{0.53}\text{Ga}_{0.47}\text{As} \), respectively. The symbols represent the measured data in [236-237], and quadratic variations fitted to the experimental data by various authors, including: Laufer et al. [238], Kuphal and Pocker [239], and Bouhafs et al. [175]. Also shown in the figure are the bandgaps formulated using Eqs. (5.5)-(5.6) with \( m_0 \) as a parameter. \( m_0 = 0.4 \) results in a very good agreement to the experimental data.

The bandgap bowing parameters for GaAsP and InAsP are adopted from [172] and [176], respectively. They are summarised along with bowing factors for other ternaries studied in this section in Table 5.6. Direct bandgap bowing factor of GaAsP in Table 5.6 agrees very well with experimental bowing factors (in the range 0.174-0.23eV) reported in [230,235], but is significantly smaller than 0.54eV measured by Kim et al. [226].

Table 5.6- Bandgap bowing factors for III-V ternary compounds of interest.

<table>
<thead>
<tr>
<th>Compound</th>
<th>( c_r ) (eV)</th>
<th>( c_x ) (eV)</th>
<th>( c_L ) (eV)</th>
<th>( c_V ) (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{In}_{1-x}\text{Al}_x\text{As} )</td>
<td>0.304</td>
<td>0.713</td>
<td>0.617</td>
<td>-0.712</td>
</tr>
<tr>
<td>( \text{Al}<em>x\text{Ga}</em>{1-x}\text{As} )</td>
<td>0.0</td>
<td>0.143</td>
<td>0.15</td>
<td>—</td>
</tr>
<tr>
<td>( \text{In}_{1-x}\text{Ga}_x\text{As} )</td>
<td>0.42</td>
<td>0.509</td>
<td>1.087</td>
<td>-0.45</td>
</tr>
<tr>
<td>( \text{In}_{1-x}\text{Ga}_x\text{P} )</td>
<td>0.67</td>
<td>0.17</td>
<td>0.41</td>
<td>-0.11</td>
</tr>
<tr>
<td>( \text{GaAs}<em>{y}\text{P}</em>{1-y} )</td>
<td>0.21</td>
<td>0.21</td>
<td>0.42</td>
<td>0.0</td>
</tr>
<tr>
<td>( \text{InAs}<em>{y}\text{P}</em>{1-y} )</td>
<td>0.286</td>
<td>0.187</td>
<td>0.115</td>
<td>-0.506</td>
</tr>
</tbody>
</table>

The direct bandgap of \( \text{In}_{1-x}\text{Ga}_x\text{As}_y\text{P}_{1-y} \) lattice-matched to InP has been studied by many groups (for a summary and list of references see [172]). Here, an appropriate value of \( m_0 \) is chosen to fit those of the above reports that are rather consistent. Figure 5.10 shows the variation of bandgap with arsenic mole fraction \( y \) for InGaAsP lattice-matched to InP reported in [175,236-239]. These authors have fitted quadratic equations to the experimental
data of bandgap for InGaAsP lattice-matched to InP substrate. Also shown in Fig. 5.10 are the variations of bandgap according to Eqs. (5.5) and (5.6) with \( m_Q \) as a parameter. It is found that the best match between the results of Eqs. (5.5) and (5.6) and experimental data is obtained using \( m_Q = 0.4 \). The bandgap of InGaAsP lattice-matched to GaAs, however, is not well studied in the literature. Figure 5.11 shows the variation of direct and indirect bandgaps of InGaAsP lattice-matched to GaAs with arsenic mole fraction, \( y \), calculated using \( m_Q = 0.4 \). The experimental data points in [240] and the two data points mentioned in [227] are also marked in this figure. The agreement between the results of this work and the data given by Cho et al. [240] is very good.

### 5.3.2. Unified Model for Electron Affinity

The behaviour of band discontinuities in semiconductors has been the subject of extensive experimental and theoretical research. Experimentally, reliable measurements exist for a number of lattice-matched III-V heterostructures; most notable of these are the Al\(_x\)Ga\(_{1-x}\)As/GaAs system [92,191] and the InP/In\(_{0.53}\)Ga\(_{0.47}\)As, In\(_{0.52}\)Al\(_{0.48}\)As/In\(_{0.53}\)Ga\(_{0.47}\)As [182,192] and In\(_{0.49}\)Ga\(_{0.51}\)P/GaAs [31] interfaces. This information is very useful in designing devices for specific characteristics. In Npn HBTs, as shown in chapter 2, it is desirable to have the majority of the bandgap discontinuity in the valence band in order to improve the emitter injection efficiency and current gain of the device. In addition, conduction band discontinuities at the B/C heterojunction can affect the output characteristics, leading to voltage offsets and/or voltage-dependent output parameters. In Pnp HBTs, on the contrary, it is most advantageous to have B/E heterojunction from a material with maximum conduction band discontinuity and a smooth valence band for hole transport.

Due to the importance of the conduction and valence band alignment at a heterojunction interface, this section presents a unified model for the conduction and valence band edges in various III-V semiconductors in an absolute energy scale. In this unified energy system the so-called vacuum energy, \( E_{\text{vac}} \), is chosen as the reference energy and all the conduction and valence band edges are measured with respect to that. Since the bandgap of III-V compounds has already been formulated, the only additional parameter introduced in this section for a complete formulation of band edges is electron affinity, \( \chi \), which is defined as the energy
difference between the vacuum energy and the conduction band minimum.

The procedure of finding the electron affinity of various materials adopted in this work relies on the available measured conduction and valence band offsets, $\Delta E_C$ and $\Delta E_V$, for the lattice-matched systems listed in the beginning of this section, plus the much studied strained In$_{1-x}$Ga$_x$As/GaAs heterojunctions. In contrast to most of the published work in the literature which assume a linear variation of valence band offsets (i.e., zero bowing parameter) with composition in ternary and quaternary alloys [241-242], in this work non-zero valence band bowing factors are used in general. No attempt has been made to fix the electron affinity of the binary compounds like GaP and GaAs with respect to each other. Instead, as is made clearer later in this section, the electron affinity of GaP is adjusted by the conduction band discontinuities in InGaAs/GaAs, InGaAs/InP, and InGaP/GaAs heterojunctions. These band offsets may themselves be adjusted independently by the users of the simulation package to various magnitudes within a reasonable experimentally reported range.

One significant question about the organisation of the present electron affinity model concerns the validity of the universal energy scale assumption for heterostructures. This assumption is valid if the experimental band alignments can be shown to conform the mathematical properties of an equivalence relation: symmetry, commutativity, and transitivity [241]. In the context of heterojunctions, symmetry means that there is no band discontinuity at the junction between a semiconductor and itself; since in this case there is no junction, this property is always satisfied. Commutativity means that the same alignment is obtained if the ordering of the materials is reversed, and transitivity means that the band alignment at a junction A/C is equal to the superposition of the alignments at junctions A/B and B/C. It appears that transitivity is satisfied in those cases for which high quality interfaces have been grown [243]. The calculations in [244] also shows that transitivity rule is well satisfied for all the II-VI and III-V binary semiconductors. However, there are experimental indications that commutativity is violated to the level of a few hundredth of eV in some systems. Vanelle et al. [225] have stated that due to the sequence of growth for GaAs grown on InGaP and InGaP on GaAs, calculated valence band offset is 320meV for the former and 355meV for the latter, and therefore, the commutativity rule breaks down. However, these authors decided that the difference is small enough to be neglected, and hence, they considered a single "effective" $\Delta E_V$ for InGaP/GaAs heterojunctions. In [245] geometrically symmetric InGaP/GaAs DHBTs were fabricated. The output characteristics of the emitter-up and emitter-down configurations were studied and shown to have asymmetries. This electrical asymmetry has been interpreted in terms of interfacial differences between the InGaP/GaAs upper and lower heterojunctions. It seemed that the conduction band spike for GaAs grown on InGaP was much larger than that of InGaP on GaAs, in agreement to the above discussion. However, the same authors related this phenomena to the possibility of the upper junction being graded due to the greater efficiency of As incorporation into the grown layers as compared to P incorporation. Therefore, this evidence will not invalidate the commutativity rule, since one of the junctions
was not abrupt. Moreover, measurements of the conduction band discontinuity for the InGaP/InGaAsP heterojunction system lattice-matched to GaAs has been made by Cho et al. [240] on both normal (InGaP on InGaAsP) and inverted (InGaAsP on InGaP) structures. Their results reveal that the conduction to valence band offset ratio for both the normal and inverted structures are very similar (in the range $\frac{18 \pm 2}{82 \pm 2}$), supporting the commutativity rule.

In conclusion, the assumptions of commutativity and transitivity seem to be sound and will be followed in this work.

Before moving on to the actual formulation of electron affinity, a final comment is necessary as to the temperature dependence of the electron affinity and band offsets. According to Adachi in [246: section 8.2.3] and [192: section 3.4], since the temperature variations of the bandgap energies are very similar among the III-V materials, the band offsets and the resultant offset ratios in the III-V heterojunction systems can be successfully assumed to be independent of temperature. Although the conduction band offset voltage (written as $\chi_1 - \chi_2$ at the interface between two semiconductors 1 and 2) is assumed temperature independent, the actual values of electron affinities in both sides of the heterojunction interface may vary with temperature. However, since the actual magnitude of the electron affinity is of no physical significance in the context of present semiconductor device modelling, it is assumed that electron affinity is a completely temperature independent parameter.

### 5.3.2.a. GaAs and Al$_x$Ga$_{1-x}$As

The electron affinity of GaAs is assumed to be 4.07 eV [173]. The electron affinity of direct bandgap Al$_x$Ga$_{1-x}$As ($0 \leq x \leq 0.4$) can be modelled based on the ratio of the discontinuity in the conduction band to the total bandgap discontinuity ($\Delta E_C / \Delta E_g$) for an abrupt AlGaAs/GaAs heterojunction. This ratio has been measured by various techniques and various groups and the majority of the recently reported values vary in the range 0.60 to 0.70 (for an excellent review see [191]). In this work, a ratio of 0.62 measured by Watanabe et al. [19] has been adopted, since their measurement technique seems to be reliable and they measured both $\Delta E_C$ and $\Delta E_V$ independently and checked the consistency of their results (i.e., $\Delta E_C + \Delta E_V = \Delta E_g$). Therefore, the electron affinity of the direct bandgap AlGaAs can be written as:

$$\chi(Al_xGa_{1-x}As) = 4.07 - 0.62 [E_g(\Gamma) - E_g(0)] \text{ for } 0 \leq x \leq 0.4$$  \hspace{1cm} (5.26)

For $0.4 < x \leq 1.0$, a linear interpolation between the electron affinities of Al$_{0.4}$Ga$_{0.6}$As and AlAs will be assumed. From the available data in the literature for the direct and indirect band offsets of AlGaAs/GaAs, Missous (in [92]) has concluded that: $\Delta E_V(Al_xGa_{1-x}As/GaAs) = (0.51 \pm 0.04) \times eV$ for the entire range of composition. Using this conclusion, $\Delta E_V(AlAs/GaAs)$ can be found as 0.51 eV, which is translated into $\Delta E_C(AlAs/GaAs) = 0.231 eV$ or $\chi(AlAs) = 3.839 eV$. 

74
5.3.2.b. In\textsubscript{1-x}Ga\textsubscript{x}As

As to the alignment of the band edges in In\textsubscript{1-x}Ga\textsubscript{x}As relative to GaAs, the band offset ratio measured for strained InGaAs on GaAs is used. The percentage of the total bandgap discontinuity existing in the conduction band (valence band) is denoted as \( Q_C \) (\( Q_V \)):

\[
Q_{C,V} = \frac{\Delta E_{C,V}}{\Delta E_g}
\]  

(5.27)

The \( Q_C \) values obtained for In\textsubscript{1-x}Ga\textsubscript{x}As/GaAs heterojunctions obtained in the literature are extensively studied and summarised by Adachi in [192: section 3.4]. The results for \( x \geq 0.5 \) show a large discrepancy and vary in the range \( Q_C = 0.60 \pm 0.24 \). These data are scattered and show no clear trend. However, the majority of the data lies in the range \( Q_C = 0.64 \pm 0.06 \). For \( x < 0.5 \) no information is available in the literature apart from \( x = 0.0 \) (i.e., InAs/GaAs heterojunction) for which three measurements are available all within a very close range: \( Q_C \) (InAs/GaAs) = 0.83, 0.84, 0.90 [192: section 3.4].

In this work, the variation of electron affinity with gallium mole fraction in In\textsubscript{1-x}Ga\textsubscript{x}As is modelled in the form of a parabolic relation with an effective bowing parameter. If the bowing of the \( \Gamma \)-bandgap is \( c_\Gamma \) (= \( c_{\Gamma-V} \)) and the bowing of the valence band with respect to the vacuum energy is \( c_V \) (= \( c_{V-vac} \)), then the bowing of electron affinity would be:

\[
c_X = c_{\text{vac}-\Gamma} = -(c_{\Gamma-V} + c_{V-vac}) = -(c_\Gamma + c_V)
\]  

(5.28)

Hence, the quadratic form of the electron affinity can be written as:

\[
\chi(\text{In}_{1-x}\text{Ga}_x\text{As}) = (1-x) \cdot \chi(\text{InAs}) + x \cdot \chi(\text{GaAs}) + x(1-x)[c_\Gamma(\text{InGaAs}) + c_V(\text{InGaAs})]
\]  

(5.29)

Note the positive sign in front of the last term on the RHS of (5.29). For \( \chi(\text{InAs}) \), here \( Q_C \) (InAs/GaAs) = 0.84 is used which gives:

\[
\chi(\text{InAs}) = \chi(\text{GaAs}) + 0.84[E_{g\Gamma}(\text{GaAs},300K) - E_{g\Gamma}(\text{InAs},300K)]
\]  

(5.30)

In (5.30) the room temperature value of \( \Delta E_g \) is used, and it is assumed that \( \chi(\text{InAs}) \) remains constant as temperature varies. This makes \( Q_C \) very slightly different from 0.84 at other temperatures. With all other parameters of Eq. (5.29) known, a value of \( c_V(\text{InGaAs}) \) is chosen in this work that gives \( Q_C = 0.64 \pm 0.06 \) for \( x \geq 0.5 \). As shown in Fig. (5.12),
c_V (InGaAs) = -0.45eV satisfies this condition. A negative c_V means upward bowing of the valence band with respect to the vacuum energy level. An upward bowing for the valence band is a common case for all the III-V ternaries studied in this work. Finally, using (5.29) and (5.30), the electron affinities of InAs and In_{0.53}Ga_{0.47}As can be calculated as 4.970 and 4.540 eV, respectively.

5.3.2.c. InP

Numerous data are available for the conduction and valence band offsets at the In_{0.53}Ga_{0.47}As/InP heterojunction interface. These data are studied and summarised by various authors, e.g. [172,182,186,192]. Majority of the reported conduction and valence band offsets are within the range (185-265)meV and (340-420)meV, respectively [192]. In this work, $\Delta E_C (\text{In}_{0.53}\text{Ga}_{0.47}\text{As}/\text{InP}) = 0.225$ eV is assumed by default, which results in:

$$\chi(\text{InP}) = \chi(\text{In}_{0.53}\text{Ga}_{0.47}\text{As}) - \Delta E_C (\text{In}_{0.53}\text{Ga}_{0.47}\text{As}/\text{InP}) = 4.315 \text{eV} \quad (5.31)$$

5.3.2.d. In_{1-x}Ga_xP

As to the electron affinity of In_{1-x}Ga_xP, the situation is much less certain. It is widely observed that In_{1-x}Ga_xP epilayers exhibit a long-term ordering under suitable growth conditions. (See also the discussion on bandgap of InGaP in subsection 5.3.1.) This long-term ordering results in a remarkable decrease of the bandgap and a splitting of the valence band in InGaP [247-248]. The amount of the bandgap reduction is approximately proportional to the degree of the atomic scale ordering. The degree and the microstructure of ordering depends strongly on the substrate orientation and growth parameters such as substrate temperature, growth rate, and III/V ratio [247]. Since the bandgap of InGaP depends on the degree of ordering, the band alignment of In_{0.49}Ga_{0.51}P/GaAs heterojunctions should be also influenced by the degree of ordering in In_{0.49}Ga_{0.51}P. This projects into a wide range of conduction band discontinuity reported in the literature for In_{0.49}Ga_{0.51}P/GaAs heterojunctions, i.e., 0.03-0.39 eV! [31,191,249-251]. Usually the ordered In_{0.49}Ga_{0.51}P gives rise to smaller bandgap and smaller $\Delta E_C$ [31].

Another possible reason for the observed discrepancy in the values of $\Delta E_C$ can be the existence of an interfacial layer of a few monolayers thickness with unknown composition in-between the InGaP/GaAs heterojunction. The interfacial layer, likely InGaAsP [225], could affect the apparent band discontinuity measured by different techniques by acting as a grading layer between InGaP and GaAs.

It is also worth mentioning that recently a type-II band alignment for ordered In_{0.49}Ga_{0.51}P/GaAs is suggested, with electrons confined in the In_{0.49}Ga_{0.51}P side and holes in the GaAs side [247,252]. This type of alignment, which is equivalent to having a negative $\Delta E_C$ according to the definitions in chapter 2, makes the choice of an appropriate electron affinity for In_{0.49}Ga_{0.51}P even more complicated.
Considering the above facts, it would be almost impossible to assign a unique electron affinity to In$_{0.49}$Ga$_{0.51}$P. The conduction band discontinuity at the In$_{0.49}$Ga$_{0.51}$P/GaAs heterojunction will remain as a user-defined parameter; here only the default value and a reasonable range of variation will be provided. As to the default value, \( \Delta E_C(\text{In}_{0.49}\text{Ga}_{0.51}\text{P}/\text{GaAs}) = 0.14\text{eV} \) suggested by Kobayashi et al. [253] and used ever since by many other authors will be adopted. This value is also very similar to that measured in our research group [254].

Based on the above default value or the user's choice of \( \Delta E_C(\text{In}_{0.49}\text{Ga}_{0.51}\text{P}/\text{GaAs}) \), the electron affinity of In$_{1-x}$Ga$_x$P will be formulated for the entire range of composition as follows. Referring to Fig. 5.13, electron affinity of In$_{1-x}$Ga$_x$P varies with composition parabolically from \( \chi(\text{InP}) \) to the electron affinity of the direct-indirect cross-over point with an effective bowing factor of \(- (c_\Gamma(\text{InGaP}) + c_\chi(\text{InGaP})) \) (see the discussion for electron affinity of In$_{1-x}$Ga$_x$As). After the cross-over point, electron affinity changes its slope and varies with a different effective bowing factor of \(- (c_\chi(\text{InGaP}) + c_\chi(\text{InGaP})) \) to reach \( \chi(\text{GaP}) \). Therefore, one can write:

\[
\begin{align*}
\chi_1(x) &= (1-x) \cdot \chi(\text{InP}) + x \cdot \left[ \chi(\text{GaP}) - E_g(\text{GaP}) + E_gx(\text{GaP}) \right] + x(1-x) \left[ c_\Gamma(\text{InGaP}) + c_\chi(\text{InGaP}) \right] \\
\chi_2(x) &= x \cdot \chi(\text{GaP}) + (1-x) \cdot \left[ \chi(\text{InP}) + E_g(\text{InP}) - E_gx(\text{InP}) \right] + x(1-x) \left[ c_\chi(\text{InGaP}) + c_\chi(\text{InGaP}) \right]
\end{align*}
\]

(5.32) \quad \text{and} \quad (5.33)

\[
\chi(\text{In}_{1-x}\text{Ga}_x\text{P}) = \max\{\chi_1(x), \chi_2(x)\} \quad (5.34)
\]

To fully describe the electron affinity of In$_{1-x}$Ga$_x$P, one needs to find appropriate values for \( c_\chi(\text{InGaP}) \) and \( \chi(\text{GaP}) \). \( c_\chi(\text{InGaP}) = -0.11\text{eV} \) is adopted from [210]. To find \( \chi(\text{GaP}) \), one can use:

\[
\chi(\text{In}_{0.49}\text{Ga}_{0.51}\text{P}) = \chi_1(0.51) = \chi(\text{GaAs}) - \Delta E_C(\text{In}_{0.49}\text{Ga}_{0.51}\text{P}/\text{GaAs}) \quad (5.35)
\]

Applying the default value of \( \Delta E_C(\text{In}_{0.49}\text{Ga}_{0.51}\text{P}/\text{GaAs}) = 0.14\text{eV} \), the electron affinity of GaP can be obtained as: \( \chi(\text{GaP}) = 3.785\text{eV} \).

### 5.3.2.e. In$_{1-x}$Ga$_x$As$_y$P$_{1-y}$

To formulate the electron affinity in InGaAsP one needs to know the valence band bowing factors for GaAs$_y$P$_{1-y}$ and InAs$_y$P$_{1-y}$. If \( c_\chi_1 \) and \( c_\chi_2 \) are defined as:

\[
\begin{align*}
c_\chi_1 &\equiv c_\Gamma + c_\chi \\
c_\chi_2 &\equiv c_\chi + c_\chi
\end{align*}
\]

77
for any ternary compound, then following the method given for the electron affinity of In$_{1-x}$Ga$_x$P one can write:

\[
\chi(\text{In}_{1-x}\text{Ga}_x\text{As}, \text{P}_{1-y}) = \max \{\chi_1(x, y), \chi_2(x, y)\} \tag{5.36}
\]

\[
\chi_1(x, y) = (1 - x)y \cdot \chi(\text{InAs}) + (1 - x)(1 - y) \cdot \chi(\text{InP}) + xy \cdot \chi(\text{GaAs}) + x(1 - y) \left[ \chi(\text{GaP}) + E_{g\chi}(\text{GaP}) - E_{g\Gamma}(\text{GaP}) \right] + x(1 - x) \cdot c_{\chi_1}(\text{InGaAs}) + x(1 - x)(1 - y) \cdot c_{\chi_1}(\text{InGaP}) + (1 - x)y(1 - y) \cdot c_{\chi_1}(\text{InAsP}) + xy(1 - y) \cdot c_{\chi_1}(\text{GaAsP}) \tag{5.37}
\]

\[
\chi_2(x, y) = (1 - x)y \cdot \left[ \chi(\text{InAs}) + E_{g\Gamma}(\text{InAs}) - E_{g\chi}(\text{InAs}) \right] + (1 - x)(1 - y) \cdot \left[ \chi(\text{InP}) + E_{g\Gamma}(\text{InP}) - E_{g\chi}(\text{InP}) \right] + xy \cdot \left[ \chi(\text{GaAs}) + E_{g\Gamma}(\text{GaAs}) - E_{g\chi}(\text{GaAs}) \right] + x(1 - y) \cdot \chi(\text{GaP}) + (1 - x)y(1 - y) \cdot c_{\chi_2}(\text{InGaAs}) + x(1 - x)(1 - y) \cdot c_{\chi_2}(\text{InGaP}) + (1 - x)y(1 - y) \cdot c_{\chi_2}(\text{InAsP}) + xy(1 - y) \cdot c_{\chi_2}(\text{GaAsP}) \tag{5.38}
\]

The valence band bowing factors of GaAsP and InAsP are taken from the calculations by Chen and Sher [176] as: $c_V(\text{GaAsP}) = 0$ and $c_V(\text{InAsP}) = -0.506$ eV. The conduction and valence band offsets of lattice-matched InGaAsP/InP and InGaAsP/GaAs heterojunctions are calculated using (5.36)-(5.38) and plotted in Fig. 5.14 against the arsenic composition. The conduction band offset ratio, $Q_C$, obtained from Fig. 5.14 for lattice-matched InGaAsP/InP heterojunction varies in the range 0.27 to 0.38, in agreement to the experimentally observed range of (0.20-0.39) [186,192]. In particular, the conduction band offset of lattice-matched InGaAsP/InP is formulated in one work (see [192]) as:

\[
\Delta E_C(y) = 0.268y + 0.003y^2 \text{ eV}
\]

and shown in Fig. 5.14 with open triangles. Obviously, in that work $\Delta E_C(\text{In}_{0.53}\text{Ga}_{0.47}\text{As}/\text{InP}) = 0.271$ eV is used.

If the same conduction band offset for In$_{0.53}$Ga$_{0.47}$As/InP is used in the present work instead of the default value, much better agreement to the above data will be obtained for the entire range of composition. The

\[\text{Figure 5.14- Conduction and valence band offsets at the lattice-matched InGaAsP/InP and InGaAsP/GaAs heterojunctions using Eqs. (5.36)-(5.38) and the default values of conduction band offsets in this work for InP/In$_{0.53}$Ga$_{0.47}$As and In$_{0.49}$Ga$_{0.51}$P/GaAs heterojunctions. Experimental data for the conduction band offsets of InGaAsP/InP (open triangles) [192] and InGaAsP/GaAs (solid circles) [240] are also shown for comparison.}\]
Figure 5.15- The relative alignment of the conduction and valence bands for some III-V binary and ternary compounds. All the mentioned energies are relative to $E_V(GaAs)$.

conduction band offsets between $In_{0.49}Ga_{0.51}P$ and $InGaAsP$ lattice-matched to $GaAs$ is also measured by Cho et al. [240] for arsenic mole fractions in the range (0.73-1.0) (solid circles in Fig. 5.14). Quite similar to the above case for the $InGaAsP/InP$ conduction band offset, if one uses $\Delta E_C(In_{0.49}Ga_{0.51}P/GaAs) = 0.092eV$ as measured by Cho et al. instead of our default value, excellent agreement between the results of this work and those measured in [240] can be obtained.

5.3.2.f. $In_{1-x}Al_xAs$

Due to the direct-to-indirect bandgap crossover in $In_{1-x}Al_xAs$, the formulation of electron affinity in this material is exactly similar to that of $In_{1-x}Ga_xP$. The only difference in this case is that the electron affinity of the two binary constituents (AlAs and InAs) are already known, and $c_v(InAlAs)$ forms the only unknown fitting parameter. The valence band bowing factor can be found from the published results of the band offsets for the lattice-matched $In_{0.52}Al_{0.48}As/In_{0.53}Ga_{0.47}As$ heterojunction. The conduction and valence band offsets at the latter heterojunction are almost consistently reported in the ranges (0.48-0.55) and (0.13-0.24), respectively [172,182,186,192]. In this work, $\Delta E_C(In_{0.52}Al_{0.48}As/In_{0.53}Ga_{0.47}As) = 0.52eV$ is adopted as the default value which is more consistently reported in the literature and also is in the middle of the above range. Using the default value of $\Delta E_C$ and the formulations similar to (5.32)-(5.35), one can find:

$$\chi(In_{0.52}Al_{0.48}As) = \chi(In_{0.53}Ga_{0.47}As) - \Delta E_C(In_{0.52}Ga_{0.48}As/In_{0.53}Ga_{0.47}As) = 4.02eV$$

and $c_v(InAlAs) = -0.712eV$, which is somehow larger than $c_v = -0.506eV$ calculated in [176].

A final comment should be made with regard to the band alignment of the $In_{0.52}Al_{0.48}As/InP$ heterojunction. The staggered (type-II) band alignment, with both the conduction and valence
bands of In\(_{0.52}\)Al\(_{0.48}\)As above those of InP, is generally agreed on. Using the transitivity rule and the above default values of conduction band discontinuity at In\(_{0.53}\)Ga\(_{0.47}\)As/InP and In\(_{0.52}\)Al\(_{0.48}\)As/InP, one can find \(\Delta E_C = -295\) meV and \(\Delta E_V = 195\) meV for In\(_{0.52}\)Al\(_{0.48}\)As/InP. Not many experimentally measured results are available in the literature for this heterojunction. Two existing reports give \(\Delta E_C = -252\) and -390 meV [192,255]; our default value is somewhere in-between the two extremes.

As a conclusion to this section, the alignment of the conduction and valence bands of various binary and ternary compounds of interest is shown in Fig. 5.15. In this figure, \(E_V(GaAs)\) is assumed to be zero.

### 5.3.3. Effective Mass and Effective Density of States

In this work, parabolic conduction and valence bands are assumed. Therefore, for each of the three conduction band minima (\(\Gamma\), \(X\), and \(L\)) one can define an effective mass which will be used to find the density of states in the corresponding band. This effective mass is therefore called *density of states (DOS) effective mass* [161]. The DOS effective masses can be defined as [161,172]:

\[
m_{nv,DOS} = \left( N_{bv}^{\nu} \cdot m_{lv}^{\nu} \cdot m_{tv}^{\nu} \right)^{1/3} \quad \nu = \Gamma, X, L
\]

where \(N_{bv}\) is the number of equivalent \(v\)-band minima in the first Brillouin zone (\(N_{b\Gamma} = 1, N_{bX} = 3, N_{bL} = 4\)), and \(m_{tv}, m_{lv}\) are the transverse and longitudinal masses of the \(v\)-minima, respectively (for \(\Gamma\)-band: \(m_{t\Gamma} = m_{l\Gamma}\)). On the other hand, the conductivity effective mass \(m_{nv,C}\), which can be used for mobility and conductivity calculations, is defined as:

\[
\frac{1}{m_{nv,C}} = \frac{1}{3} \left( \frac{2}{m_{tv}} + \frac{1}{m_{lv}} \right)
\]

Since \(m_{t\Gamma} = m_{l\Gamma}\), the relation \(m_{n\Gamma,DOS} = m_{n\Gamma,C}\) is always sustained. As shown in section 5.4, the formulation of mobility is based on empirical relations rather than theoretical calculations using conductivity effective mass. Therefore, in this work only the DOS effective mass will be dealt with, and the subscript DOS will be dropped hereafter.

Similarly, the density of states in the valence band of the III-V compounds can be characterised by the effective masses of three subbands, all in the centre of the Brillouin zone (i.e., \(\Gamma\)-point): two degenerate subbands of light and heavy holes, and the split-off subband. However, the split-off band, being depressed in energy, is only sparsely populated and is often ignored for unstrained semiconductors [161]. Therefore, the valence band density of states effective mass \(m_p^*\) will be formulated using the light and heavy hole effective masses, \(m_{lh}\) and \(m_{hh}\). But before turning to the actual formulation of the density of states in terms of DOS effective masses, appropriate values will be assigned to \(m_{n\Gamma}, m_{nX}, m_{nL}, m_{lh},\) and \(m_{hh}\) for the binary, ternary, and quaternary compounds of interest.
5.3.3.a. Effective Mass of Binary Compounds

Table 5.7 summarises the reported values of effective masses for the five binary compounds of interest (i.e., AlAs, GaAs, InAs, InP, and GaP). In this work, it is tried to choose those values of the effective masses which are most consistently reported in the literature.

Table 5.7- Reported values of the effective masses for GaAs, AlAs, InAs, InP, and GaP. The last column shows the effective masses chosen in this thesis.

<table>
<thead>
<tr>
<th>Effective Mass</th>
<th>Reported Values</th>
<th>References</th>
<th>Used Here</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_{nf}$ (GaAs)</td>
<td>0.063-0.067, 0.071</td>
<td>[92,191,225]</td>
<td>0.065</td>
</tr>
<tr>
<td>$m_{nh}$ (GaAs)</td>
<td>0.85, 0.98, 1.02, 1.17</td>
<td>[92,183]</td>
<td>0.85</td>
</tr>
<tr>
<td>$m_{nl}$ (GaAs)</td>
<td>0.56, 0.70, 0.74</td>
<td>[92,183]</td>
<td>0.56</td>
</tr>
<tr>
<td>$m_{nh}$ (GaAs)</td>
<td>0.074-0.094</td>
<td>[92,174,176]</td>
<td>0.080</td>
</tr>
<tr>
<td>$m_{hh}$ (GaAs)</td>
<td>0.33-0.40, 0.48-0.51, 0.62</td>
<td>[92,181,184,225]</td>
<td>0.48</td>
</tr>
<tr>
<td>$m_{nf}$ (AlAs)</td>
<td>0.11, 0.124, 0.135, 0.15, 0.172, 0.22</td>
<td>[92,176,183]</td>
<td>0.15</td>
</tr>
<tr>
<td>$m_{nh}$ (AlAs)</td>
<td>0.5, 0.71, 0.79</td>
<td>[92]</td>
<td>0.71</td>
</tr>
<tr>
<td>$m_{nl}$ (AlAs)</td>
<td>0.66, 0.78</td>
<td>[92,183]</td>
<td>0.66</td>
</tr>
<tr>
<td>$m_{nh}$ (AlAs)</td>
<td>0.15-0.18, 0.208</td>
<td>[92,182]</td>
<td>0.16</td>
</tr>
<tr>
<td>$m_{hh}$ (AlAs)</td>
<td>0.40-0.51, 0.75-0.81</td>
<td>[92,182]</td>
<td>0.76</td>
</tr>
<tr>
<td>$m_{nf}$ (InAs)</td>
<td>0.021-0.024</td>
<td>[172,176]</td>
<td>0.023</td>
</tr>
<tr>
<td>$m_{nh}$ (InAs)</td>
<td>0.64</td>
<td>[184]</td>
<td>0.64</td>
</tr>
<tr>
<td>$m_{nl}$ (InAs)</td>
<td>0.29</td>
<td>[184]</td>
<td>0.29</td>
</tr>
<tr>
<td>$m_{nh}$ (InAs)</td>
<td>0.025-0.027, 0.030</td>
<td>[176,256-257]</td>
<td>0.026</td>
</tr>
<tr>
<td>$m_{hh}$ (InAs)</td>
<td>0.34-0.41, 0.6</td>
<td>[184,225]</td>
<td>0.45</td>
</tr>
<tr>
<td>$m_{nh}$ (InP)</td>
<td>0.068, 0.076-0.082</td>
<td>[183,186]</td>
<td>0.079</td>
</tr>
<tr>
<td>$m_{nl}$ (InP)</td>
<td>0.66, 0.676</td>
<td>[184,205]</td>
<td>0.676</td>
</tr>
<tr>
<td>$m_{nh}$ (InP)</td>
<td>0.63, 0.655</td>
<td>[184,205]</td>
<td>0.655</td>
</tr>
<tr>
<td>$m_{hh}$ (InP)</td>
<td>0.081, 0.089, 0.120-0.123</td>
<td>[184,186,210]</td>
<td>0.120</td>
</tr>
<tr>
<td>$m_{nh}$ (InP)</td>
<td>0.45-0.61, 0.85</td>
<td>[186,225]</td>
<td>0.56</td>
</tr>
<tr>
<td>$m_{nh}$ (GaP)</td>
<td>0.0925, 0.126, 0.17, 0.254</td>
<td>[183,225]</td>
<td>0.126</td>
</tr>
<tr>
<td>$m_{nh}$ (GaP)</td>
<td>0.79, 0.82, 1.04</td>
<td>[181,183-184]</td>
<td>0.82</td>
</tr>
<tr>
<td>$m_{hh}$ (GaP)</td>
<td>0.326, 0.45, 0.54, 0.79</td>
<td>[182,225]</td>
<td>0.54</td>
</tr>
</tbody>
</table>

5.3.3.b. Effective Mass of Ternary Compounds

A linear interpolation of the form (5.1) and (5.3) is used in this thesis to find the effective masses of the ternary and quaternary compounds from their constituent binaries. This gives, for instance:

$$m(\text{In}_{1-\chi}\text{Ga}_{\chi}\text{As}_{\chi}\text{P}_{1-\chi}) =$$

\[
(l - x)y \cdot m(\text{InAs}) + (l - x)(l - y) \cdot m(\text{InP}) + xy \cdot m(\text{GaAs}) + x(l - y) \cdot m(\text{GaP})
\]

This is to be compared against the interpolation of the form:
\[
\frac{1}{m(\text{In}_{1-x}\text{Ga}_x\text{As}_y\text{P}_{1-y})} = \frac{(1-x)y}{m(\text{InAs})} + \frac{(1-x)(1-y)}{m(\text{InP})} + \frac{xy}{m(\text{GaAs})} + \frac{x(1-y)}{m(\text{GaP})}
\]  

(5.43)

which has been claimed by Adachi [172] to be more physically justified. Comparison between the results obtained from Eqs. (5.42) and (5.43) and those reported in the literature for \(m_r\), \(m_h\), and \(m_{hh}\) in \(\text{In}_{0.53}\text{Ga}_{0.47}\text{As}\), and \(m_r\) in \(\text{In}_{0.49}\text{Ga}_{0.51}\text{P}\) and \(\text{In}_{0.52}\text{Al}_{0.48}\text{As}\) is made in Table 5.8. It is clear from Table 5.8 that the linear interpolation scheme explained by Eq. (5.42) gives much better agreement to the reported data in the literature.

<table>
<thead>
<tr>
<th>Effective Mass</th>
<th>Reported Range</th>
<th>Reference(s)</th>
<th>Using (5.42)</th>
<th>Using (5.43)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(m_r(\text{In}<em>{0.53}\text{Ga}</em>{0.47}\text{As}))</td>
<td>0.038-0.044</td>
<td>[186,258]</td>
<td>0.043</td>
<td>0.033</td>
</tr>
<tr>
<td>(m_h(\text{In}<em>{0.53}\text{Ga}</em>{0.47}\text{As}))</td>
<td>0.050-0.052</td>
<td>[186,258]</td>
<td>0.051</td>
<td>0.038</td>
</tr>
<tr>
<td>(m_{hh}(\text{In}<em>{0.53}\text{Ga}</em>{0.47}\text{As}))</td>
<td>0.46-0.47, 0.50</td>
<td>[186,192,257]</td>
<td>0.464</td>
<td>0.464</td>
</tr>
<tr>
<td>(m_r(\text{In}<em>{0.49}\text{Ga}</em>{0.51}\text{P}))</td>
<td>0.092-0.105,0.118</td>
<td>[187,250,259-260]</td>
<td>0.103</td>
<td>0.098</td>
</tr>
<tr>
<td>(m_r(\text{In}<em>{0.52}\text{Al}</em>{0.48}\text{As}))</td>
<td>0.070-0.075,0.081</td>
<td>[192,258]</td>
<td>0.084</td>
<td>0.039</td>
</tr>
</tbody>
</table>

Figure 5.16 shows the available experimental data collected in [92] and theoretically calculated values by Chen and Sher [210] for \(m_r(\text{Al}_x\text{Ga}_{1-x}\text{As})\). The interpolated effective masses using Eqs. (5.42) and (5.43) are also shown in this figure. It can be observed that the linear interpolation scheme gives a very good agreement to the experimental data by Emlin et al. and Zachau et al. in [92] and calculated results obtained by Chen and Sher [210]. However, Eq. (5.43) gives lower values of effective masses compared with the available data. A similar comparison has been made in Fig. 5.17 for the available data of \(m_r(\text{In}_{1-x}\text{Ga}_x\text{As})\) with the interpolated values according to (5.42) and (5.43). Again a very good agreement to the experimental data can be obtained using a linear interpolation. It can also be shown that Eq. (5.42) gives a good fit to the data for the electron effective mass in \(\text{InGaAsP}\) lattice-matched to \(\text{InP}\) summarised in [172].

### 5.3.3.c. External Perturbations on Effective Mass

The electron effective mass is almost directly proportional to the energy gap [172]. Thus, the electron mass decreases with increasing temperature. Studies of the temperature effect on \(m_r(\text{GaAs})\) suggest that the effective mass changes by almost 2-7% in the temperature range \(60 < T < 300\text{K}\) [92]. In this work, however, due to the small change of the effective mass and also due to the lack of temperature dependent data in materials other than GaAs and InP, an effective mass independent of temperature will be assumed.

The DOS effective mass is also influenced by the carrier concentration. Higher doping levels usually give larger electron effective masses (see e.g., [260]). Such a change in the electron mass is due to the band filling and non-parabolicity of the conduction band [92,260]. To take this effect into account one needs to introduce the conduction band non-parabolicity factors.
Finally, increasing the pressure increases the effective mass, due mainly to the increase in the energy bandgap [172]. But the effect of pressure on material parameters is not included in this work.

5.3.3.d. Conduction and Valence Band Density of States

The relations between the effective masses and the densities of states in the corresponding bands are in the form [161]:

\[ N_V = 2 \left( \frac{2\pi \cdot m_p^* \cdot kT}{\hbar^2} \right)^{3/2} \]  
\[ N_C = 2 \left( \frac{2\pi \cdot m_n^* \cdot kT}{\hbar^2} \right)^{3/2} \]

Considering first the valence band, one has to take into account both the light and heavy hole subbands. The effect of the two subbands can be combined as follows:

\[ N_V = N_{V_l} + N_{V_h} \]  

where \( N_{V_l} \), \( N_{V_h} \) are the density of states in the light and heavy hole valence bands, respectively; their relations with \( m_{lh} \) and \( m_{hh} \) are similar to (5.44). Therefore, the hole effective mass, \( m_p^* \) in Eq. (5.44) can be written as:

\[ m_p^* = \left( m_{lh}^{3/2} + m_{hh}^{3/2} \right)^{2/3} \]
As to the conduction band density of states, the situation is slightly more complicated. In direct bandgap binaries such as GaAs, InAs, and InP, \( m_n^* = m_n \Gamma \) can be assumed. Similarly for GaP and AlAs, \( m_n^* = m_n X \). But near the direct-to-indirect crossover in materials such as In_{1-x}Ga_xP and Al_xGa_{1-x}As two or more conduction band minima may simultaneously contribute to the total conduction band density of states. In such circumstances, the relation between \( m_n^* \) and the DOS effective masses of individual conduction bands can be developed as follows. First, the electron concentration in each conduction band minimum can be written as:

\[
n_n = N_{CV} \cdot \frac{F_{1/2} \left( \frac{E_F - E_n}{kT} \right)}{\gamma_{nY} \cdot \exp \left( \frac{E_F - E_n}{kT} \right)}
\]

where again \( v \) stands for \( \Gamma \), \( X \), or \( L \), and using (5.45), \( N_{CV} \) is expressed as:

\[
N_{CV} = 2.509 \times 10^{19} \left( \frac{T}{300K} \right)^{3/2} \left( \frac{m_{nY}}{m_0} \right)^{3/2} \text{ in (cm}^{-3}\text{)}
\]

(Note that the second equality in Eq. (5.48) can be viewed as the definition of \( \gamma_{nY} \).) Now, the total electron concentration can be written as:

\[
N_C \cdot \gamma_n \cdot \exp \left( \frac{E_F - E_C}{kT} \right) \equiv n = n_{\Gamma} + n_{X} + n_{L} =
\]

\[
N_{C\Gamma} \cdot \gamma_{n\Gamma} \cdot \exp \left( \frac{E_F - E_{\Gamma}}{kT} \right) + N_{CX} \cdot \gamma_{nX} \cdot \exp \left( \frac{E_F - E_{X}}{kT} \right) + N_{CL} \cdot \gamma_{nL} \cdot \exp \left( \frac{E_F - E_{L}}{kT} \right)
\]

where \( E_C = \text{Min}\{E_\Gamma, E_X, E_L\} \). Therefore, from the above relations \( N_C \) can be defined as:

\[
N_C = N_{C\Gamma} \cdot \gamma_{n\Gamma} \cdot \exp \left( \frac{E_C - E_{\Gamma}}{kT} \right) + N_{CX} \cdot \gamma_{nX} \cdot \exp \left( \frac{E_C - E_{X}}{kT} \right) + N_{CL} \cdot \gamma_{nL} \cdot \exp \left( \frac{E_C - E_{L}}{kT} \right)
\]

![Figure 5.18- Variation of the effective electron masses in Al_xGa_{1-x}As, In_{1-x}Al_xAs, and In_{1-x}Ga_xAs with composition.](image-url)
Once $N_C$ is known, $m^*$ can be directly obtained from (5.45).

Figure 5.18 shows the variation of the effective electron masses ($m_{n\Gamma}$, $m_{nX}$, $m_{nL}$, and $m^*$) in $\text{Al}_{x}\text{Ga}_{1-x}\text{As}$, $\text{In}_{1-x}\text{Al}_x\text{As}$, and $\text{In}_{1-x}\text{Ga}_x\text{As}$ with composition, $x$. It can be observed from Fig. 5.18 that due to the contribution of two or more conduction band minima in the total density of states, $m^*$ finds a local maximum near the direct-to-indirect crossover in $\text{AlGaAs}$ and $\text{InAlAs}$. Otherwise, $m^*_n = m_{n\Gamma}$ or $m_{nX}$ for direct or indirect bandgap materials, respectively.

5.4. Carrier Mobility

5.4.1. Low-Field Mobility Model

Electron and hole mobilities, $\mu_n$ and $\mu_p$, are two of the most important parameters for characterising the transport of charged carriers and formulating the current in semiconductors. A great deal of attention has been paid in the literature to these parameters and many authors have measured the carrier mobilities for various III-V compounds by Hall effect measurement technique (for a detailed discussion on Hall effect measurement see [261]). Plenty of effort has also been focused on theoretical formulation of minority and majority carrier mobilities in semiconductors (see, for instance, [172,262-263]). In these analyses, a detailed physical formulation of various scattering mechanisms (such as ionised impurity scattering, acoustic phonon scattering, polar and non-polar optical phonon scatterings, carrier-carrier scattering, and alloy scattering) has been made to accurately determine the variation of carrier mobility with carrier concentration, temperature and, in some cases [262], the compensation ratio. However, these approaches require accurate values of some physical constants, like deformation potentials and alloy scattering potential, which are only available for widely studied semiconductors such as $\text{Si}$, $\text{GaAs}$, and $\text{InP}$. For ternary and quaternary compounds, the common approach is to use linear interpolations between the values for corresponding binaries [173-174]. But since the linear interpolation is proved to be not too realistic for some physical parameters (e.g., bandgap and thermal conductivity), these parameters will finally be treated as adjustable fitting parameters. One common example of this case is alloy scattering potential which is a totally empirical parameter [172,246]. Additionally, the above techniques are sometimes too much mathematically involved and time-consuming to be suitable for device simulation packages. Finally, some of the available scattering models are only applicable to limited ranges of carrier concentration and temperature. One good example of the latter is the Brooks-Herring model for ionised impurity scattering. In this model, which is used in almost all of the existing theoretical formulations of carrier mobility in the literature, the ionised impurity scattering limited mobility, $\mu_{II}$, can be written as [264-265]:

$$
\mu_{II} = \frac{64e^2}{q^3N_{II}} \sqrt{\frac{\pi}{m^*}} g_{BH} \left( \frac{24m^* e_s (kT)^2}{q^2h^2N_{II}} \right)^{3/2}
$$

(5.51)

where $N_{II}$ is the ionised impurity concentration and all other parameters have their usual
meanings. $g_{BH}$ is the Brooks-Herring function of the form:

$$g_{BH}(\gamma) = \left[ \ln(1+\gamma) - \frac{\gamma}{1+\gamma} \right]^{-1}$$  \hspace{1cm} (5.52)

It is straightforward to prove that in the limit of $\gamma \to 0$ (i.e., very small $T$ or large $N_I$), the Brooks-Herring model generates some strange asymptotic behaviour:

$$\mu_II \propto \frac{N_I}{T^{5/2}(m^*)^{5/2}} \text{ for } \gamma \to 0$$  \hspace{1cm} (5.53)

which means that at low temperatures (typically $T < 50K$) $\mu_II$ should increase with further reduction of temperature or, more strangely, with increasing the doping concentration! Therefore, it is well-known that the above model should not be used for very small temperatures or high doping concentrations.

Having the above disadvantages of the accurate theoretical modelling in mind, in this work it has been decided to use semi-empirical models for carrier mobilities. Semi-empirical models are functions which produce the correct trends of mobility variation with temperature, doping concentration, electric field, and alloy composition, but their parameters will be treated as totally empirical giving a good fit to the available experimental data. Models by Klaassen [266], Caughey and Thomas [267], and Arora et al. [268] are examples of semi-empirical mobility models. In this thesis, a Caughey-Thomas-like mobility model with temperature dependent parameters is used. Parameters of this model are best fitted to the available Hall data for a wide range of carrier concentration and temperature using least-squares method. Additionally, these parameters are found for a large number of III-V binary, ternary, and quaternary compounds using both the available data and appropriate interpolation schemes. The low-field mobility model is expressed as:

$$\mu_{LF}(N,T) = \mu_{\text{min}} + \frac{\mu_{\text{max}}(300K) \cdot (300K/T)^{\theta_1} - \mu_{\text{min}}}{1 + \left( \frac{N}{N_{\text{ref}}(300K) \cdot (T/300K)^{\theta_2}} \right)^{\Lambda}}$$  \hspace{1cm} (5.54)

where all the fitting parameters are non-negative. First we consider the physical sense of (5.54):

1. At very low doping concentration, mobility saturates at $\mu_{\text{max}}(T)$, the lattice-limited mobility, which itself reduces with increasing temperature ($\theta_1 > 0$).

2. At very high doping concentration, mobility saturates at $\mu_{\text{min}}$, which is temperature independent. This is consistent with the fact that degenerate semiconductors show an almost temperature insensitive metal-like mobility [191-192]. In particular, Lovejoy et al. [269] have shown that for n-type (p-type) GaAs samples doped to $\sim4\times10^{18}$ cm$^{-3}$ the majority electron (hole) mobility is almost constant at $\sim2000$ ($\sim120$) for the temperature
range $77K < T < 300K$.

3. $N_{\text{ref}}(T) = N_{\text{ref}}(300K) \cdot (T/300K)^{\theta_2}$ is the doping concentration at which mobility reduces to almost half of its maximum value at low doping. At higher temperatures lattice scattering mainly dominates, and therefore, the contribution of ionised impurity scattering is expected to start at larger doping concentrations. This is also consistent with the positive temperature power for $N_{\text{ref}}(T)$ in (5.54) ($\theta_2 > 0$).

4. At high temperatures or low enough doping concentrations the temperature dependence of mobility is mainly determined by $\theta_1$, which dictates a reduction of mobility with increasing $T$. However, as temperature reduces, the term inside the bracket of the denominator in (5.54) starts to become important, even at relatively low doping concentration. Under this condition, the temperature power of $\mu_{\text{LF}}(T)$ becomes $\sim (\lambda \theta_2 - \theta_1)$. Consequently, mobility saturates at low $T$, and if $\lambda \theta_2 > \theta_1$ it finds a maximum, after which the mobility will be reduced with further reduction of temperature due to the dominance of ionised impurity scattering. This trend is exactly what the experimental Hall data for various semiconductors suggest. This property of the empirical function (5.54), together with other properties addressed above, can be clearly seen in Fig. 5.19 in which $\mu_{\text{LF}}(\text{In}_{0.53}\text{Ga}_{0.47}\text{As})$ is plotted as a function of temperature for three different doping concentrations.

5. The composition dependence of $\mu_{\text{LF}}$ is modelled by considering all the fitting parameters in (5.54) including $\mu_{\text{min}}$, $\mu_{\text{max}}(300K)$, $N_{\text{ref}}(300K)$, $\lambda$, $\theta_1$, and $\theta_2$ functions of composition $x$ and $y$. Simplest possible interpolation schemes (linear, power, or quadratic interpolations) are used to find the ternary and quaternary parameters in the entire range of composition. These forms can be later modified to more complex functions as more data for intermediate ternary and quaternary materials become available. Of particular attention in this case must be the contribution of alloy scattering in ternary and quaternary compounds.

Alloy scattering is shown to have a smoother variation with temperature than various mechanisms of lattice scattering (i.e., polar and non-polar optical phonon and acoustic phonon scatterings) (see, for instance, [172: pp.227-236]). Therefore, alloy scattering is expected to reduce both $\mu_{\text{max}}(300K)$ and $\theta_1$ compared to the case where alloy scattering is ignored. As will be seen in the forthcoming parts, both $\mu_{\text{max}}(300K)$ and $\theta_1$ will have a downward bowing (in the majority of materials) as the composition of ternary and quaternary

---

**Figure 5.19** Variation of low-field electron mobility in In$_{0.53}$Ga$_{0.47}$As with temperature with three different doping concentrations according to Eq. (5.54) and the parameters listed in Table 5.9.
5.4.2. Limitations of the Present Mobility Model

In the following, the limitations and the range of applicability of the above mobility model will be discussed:

1. The mobility data available in the literature are mainly measured using Hall effect technique and they give the majority carrier mobility versus majority carrier concentration. There are, however, instances where the minority carrier mobility would be of real concern, as in the base region of bipolar transistors. But, the techniques to measure the minority mobility are not straightforward, and consequently, their data are scarce and limited to only some well-studied semiconductors like Si, GaAs, and Al_xGa_{1-x}As. Therefore, in this work it has been assumed that the minority carrier mobility is the same as majority one. This assumption can cause some large errors, since the variation of minority mobility with both doping concentration and temperature are shown to be significantly different from that of majority mobility. In particular, the minority electron mobility data for p-GaAs are summarised in [246: section 14.3]. These data show that minority electron mobility is smaller than the majority one for N_A < 10^{19} cm^{-3}. But as the acceptor concentration increases above 10^{19} cm^{-3}, the minority electron mobility starts to increase. The same trend has recently been observed for p-AlGaAs [111]. Theoretical calculation by Bennett [263] has also shown a good agreement to the above trend for GaAs and AlGaAs.

2. The measured ‘Hall’ mobility (\mu_H) can be significantly different from the ‘drift’ mobility (\mu_{\text{drift}}) when transport is due to carriers in two or more subbands [246: p.565] [270]. The effective Hall factor determined as \tau_H = \mu_H / \mu_{\text{drift}} can be as large as 4 for p-type III-V compounds in which hole transport occurs in the heavy and light hole subbands. Also for materials like Al_xGa_{1-x}As and In_{1-x}Ga_xP, where a direct-to-indirect crossover may occur, the effective Hall factor for electrons increases near the crossover compositions [246: p.565]. However, since the materials used in the majority of device applications are normally direct bandgap materials sufficiently away from crossover composition, the effective Hall factor for electrons should not be of a prime concern. Additionally, in Npn HBTs the only p-type region is the base layer which is normally very highly doped. It is shown by Wenzel et al. [270] that the hole effective Hall factor approaches unity as the doping concentration increases. This makes the assumption \mu_H = \mu_{\text{drift}} valid for Npn HBT simulation. However, this assumption should be reconsidered when compositions near direct-to-indirect crossover and/or lowly doped p-type materials are used.

3. Dopant compensation is ignored in this work. Considering the compensation makes the mobility model very complicated. Tables of mobility values for some III-V materials like GaAs [271], and InP [272] are provided as a function of compensation ratio and total doping concentration at 77K and 300K. But, unavailability of a similar set of data for other materials and also the disagreement of these tables with available experimental data in
some ranges of doping concentration [273] stops us from finding a general relation for compensation dependent mobilities. Consequently, uncompensated materials will be assumed throughout this work and $N_D + N_A$ will be used in place of $N$ in Eq. (5.54).

4. Mobility is considered independent of the dopant species used. In fact, Anderson et al. [273] have measured the electron Hall mobilities for a large number of InP samples doped with Si, S, Sn, Se, and Ge to the levels between $10^{15}$ to $10^{19}$ cm$^{-3}$. Their results showed nearly identical electrical behaviour for all n-type dopants. On the other hand, the study of hole mobility in p-GaAs has revealed that C-doped GaAs has almost 20-30% higher mobility than Zn- or Be-doped GaAs with the same amount of hole concentration [191]. These strikingly different results to those of InP may be due to the different compensation ratio in C-doped GaAs compared to other p-type dopants.

5. It is also assumed that mobility is independent of the growth method. In the next subsection it will be shown that MOCVD grown n-In$_{0.49}$Ga$_{0.51}$P generally has higher electron mobility than MBE, GSMBE, and LPE samples. This is contradicting the present assumption, but again can be due to different compensation levels for various growth methods.

6. Effects like carrier-carrier scattering and surface scattering (very important for FET devices) are not considered.

7. The model is valid for relatively thick epitaxially grown materials. Effects like mobility enhancement due to modulation doping in MODFETs which form two-dimensional electron gas (2DEG) are not considered.

8. Although Eq. (5.54) generates a reasonable trend for variation of mobility at low temperatures (i.e., saturation of mobility and sometimes reduction of it with further reducing temperature), it is not particularly attempted to fit this equation to the available experimental data for $T < 100$-150K. This attempt may, for example, need a temperature dependent $\lambda$ in Eq. (5.54). Therefore, this mobility model should be used with caution at temperatures lower than 100-150K.

9. The high field mobility (to be formulated later in this chapter) is assumed to be only a function of local field. Also hot carrier effects like velocity overshoot and ballistic transport of carriers are not considered in the present formulation of carrier mobility.

The above assumptions may sometimes seem very harsh and the reader may conclude that the final product will be only of a very limited application. However, it should be considered that many of the above assumptions may be completely satisfied for a particular device, and those which are not, may act in opposing directions. Indeed, as is shown in the next subsection, Eq. (5.54) creates excellent fits to the available data for a wide range of III-V materials, temperature, and doping concentration. Additionally, it will be shown in chapter 7 that the mobility model developed in this section works very well for various types of HBTs with different material structure. To the best of author's knowledge, the present mobility model addresses the largest number of III-V compounds ever reported in the literature.
5.4.3. Fitting Parameters for Low-Field Mobility

Table 5.9 summarises the fitting parameters of Eq. (5.54) for various binary and ternary compounds obtained in this work. These parameters together with their variation with composition for ternary and quaternary compounds will be discussed individually in the forthcoming subsections.

**Table 5.9- Fitting parameters for the low-field mobility model given in Eq. (5.54).**

<table>
<thead>
<tr>
<th>Material</th>
<th>electron or hole</th>
<th>( \mu_{\text{max}}(300\text{K}) ) (cm²/V.s)</th>
<th>( \mu_{\text{min}}(300\text{K}) ) (cm²/V.s)</th>
<th>( N_{\text{ref}}(300\text{K}) ) (cm⁻³)</th>
<th>( \lambda )</th>
<th>( \theta_1 )</th>
<th>( \theta_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>AlAs electron</td>
<td>400</td>
<td>10</td>
<td>5.46x10¹⁷</td>
<td>1.00</td>
<td>2.1</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>hole</td>
<td>200</td>
<td>10</td>
<td>3.84x10¹⁷</td>
<td>0.488</td>
<td>2.24</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>GaAs electron</td>
<td>9400</td>
<td>500</td>
<td>6.0x10¹⁶</td>
<td>0.394</td>
<td>2.1</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>hole</td>
<td>491.5</td>
<td>20</td>
<td>1.48x10¹⁷</td>
<td>0.38</td>
<td>2.2</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>InAs electron</td>
<td>34000</td>
<td>1000</td>
<td>1.1x10¹⁸</td>
<td>0.32</td>
<td>1.57</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>hole</td>
<td>530</td>
<td>20</td>
<td>1.1x10¹⁷</td>
<td>0.46</td>
<td>2.3</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>InP electron</td>
<td>5200</td>
<td>400</td>
<td>3.0x10¹⁷</td>
<td>0.47</td>
<td>2.0</td>
<td>3.25</td>
<td></td>
</tr>
<tr>
<td>hole</td>
<td>170</td>
<td>10</td>
<td>4.87x10¹⁷</td>
<td>0.62</td>
<td>2.0</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>GaP electron</td>
<td>152</td>
<td>10</td>
<td>4.4x10¹⁸</td>
<td>0.80</td>
<td>1.60</td>
<td>0.71</td>
<td></td>
</tr>
<tr>
<td>hole</td>
<td>147</td>
<td>10</td>
<td>1.0x10¹⁸</td>
<td>0.85</td>
<td>1.98</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>Al₀.₃Ga₀.₇As hole</td>
<td>240</td>
<td>5</td>
<td>1.0x10¹⁷</td>
<td>0.324</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>In₀.₅₂Al₀.₄₈As electron</td>
<td>4800</td>
<td>800</td>
<td>3.0x10¹⁶</td>
<td>1.10</td>
<td>—</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>hole</td>
<td>14000</td>
<td>300</td>
<td>1.3x10¹⁷</td>
<td>0.48</td>
<td>1.59</td>
<td>3.68</td>
<td></td>
</tr>
<tr>
<td>In₀.₄₉Ga₀.₅₁P electron</td>
<td>4300</td>
<td>400</td>
<td>2.0x10¹⁶</td>
<td>0.70</td>
<td>1.66</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>hole</td>
<td>150</td>
<td>15</td>
<td>1.5x10¹⁷</td>
<td>0.80</td>
<td>2.0</td>
<td>—</td>
<td></td>
</tr>
</tbody>
</table>

5.4.3.a. GaAs

Electron and hole Hall mobilities of GaAs have been measured by many groups and a large number of data are available in the literature. References [191,246,274], in particular, include excellent compilations of the data and useful references for further studies. Additionally, measured data by various authors show a relatively good consistency in the carrier concentration range \( 1x10^{13}-2x10^{19} \) cm⁻³ for n-type and \( 1x10^{14}-2x10^{21} \) cm⁻³ for p-type GaAs. This allows us to find an accurate set of parameters for low-field mobilities in GaAs. Figures 5.20 and 5.21 show some of the available room temperature data for n- and p-type GaAs, respectively, together with the empirical fits made by the parameter sets given in Table 5.9.

As mentioned previously in section 5.4.2, the hole mobility of GaAs doped with C is almost 20-30 percent higher than that of Zn- or Be-doped GaAs [191]. The parameters obtained in this work best fit the data for C-doped GaAs at concentrations above \( 10^{18} \) cm⁻³.
The parameter $\theta_{n1}$ and $\theta_{p1}$ are found using the temperature dependence of mobility for relatively pure materials. It is generally observed that $\mu_n$(GaAs) drops sharply with increasing temperature ($\theta_{n1} = 2.59$) in the temperature range $70K < T < 300K$, but the fall-off becomes slower for $300K < T < 900K$ [172: p.225].

Values of $\theta_{n1}$ in the range 2.0-2.3 can be fitted to the data available in [191] [184: p.84] for the medium range of temperature. $\theta_{n1} = 2.1$ used in this thesis seems to produce very good fit to the available data for $100K \leq T \leq 400K$, and even above 400K the error is not large. $\theta_{p1}$ in the range 2.0-2.5 can be best fitted to the data in [172: p.235] [184: p.86] [191] [246: p.570], and is also mentioned by other authors [279: p.82] [280]. $\theta_{p1} = 2.2$ is chosen in this work for the temperature range $100K \leq T \leq 400K$.

In order to find $\theta_{p2}$, one needs to have the concentration dependence of mobility at temperatures other than 300K. Unfortunately, the only other temperature for which reliable concentration dependent data are available for majority of III-V materials is 77K, that is not lying within the range of validity of the mobility models in this work. However, assuming that the temperature dependence of $N_{\text{ref}}$ remains the same outside the range of validity of our mobility model, one can use $N_{\text{ref}}$(77K) to find an estimate of $\theta_{p2}$. Using the data compiled in [274] one finds $N_{n,\text{ref}}$(GaAs,77K) = $1.0\times10^{15}$ cm$^{-3}$, which results in $\theta_{n2} = 3.0$. This value is similar to the one obtained for silicon [281], and will be used as default throughout this work wherever reliable data is not available. Since a similar 77K data set is not available for $\mu_p$(GaAs), the default value of 3.0 will be used for $\theta_{p2}$.

![Figure 5.20: GaAs electron mobility data versus doping (carrier) concentration at room temperature, together with the empirical fitting obtained in this work.](image)

![Figure 5.21: GaAs hole mobility data versus doping (carrier) concentration at room temperature, together with the empirical fit obtained in this work. The experimental data are taken from numerous sources: [111,191,270,275-278].](image)
Chapter 5  

5.4.3.b. AlAs and Al$_x$Ga$_{1-x}$As

The parameters for the electron mobility in AlAs are chosen to give the best least-square fits to the data in [246: p.556], and the single data points in [177,275]. Due to the lack of experimental data, it is assumed that $\theta_{n1}$ and $\theta_{n2}$ in AlAs are the same as those in GaAs. As to the electron mobility in Al$_x$Ga$_{1-x}$As as a function of Al mole fraction, the situation is a bit more complicated. Since the electron effective mass of AlGaAs increases almost abruptly as the Al mole fraction, $x$, approaches the direct-to-indirect crossover point ($x \sim 0.4$), a sudden drop of electron mobility near the same composition level is expected theoretically (and observed experimentally, as the following results suggest). This form of mobility variation can be taken into account by a similar variation in $\mu_{n,\text{max}}(300K)$ and $\mu_{n,\text{min}}$. The formulation of the latter two parameters in this work follows that suggested by Sutherland and Hauser [282]. When multi-valley conduction takes place, an effective mobility can be defined as:

$$\mu_M(x) = \frac{n_{\Gamma}(x) \cdot \mu_D(x) + [n_X(x) + n_L(x)] \cdot \mu_I(x)}{n}$$  (5.55)

where $\mu_D(x)$ and $\mu_I(x)$ represent the effective electron mobility in the direct and indirect conduction valleys, respectively, and $\mu_M$ can stand for both $\mu_{n,\text{max}}(300K)$ and $\mu_{n,\text{min}}$ in Al$_x$Ga$_{1-x}$As. For the formulation of the direct band mobility, it is assumed that mobility of AlGaAs is dominated by polar optical phonon scattering [282], and therefore:

$$\mu_D(x) \propto \frac{1}{(m_{n\Gamma}(x))^{3/2} \cdot [\epsilon_h^{-1}(x) - \epsilon_i^{-1}(x)]}$$  (5.56)

Hence, using the mobility of (direct bandgap) GaAs as one end of this composition range, one can write:

$$\mu_D(Al_xGa_{1-x}As) = \mu_M(GaAs) \cdot \left[ m_{n\Gamma}(GaAs) \right]^{3/2} \cdot \frac{\epsilon_h^{-1}(GaAs) - \epsilon_i^{-1}(GaAs)}{\epsilon_h^{-1}(GaAs) - \epsilon_i^{-1}(GaAs)}$$  (5.57)

A similar formulation can be used to find $\mu_I(Al_xGa_{1-x}As)$, this time replacing the parameters of GaAs with those of (indirect bandgap) AlAs. But since the electron mobility at the X-valley for both AlAs (see Table 5.9) and GaAs (see below) are known, a linear interpolation between the two binary ends would be a reasonable choice. It is well-known that at pressures above $\sim$40kbar the X-valley of GaAs lies below the $\Gamma$-valley, and therefore, the electron mobility measured at pressures extending 40kbar would represent the X-valley mobility in this material. A plot of GaAs X-valley Hall mobility at 50kbar and room temperature is shown in [246: p. 555] as a function of impurity concentration. It can be seen from this figure that the variation of $\mu_{nX}(GaAs)$ with doping is very similar to that of $\mu_n(AlAs)$. Therefore, a linear interpolation between the values of the two binaries will result in:

$$\mu_I(Al_xGa_{1-x}As) = \mu_M(AlAs)$$  (5.58)
As mentioned previously, the above formulation can be used to find both $\mu_{n,max}(300K)$ and $\mu_{n,min}$. The two parameters $\lambda_n$ and $\theta_n$ are linearly interpolated between the corresponding values of the binaries. Since the range of variation for $N_{n,ref}$ is usually much wider than other parameters, a power interpolation of the following form may be used:

$$N_{n,ref}(Al_xGa_{1-x}As) = [N_{n,ref}(GaAs)]^{(1-x)} \cdot [N_{n,ref}(AlAs)]^x$$

(5.59)

To take the downward bowing of $\theta_{n1}(x)$ due to alloy scattering into account, an interpolation formula of the form:

$$\theta_{n1}(Al_xGa_{1-x}As) = \frac{(1-x) \cdot \theta_{n1}(GaAs) + x \cdot \theta_{n1}(AlAs)}{1 + m \cdot x \cdot (1-x)}$$

(5.60)

can be suggested. It is found that $m = 1$ gives a reasonable agreement to the temperature dependent electron mobility data for $x = 0.32, 0.36$ in [246: pp. 566-567].

As a confirmation of the above mobility formulation, the room temperature electron mobility of $Al_xGa_{1-x}As$ with $7 \times 10^{16}$ cm$^{-3}$ doping concentration is plotted in the entire composition range in Fig. 5.22. Also shown in this figure are the measured Hall data in the literature (summarised in [92]) for the doping concentration range $(5 \times 10^{16} - 1 \times 10^{17})$ cm$^{-3}$. An excellent agreement has been found for the entire range of composition, considering the fact that the doping concentration of the experimental data may slightly change above or below the value used for calculation in this work.

The parameters of $\mu_p(Al_{0.3}Ga_{0.7}As)$ in Table 5.9 are obtained from fitting to the room temperature mobility data in [246: p. 579]. To the best of author's knowledge no Hall mobility data is available for p-type AlAs in the literature. However, due to the similarities between the band structure of AlAs and GaP, it is expected that AlAs has a similar hole mobility to GaP (to be discussed later in this section). The room temperature hole mobility parameters of AlAs in this work are chosen such that quadratic interpolation between the parameters of GaAs, Al$_{0.3}$Ga$_{0.7}$As, and AlAs results in good agreement to the available data for $0 < x < 1$ (see Fig. 5.23). Figure 5.23 shows the available experimental hole mobilities of AlGaAs ($0 \leq x \leq 0.8$) for the doping ranges $(1.5-2.5) \times 10^{17}$ and $(1.5-2.5) \times 10^{18}$ cm$^{-3}$ [92: p.170] [283], together with the calculated values in this work for $N_A = 2 \times 10^{17}$ and $2 \times 10^{18}$ cm$^{-3}$, respectively. The agreement is very good. As mentioned
earlier, the room temperature parameters of \( \mu_p(AlGaAs) \), i.e., \( \mu_{p,\text{max}}(300\text{K}) \), \( \mu_{p,\text{min}} \), \( \lambda_p \), and \( \log_{10}(N_{p,\text{ref}}) \), are obtained by quadratic interpolations between the corresponding parameters of the GaAs, AlAs, and \( Al_{0.3}Ga_{0.7}As \). \( \theta_{p1}(AlAs) = 2.24 \) can be obtained from the calculated temperature dependent figure in [246: p.575] for \( 100\text{K} < T < 400\text{K} \), and \( \theta_{p1}(AlGaAs) \) can be found similar to \( \theta_{n1}(AlGaAs) \) (Eq. (5.60)) using \( m = 1 \). Due to the lack of experimental data, \( \theta_{p2}(AlGaAs) \) is assumed to be equal to \( \theta_{p2}(GaAs) \).

### 5.4.3.c. InP

Electron Hall mobility of InP has been measured and compiled by various authors [186,273, 275,284-286]. Anderson et al. [273] have extensively studied the electron mobility in numerous InP samples doped with Si, S, Se, Ge, or Sn, in a wide range of concentration (\( 10^{15}-5\times10^{19} \))\text{ cm}^{-3} at \( T = 77\text{K} \) and 300K. Their data for various dopants shows a very good consistency in the entire range of doping concentration, and therefore, it is the tabulated average of these data which forms the basis for empirical fitting in this work. The average of the 300K data in [273] together with other data in the literature and the empirical fit obtained in this work are shown in Fig. 5.24. Additionally, from fitting to the 77K data in [273] one can obtain \( N_{n,\text{ref}}(InP, 77\text{K}) = 3.6\times10^{15} \text{ cm}^{-3} \), which together with the room temperature value of this parameter results in \( \theta_{n2}(InP) = 3.25 \). Also \( \theta_{n1}(InP) = 2.0 \) (100K < \( T < 300\text{K} \)) is used in this work as suggested in [172: p.225] [184: p.175].

The hole mobility in InP has also received considerable attention in the literature. Figure 5.25 summarises the data from [172,177,184,186,270,287-288]. Again the empirical fit in this work shows a very good agreement to the above data. Values of \( \theta_{p1}(InP) \) in the range 1.9-2.2 can be found from the data in [172: p.235] [184: p.177] [186: p.87], and \( \theta_{p1}(InP) = 2.0 \) (100K < \( T < 300\text{K} \)) is adopted in this work. Also \( \theta_{p2}(InP) = 3.0 \) (default value) is assumed.

### 5.4.3.d. GaP

The electron and hole mobilities in GaP are measured as a function of both temperature and carrier concentration by Kao and Eknoyan [289], and these data are used in this work to obtain all the fitting parameters in Eq. (5.54) for GaP. The empirical fit explained by Eq. (5.54) together with the parameter sets given in Table 5.9 agree excellently to the measured \( \mu_n(GaP) \) and \( \mu_p(GaP) \) in a wide range of temperature and doping concentration, as shown in
Figs. 5.26 and 5.27. Also the room temperature variation of $\mu_p(GaP)$ with doping concentration in Fig. 5.27 agrees very well with the data from other resources [270,275] (not shown in this figure).

![Figure 5.24- Room temperature electron mobility data for InP from various sources, together with the empirical fit obtained in this work.](image)

![Figure 5.25- Room temperature hole mobility data for InP from various sources in the literature, together with the empirical fit obtained in this work.](image)

![Figure 5.26- The measured Hall electron mobility (symbols) from [289] and the empirical fit obtained in this work (solid lines) for GaP as a function of both temperature and doping concentration.](image)

![Figure 5.27- The measured Hall hole mobility (symbols) from [289] and the empirical fit obtained in this work (solid lines) for GaP as a function of both temperature and doping concentration.](image)

5.4.3. e. In$_{1-x}$Ga$_x$P

The room temperature electron mobility of In$_{0.49}$Ga$_{0.51}$P is measured by various groups [227, 259,290-308] and they are summarised in Fig. 5.28. As can be seen in this figure, the MOCVD grown samples clearly offer larger mobilities than their MBE, GSMBE, and LPE counterparts. Since the majority of the samples used in this work are grown by MOCVD, an empirical fit to the data for this method of growth is given in Table 5.9. From the temperature
dependent data in [294], 
\[ \theta_{n1}(\text{In}_{0.49}\text{Ga}_{0.51}\text{P}) = 1.66 \quad (150\text{K} < T < 300\text{K}) \] can be obtained.

\[ \mu_p(\text{In}_{0.49}\text{Ga}_{0.51}\text{P}) \] has also been reported by various authors [227,291,293-294, 296,298-299,308-310], and in this case samples with different growth techniques do not show any striking difference in their hole mobility. Room temperature parameters listed in Table 5.9 best fit to the above data. From the hole mobility data for \text{In}_{0.49}\text{Ga}_{0.51}\text{P} with \[ 5.14 \times 10^{17} \text{ cm}^{-3} \] hole concentration [294], a temperature dependence proportional to \( (300\text{K}/T)^{1.76} \) could be obtained for \( 150\text{K} < T < 300\text{K} \). But considering the fact that lower doped samples show sharper variation of mobility with temperature, a value of \( \theta_{p1}(\text{In}_{0.49}\text{Ga}_{0.51}\text{P}) = 2.0 \) seems appropriate for relatively low doped materials.

As to the parameters of \text{In}_{1-x}\text{Ga}_x\text{P}, a linear interpolation between the values of the binaries will be used for finding \( \theta_{n2} \) and \( \theta_{p2} \). The parameters \( \theta_{n1}, \theta_{p1}, \lambda_n, \lambda_p, \log_{10}(N_{n,\text{ref}}(300\text{K})), \log_{10}(N_{p,\text{ref}}(300\text{K})), \mu_{p,\text{min}} \), and \( \mu_{p,\text{max}}(300\text{K}) \) are found using a quadratic interpolation between the corresponding values for \text{InP}, \text{In}_{0.49}\text{Ga}_{0.51}\text{P}, and \text{GaP}. \mu_{n,\text{max}}(\text{In}_{1-x}\text{Ga}_x\text{P}, 300\text{K}) \) is calculated in a way similar to \text{Al}_x\text{Ga}_{1-x}\text{As}:

\[ \mu_n(x) = \frac{n_{\Gamma}(x) \cdot \mu_{\Gamma}(x) + [n_X(x) + n_L(x)] \cdot \mu_1(x)}{n} \]  

(5.61)

However, in this case \( \mu_D(x) \) is obtained by a linear interpolation between the maximum mobilities in the direct bandgap materials \text{InP} and \text{In}_{0.49}\text{Ga}_{0.51}\text{P}:

\[ \mu_D(x) = \left( 1 - \frac{x}{0.51} \right) \cdot \mu_{n,\text{max}}(\text{InP},300\text{K}) + \left( \frac{x}{0.51} \right) \cdot \mu_{n,\text{max}}(\text{In}_{0.49}\text{Ga}_{0.51}\text{P},300\text{K}) \]  

(5.62)

and a theoretical mobility variation is used for the indirect valley mobility, \( \mu_1 \), based on the assumption that electron mobility in \text{InGaP} is dominated by polar optical phonon scattering:

\[ \mu_1(x) = \mu_{n,\text{max}}(\text{GaP},300\text{K}) \cdot \left[ \frac{m_{n,\text{indir}}(\text{GaP})}{m_{n,\text{indir}}(x)} \right]^{3/2} \cdot \frac{\varepsilon_{h1}(\text{GaP}) - \varepsilon_{l1}(\text{GaP})}{\varepsilon_{h1}(x) - \varepsilon_{l1}(x)} \]  

(5.63)

where \( m_{n,\text{indir}}(\text{GaP}) = m_n(\text{GaP}) \) and \( m_{n,\text{indir}}(x) \) is an appropriately defined effective mass which takes into account the relative position of the indirect valleys X and L in \text{In}_{1-x}\text{Ga}_x\text{P} as a
function of x (see subsection 5.3.3.d for more details). A similar approach to the above is used to find $\mu_{n,\text{min}}(\text{In}_{1-x}\text{Ga}_x\text{P})$.

### 5.4.3.f. InAs

The room temperature electron and hole mobility parameters of InAs in Table 5.9 are chosen to give reasonable agreement to the data mainly in [184: p.153] and few others reported elsewhere [177,275,311]. The large value of $N_{n,\text{ref}} = 1.1 \times 10^{18}$ cm$^{-3}$ and the small value of $\lambda_n = 0.32$ are consistent with the large number of data shown in [184] which suggest that for $N_D < 4 \times 10^{17}$ cm$^{-3}$, $\mu_n$(InAs) does not change significantly with doping. $\theta_{n1}$(InAs) = 1.57 can reproduce the temperature dependent electron mobility data in [184: p.153] reasonably well in the temperature range $150K < T < 900K$. Also $\theta_{p1}$(InAs) = 2.3 is reported in [172: p.235] for $T > 250K$. Finally, due to the lack of data, the default value of 3.0 is allocated to both $\theta_{n2}$ and $\theta_{p2}$ in InAs.

### 5.4.3.g. In$_{1-x}$Ga$_x$As

First, the fitting parameters of the composition lattice-matched to InP are considered. The electron mobility parameters in In$_{0.53}$Ga$_{0.47}$As are fitted to the average of the low doping data in [186,192,217,284] and mainly to the high doping data in [312]. $\theta_{n1}$(In$_{0.53}$Ga$_{0.47}$As) = 1.59 (100K $< T < 300K$) can be fitted to the temperature dependent mobility data in [172: p.228]. A value of $N_{n,\text{ref}}$(In$_{0.53}$Ga$_{0.47}$As, 77K) = $8.7 \times 10^{14}$ cm$^{-3}$ can be obtained from fitting to the 77K Hall data mainly in [192,284], which results in $\theta_{n2}$(In$_{0.53}$Ga$_{0.47}$As) = 3.68. The room temperature hole mobility parameters of In$_{0.53}$Ga$_{0.47}$As are fitted to the rather limited data in [192,217]. $\theta_{p1} = \theta_{n1}$ is also assumed for In$_{0.53}$Ga$_{0.47}$As.

All the electron and hole mobility parameters of In$_{1-x}$Ga$_x$As are quadratically interpolated between the corresponding parameters of GaAs, InAs, and In$_{0.53}$Ga$_{0.47}$As, except for $\theta_{p2}$, $N_{n,\text{ref}}$(300K), and $N_{p,\text{ref}}$(300K). $\theta_{p2}$(In$_{1-x}$Ga$_x$As) = $\theta_{p2}$(GaAs) is assumed, and the logarithms of the other two parameters are quadratically interpolated between the known values for the three compositions. It is worth noting that the values of $\mu_{\text{max}}$(300K) and $\theta_1$ (both for electrons and holes) given in Table 5.9 automatically produce downward bowing for these parameters in the ternary compound In$_{1-x}$Ga$_x$As, taking the effects of alloy scattering into account.

### 5.4.3.h. In$_{1-x}$Al$_x$As

Recently Goto et al. [313] presented a large number of data for electron mobility of In$_{0.52}$Al$_{0.48}$As in samples with Hall electron concentrations in the range ($5 \times 10^{16} - 6 \times 10^{18}$) cm$^{-3}$. Although these data show some discrepancy due to the variation of growth parameters, they together with few other experimental and theoretical data available in [192: p.112] form the basis for the selection of room temperature parameters of $\mu_n$(In$_{0.52}$Al$_{0.48}$As) in Table 5.9.

As to the parameters for $\mu_n$(In$_{1-x}$Al$_x$As), the maximum mobility is found using a method similar to that used for In$_{1-x}$Ga$_x$P. However, due to the large difference in the electron
mobilities for InAs and In$_{0.52}$Al$_{0.48}$As, a linear interpolation between the mobilities in these two materials results in negative mobility in some range of composition. Therefore, a power interpolation of the following form is adopted:

$$\mu_D(IN_{1-x}Al_xAs) = (\mu_{n,max}(InAs,300K))^{(1-x/0.48)} \cdot (\mu_{n,max}(In_{0.52}Al_{0.48}As,300K))^{x/0.48} \quad (5.64)$$

A similar formulation is used for $\mu_{n,min}(IN_{1-x}Al_xAs)$. A quadratic interpolation is used for $\lambda_n$ and Log$_{10}(N_{n,ref}(300K))$ in this material, while $\theta_{n2}$ is linearly interpolated between the corresponding values for InAs and AlAs, and an interpolation scheme similar to (5.60) with $m = 1$ is used for $\theta_{n1}$.

No reliable data is available in the literature for $\mu_p(IN_{1-x}Al_xAs)$. Therefore, a linear interpolation is used for $\mu_p$, $\log_{10}(N_{p,ref}(300K))$, $\lambda_p$, and $\theta_{p2}$. To take the effect of alloy scattering into account, (5.60) may be used for both $\mu_{p,max}(300K)$ and $\theta_{p1}$.

5.4.3.i. $IN_{1-x}Ga_xAs_yP_{1-y}$

Majority of the electron and hole mobility data in the literature for the quaternary material $IN_{1-x}Ga_xAs_yP_{1-y}$ belong to the compositions lattice-matched to InP [172 (and references therein)] [287,314]. Due to the lack of knowledge about the Hall mobilities of GaAs$_yP_{1-y}$ and InAs$_yP_{1-y}$, only the mobility parameters of $IN_{1-x}Ga_xP$ and $IN_{1-x}Ga_xAs$ are used for interpolation schemes (see Eq. (5.7)). For instance, $\lambda_n$ is found using:

$$\lambda_n(IN_{1-x}Ga_xAs_yP_{1-y}) = y \cdot \lambda_n(IN_{1-x}Ga_xAs) + (1 - y) \cdot \lambda_n(IN_{1-x}Ga_xP) \quad (5.65)$$

A similar linear interpolation is used for $\theta_{n2}$, Log$_{10}(N_{n,ref}(300K))$, $\lambda_p$, $\mu_{p,min}$, $\theta_{p2}$, and Log$_{10}(N_{p,ref}(300K))$. $\mu_{n,max}(300K)$ is interpolated using:

$$\mu_{n,max}(IN_{1-x}Ga_xAs_yP_{1-y},300K) = \frac{y \cdot \mu_{n,max}(IN_{1-x}Ga_xAs,300K) + (1 - y) \cdot \mu_{n,max}(IN_{1-x}Ga_xP,300K)}{1 + m \cdot y \cdot (1 - y)} \quad (5.66)$$

with $m = 6$. Similarly, (5.66) is used for $\mu_{n,min}$ ($m = 6$), $\theta_{n1}$ ($m = 1$), $\mu_{p,max}(300K)$ ($m = 6$), and $\theta_{p1}$ ($m = 1$). The above values of $m$ are chosen to give a reasonable agreement to the electron and hole mobility data for InGaAsP lattice-matched to InP available in the literature (see below). For example, the electron mobility data of InGaAsP lattice-matched to InP for the doping range (1-5)$\times10^{16}$ cm$^{-3}$ [287,315] and for relatively pure samples [172] are shown in Fig. 5.29 as a function of arsenic mole fraction, $y$. Also plotted in this figure are the empirical variations obtained in this work for $N = 10^{15}$, $10^{16}$, and $5\times10^{16}$ cm$^{-3}$. It is clear from this figure that the empirical formula used in this work predicts the electron mobility data of InGaAsP quite well; the majority of measured data points are located in the area between the two empirical curves for $10^{15}$ and $5\times10^{16}$ cm$^{-3}$ doping concentrations. A similar comparison has been made in Fig. 5.30 for the hole mobility data in InGaAsP lattice-matched to InP. Again the empirical fits agree quite well with the data points summarised in [287,315].
Additionally, using $m = 1$ for $\theta_n^1(\text{In}_{1-x}\text{Ga}_x\text{As}_y\text{P}_{1-y})$ has resulted in a temperature dependence of mobility very similar to those reported in [314] for various compositions of InGaAsP lattice-matched to InP.

![Figure 5.29- Electron mobility data of InGaAsP lattice-matched to InP versus arsenic mole fraction for the doping concentrations (1-5)×10^{16} cm^{-3} in [287], for $N < 3\times10^{16}$ cm^{-3} in [315], and for relatively pure samples in [172]. The experimental Hall data are compared with the empirical curves found in this work.

![Figure 5.30- Hole mobility data of InGaAsP lattice-matched to InP versus arsenic mole fraction for the doping concentrations 4×10^{16} and 2×10^{18} cm^{-3} in [287,315] are compared against the empirical curves found in this work for almost similar doping densities.

**5.4.4. High-Field Mobility Model**

The velocity-field curves for GaAs and many other direct bandgap compound semiconductors show a reduction in velocity at high fields after passing through a maximum. The explanation for this is provided by the inter-valley charge transfer, i.e., the Gunn effect [316: Chap.11]. The lowest conduction band minimum for GaAs is at the zero momentum $\Gamma$-point, which is where the electrons are located when they are not accelerated. This conduction band minimum has the lowest effective mass and therefore, the highest electron mobility. The next conduction band minimum (L-point) is ~0.3eV higher in energy and has a much higher effective mass, indicating that electrons which are capable of entering this part of the conduction band will move more slowly in the same electric field. As the electric field accelerates the electrons in the $\Gamma$-band, they become hotter. This raises the probability of their excitation to the upper L-band where they slow down. At very high fields, the effect is so marked that the otherwise upward trend in velocity with field is overtaken by the inter-valley transfer effect (see Fig. 5.31).

Many authors have suggested various models to take the above field-dependent behaviour of electron mobility into account [317-320]. In this work, the following model from Kramer and Mircea [317] which has a minimum number of parameters is adopted:
where $E_{\text{crit}}$ is a critical electric field (almost equal to the electric field where velocity is maximum), $m > 1$ is a fitting exponent which determines the steepness of the negative differential mobility region between the peak velocity and saturation velocity, and $E_n$ is an effective electric field for electrons defined by:

$$E_n = -\frac{d\psi_n}{dz}$$

(5.68)

Table 5.10- Parameters of the high-field mobility model explained by Eq. (5.67) for some binary and ternary compounds ($T = 300 K$).

<table>
<thead>
<tr>
<th>Material</th>
<th>$E_{\text{crit}}$ (V/cm)</th>
<th>$v_{n,\text{sat}}$ (cm/s)</th>
<th>$m$</th>
<th>$v_{p,\text{sat}}$ (cm/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GaAs</td>
<td>$4 \times 10^3$</td>
<td>$1.0 \times 10^7$</td>
<td>2</td>
<td>$1.0 \times 10^7$</td>
</tr>
<tr>
<td>AlAs</td>
<td>$4 \times 10^3$</td>
<td>$0.8 \times 10^7$</td>
<td>2</td>
<td>$1.0 \times 10^7$</td>
</tr>
<tr>
<td>InP</td>
<td>$10 \times 10^3$</td>
<td>$1 \times 10^7$</td>
<td>3</td>
<td>$0.8 \times 10^7$</td>
</tr>
<tr>
<td>GaP</td>
<td>$10 \times 10^3$</td>
<td>$1 \times 10^7$</td>
<td>2</td>
<td>—</td>
</tr>
<tr>
<td>InAs</td>
<td>$2.2 \times 10^3$</td>
<td>$1 \times 10^7$</td>
<td>4</td>
<td>—</td>
</tr>
<tr>
<td>In$<em>{0.49}$Ga$</em>{0.51}$P</td>
<td>$10 \times 10^3$</td>
<td>$0.8 \times 10^7$</td>
<td>—</td>
<td>$0.8 \times 10^7$</td>
</tr>
<tr>
<td>In$<em>{0.53}$Ga$</em>{0.47}$As</td>
<td>$4 \times 10^3$</td>
<td>$0.85 \times 10^7$</td>
<td>—</td>
<td>$0.6 \times 10^7$</td>
</tr>
<tr>
<td>In$<em>{0.52}$Al$</em>{0.48}$As</td>
<td>$8 \times 10^3$</td>
<td>$0.7 \times 10^7$</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Values of the parameters used in Eq. (5.67) for some of the binary and ternary materials are listed in Table 5.10. These values are taken from the graphs and data in [172, 178, 183-184, 187, 192, 318-325]. In choosing the saturation velocity, $v_{n,\text{sat}}$, from the data in the above references, the largest of all the reported data is selected. This choice has several reasons:
1. (5.67) is a steady-state velocity-field relationship; it does not take the velocity overshoot of electrons into account. As a result, (5.67) largely underestimates the average velocity of electrons inside the base and collector regions of HBTs when they transport from emitter to collector, and results in a calculated forward transit time which is much larger than the measured data (see chapters 7 and 10 for more details). The above choice of the saturation velocity is meant to partly compensate this shortcoming, although the experimentally observed average velocities of electrons due to the velocity overshoot are still significantly larger than the steady-state values of $v_{n,sat}$ given in Table 5.10 (see [326], for instance).

2. The velocity-field characteristics of minority electrons are observed to be largely different from those of majority electrons. For example, Degani et al. [327] reported the steady-state v-E characteristics for photoexcited minority electrons in p-In$_{0.53}$Ga$_{0.47}$As. They measured a steadily increasing velocity up to the value of $2.6 \times 10^7$ cm/s at 7.5 kV/cm electric field, with no evidence of transferred electron effects. They suggested that this could result from energy exchanging electron-hole collisions. In this way the minority electron system cannot be heated to the extent that would occur in n-type material. Thus, the fraction of carriers which can be scattered to the satellite valley is reduced, and the drift velocity is limited by saturation process in the Γ-valley alone. In principle this leads to an enhancement of electron drift velocity inside the base region of HBTs which reduces $\tau_B$ (see Eq. (10.4)).

3. Some measurement techniques which have resulted in low values of saturation velocities are based on assumptions which are not physically justifiable. For instance, Liu et al. [324] have obtained the electron saturation velocity of In$_{0.49}$Ga$_{0.51}$P by measuring the cutoff frequency of an InGaP/GaAs DHBT under various bias conditions. They calculated a low saturation velocity of $4.4 \times 10^6$ cm/s for In$_{0.49}$Ga$_{0.51}$P. However, their delay time analysis did not consider the carrier blocking effect of the B/C heterojunction in DHBTs which results in a significant enhancement of the forward transit time in DHBTs. For more detailed discussion on this issue see Chapter 11.

For the high-field mobility of holes the formulation is much simpler, since the inter-valley charge transfer does not occur inside the valence band. The high-field hole mobility can be expressed as:

$$\mu_p = \frac{\mu_{p,LF}}{1 + \frac{\mu_{p,LF} |E_p|}{v_{p,sat}}}$$

which has only one parameter, $v_{p,sat}$, also listed in Table 5.10. The effective field for holes, $E_p$, can be expressed similar to $E_n$ as:

$$E_p = -d\psi_p / dz$$

However, there is an argument that when the current is flowing opposite to the electric field
driving force (as in the space-charge region of the B-E junction), the choice of the \( E_n \) and \( E_p \) given by Eqs. (5.68) and (5.70) as the driving force for velocity saturation is questionable [56,281,328]. This argument is discussed in more details in section 7.1.

Parameters of the high-field mobility for ternary compounds can be found by linear or quadratic interpolation between the known values for these parameters in that material system. For example, \( E_{\text{crit}} \) and \( v_{n,\text{sat}} \) for In\(1-x\)Ga\(_x\)P can be found by a quadratic interpolation between the corresponding values for In\(0.49\)Ga\(_{0.51}\)P, GaP, and InP, while \( m \) in the same material system is obtained by a linear fit between the values for the binaries. \( v_{p,\text{sat}} \) in In\(_{1-x}\)Ga\(_x\)As is an exception, since a linear interpolation between the available values for GaAs and In\(_{0.53}\)Ga\(_{0.47}\)As results in an extremely low value of \( v_{p,\text{sat}}(\text{InAs}) \). Therefore, for this material system, \( v_{p,\text{sat}} \) is linearly interpolated between the corresponding values for GaAs and In\(_{0.53}\)Ga\(_{0.47}\)As only for \( 0.47 \leq x \leq 1 \). For gallium mole fractions less than 0.47, \( v_{p,\text{sat}} \) is assumed to be constant, i.e., \( v_{p,\text{sat}}(\text{InAs}) = v_{p,\text{sat}}(\text{In}_0.53\text{Ga}_0.47\text{As}) \). As to the parameters for the quaternary material InGaAsP, a linear fit between the values for InGaAs and InGaP similar to Eq. (5.7) is used.

Electron and hole saturation velocities are generally observed to reduce at higher temperatures [172,184,319,321]. A linear expression is suggested for the variation of saturation velocities with temperature:

\[
v_{n(p),\text{sat}}(T) = v_{n(p),\text{sat}}(0K) \cdot (1 - a_n(p)T)
\]

(5.71)

with \( a_n(\text{GaAs}) = 1.17 \times 10^{-3} \text{ K}^{-1} \), \( a_n(\text{InP}) = 0.71 \times 10^{-3} \text{ K}^{-1} \), \( a_n(\text{In}_0.53\text{Ga}_0.47\text{As}) = 0.69 \times 10^{-3} \text{ K}^{-1} \), \( a_n(\text{InGaAsP}) = 0.48 \times 10^{-3} \text{ K}^{-1} \) suggested in [172,329]. In this work, a unique value for \( a_n \) is used (\( a_n = 0.7 \times 10^{-3} \text{ K}^{-1} \)) for all the materials. Using this value of \( a_n \), (5.71) can be rewritten as:

\[
v_{n,\text{sat}}(T) = v_{n,\text{sat}}(300K) \cdot (1.266 - 8.86 \times 10^{-4} T)
\]

(5.72)

Similarly, using only the \( T > 150K \) part of the graph in [184: p.90] for GaAs, one can write:

\[
v_{p,\text{sat}}(T) = v_{p,\text{sat}}(300K) \cdot (1.82 - 2.73 \times 10^{-3} T)
\]

(5.73)

Due to the insufficient experimental data for other materials, (5.73) is used for all the III-V compounds in this work.

The electron and hole velocities of various undoped materials at room temperature are plotted as a function of effective electric field in Figs. 5.31 and 5.32, respectively. As can be seen in Fig. 5.31, direct bandgap materials show a maximum electron velocity at \( E_n = E_{\text{crit}} \). But indirect bandgap materials such as GaP do not have a negative differential mobility region. In fact, it is shown in Fig. 5.33 that as the low-field mobility of GaAs decreases by increasing the doping concentration, the peak in the velocity-field characteristics disappears. The velocity-field characteristics shown in Figs. 5.31-5.33 are in general agreement to those observed experimentally (see, for instance, [172,184]) or calculated theoretically (see, for
5.5. Recombination and Generation Mechanisms

When a semiconductor is perturbed from the equilibrium state there is typically an attendant modification in the carrier densities inside the semiconductor. Recombination-generation (R-G) is nature’s order-restoring mechanism, the means whereby the carrier excess or deficit inside the semiconductor is stabilised (if the perturbation is maintained) or eliminated (if the perturbation is removed). Since the non-equilibrium conditions prevail during device operation, R-G plays a major role in shaping the characteristics exhibited by a device. This section deals with the most important R-G mechanisms for the bipolar transistors and theoretical or empirical methods to model them.

5.5.1. Band-to-Band (Radiative) Recombination-Generation

Band-to-band recombination, also referred to as radiative recombination, is conceptually the simplest of all recombination processes. It involves the direct annihilation of a conduction band electron and a valence band hole, the electron falling from an allowed conduction band state into a vacant valence band state. The process is typically radiative, with the excess energy released during the process going into the production of a photon. The above recombination process can be reversed to generate carriers. The band-to-band mechanism for generation-recombination is important for direct bandgap semiconductors. In Si, Ge, and GaP band-to-band R-G is insignificant for all imaginable conditions.

The net radiative recombination rate can be given by the following expression [170]:

$$R_{\text{rad}} = B_{\text{rad}} (n_0 - n_p^2)$$

(5.74)

where $B_{\text{rad}}$ is the radiative constant. $B_{\text{rad}}$ is a function of temperature and band parameters of the material. For direct bandgap materials it is shown that [330]:

$$B_{\text{rad}} \propto \left[ \frac{(m_{hh}/(m_n + m_{lh}))^{3/2} + (m_{lh}/(m_n + m_{lh}))^{3/2}}{m_n \cdot (m_{hh}^{3/2} + m_{lh}^{3/2})} \right] \cdot E_g^2 \cdot T^{-3/2}$$

Since the term inside the square bracket in the above relation is a slowly varying function of composition, this relation is rewritten as:

$$B_{\text{rad}}(x,y,T) = B_{\text{rad}}(\text{GaAs,300K}) \cdot f_1(x,y) \cdot (300/T)^{3/2}$$
where \( f_i(x,y) \) is a slowly varying function of composition. Using the effective masses and bandgaps of the binaries as given in Table 5.1, \( f_i(x,y) \) can be evaluated as 1, 0.21, 0.60, 1.39, and 0.75 for GaAs, InAs, InP, GaP, and AlAs, respectively. However as mentioned earlier, the indirect bandgap materials are expected to have negligible direct band-to-band recombination. Zarem et al. [331] observed a sharp increase of the effective lifetime in \( Al_xGa_{1-x}As \) near the direct-to-indirect crossover composition and attributed this to a decrease in the relative population of the direct valley. Parallel to the approach used in [331], the effective radiative constant in this work is scaled by the factor \( \left( \frac{n_n}{n} \right) \), which results in:

\[
B_{\text{rad}}(x,y,T) = B_{\text{rad}}(\text{GaAs,300K}) \cdot f_i(x,y) \cdot \exp \left( \frac{E_g - E_g^r}{kT} \right) \left( \frac{300}{T} \cdot \frac{m_{n}^r}{m_n} \right)^{3/2} \tag{5.75}
\]

An expression similar to (5.75) has recently been used to explain the effects of Al incorporation in \( (Al_xGa_{1-x})_{0.51}In_{0.49}P \) light emitters [332]. The value of \( B_{\text{rad}}(\text{GaAs,300K}) \) and the interpolation scheme for finding \( f_i(x,y) \) in ternary and quaternary materials is discussed in subsection 5.5.4. It is worth mentioning that a reduction of the radiative constant at high levels of injection is observed in semiconductor lasers [333]. However, this effect is ignored in the present work, since high-level injection does not occur inside the base nor the emitter of HBTs [334].

5.5.2. Auger Recombination

In the radiative recombination mechanism, as discussed above, the energy released during the annihilation of an electron-hole pair is producing a photon. In the Auger mechanism this energy is transferred to another electron (or hole), which gets excited to a higher energy state in the band. This electron or hole then relaxes back to achieve thermal equilibrium by losing its energy to lattice vibrations or phonons. For instance, the energy released during recombination may excite an electron to a higher state in the conduction band, may send a hole in the heavy-hole band to the light-hole band, or may transfer a hole in the heavy-hole band to the split-off valence band. The above processes are conventionally labelled CHCC, CHHL, and CHHS, respectively, where C, H, L, and S stand for conduction, heavy-hole, light-hole, and split-off bands, respectively [287]. Calculations have shown that CHCC and CHHS are the dominant processes for n- and p-type semiconductors, respectively [287]. The first process is proportional to \( n^2p \), while the second is proportional to \( np^2 \). The Auger R-G rate can thus be written as:

\[
R_{\text{Aug}} = (C_{n,\text{Aug}} \cdot n + C_{p,\text{Aug}} \cdot p)(np - n_n^2 \gamma_n \gamma_p) \tag{5.76}
\]

where \( C_{n,\text{Aug}} \) and \( C_{p,\text{Aug}} \) are called the electron and hole Auger coefficients.

Generally speaking, the Auger coefficients are larger for narrower bandgap materials and at higher temperatures [287]. In the direct type of Auger processes the conservation laws of energy and momentum should be both satisfied for the system of free particles involved in the
process. This leads to a temperature and bandgap dependence of the form
\[ C_{n(p),\text{Aug}} \propto \exp(-\delta \cdot E_g / kT), \]
where \( \delta \) depends on the effective masses of the material [287].

This direct Auger process becomes highly improbable for indirect bandgap materials due to
the large momentum difference between the direct and indirect valleys. Instead, another
mechanism named phonon-assisted Auger process becomes dominant in indirect bandgap
materials, where conservation of momentum is satisfied through phonon participation.
Phonon-assisted Auger process is also possible for direct bandgap materials, and it becomes
the dominant Auger mechanism for larger bandgaps and at lower temperatures [287,335]. The
temperature and bandgap dependence of this process is shown [336] to be much weaker than
the direct Auger process \( (C_{n(p),\text{Aug}} \propto T \cdot E_g^{-3/2}) \). Therefore, in the present work the Auger
coefficients are expressed as:

\[
C_{p,\text{Aug}} = C_2 \cdot f_2(x,y) \cdot \exp\left(\frac{-g(x,y)E_g}{NkT}\right) \tag{5.77a}
\]
\[
C_{n,\text{Aug}} = C_3 \cdot f_3(x,y) \cdot \exp\left(\frac{-g(x,y)E_g}{NkT}\right) \tag{5.77b}
\]

where \( C_2, C_3, \) and \( N \) are constants, and \( f_2(x,y), f_3(x,y), \) and \( g(x,y) \) are slowly varying functions
of composition. The different temperature dependence for the two Auger mechanisms is taken
into account by choosing a smaller \( g \) for indirect bandgap materials. \( g \) is assumed to be equal
to unity for GaAs, InAs, and InP and equal to 0.6 for GaP and AlAs. Values of the constants
in Eqs. (5.77) and variation of \( f_2, f_3, \) and \( g \) with composition are discussed in subsection 5.5.4.

### 5.5.3. Indirect Thermal Recombination-Generation

Certain impurity atoms can introduce allowed energy levels \( (E_T) \) into the midgap region of a
semiconductor. Crystal defects, particularly defects decorated with impurity atoms, can also
give rise to deep-level states. The R-G centres thereby created act as intermediaries in the
recombination process. First, one type of carrier and then the other type is attracted to the R-G
centre. The capture of an electron and a hole at the same site leads to the annihilation of the
electron-hole pair. Alternatively, the process may be described in terms of the state-to-state
transition of a single carrier: a carrier is first captured at the R-G site and then makes an
annihilating transition to the opposite carrier band. Indirect thermal recombination, also called
SRH (Shockley, Read, Hall\(^\dagger\)) recombination, is characteristically non-radiative. Thermal
energy is released during this process, or equivalently, lattice vibrations (phonons) are
produced. In general, SRH recombination exists at any region within the device under any
working condition, and thus, is the most common mechanism of recombination-generation.
Again the SRH generation has a mechanism that is exactly reverse to that of SRH
recombination.

\(^\dagger\) W. Shockley and W.T. Read, Jr., and independently R.N. Hall, were the first to model and investigate this
process.
The SRH recombination rate due to the recombination centres with a single energy level in the bandgap is expressed as [161]:

\[
R_{\text{SRH}} = \frac{n_p - n_i^2 \gamma_n \gamma_p}{\tau_p,\text{SRH}(n + n_i) + \tau_n,\text{SRH}(p + p_i)}
\]  

(5.78)

where \( \tau_p,\text{SRH} \) and \( \tau_n,\text{SRH} \) are the lifetimes for holes and electrons as minority carriers, respectively. \( n_i \) and \( p_i \) are the carrier concentrations when the Fermi level is located at the recombination centre energy level \( (E_T) \). The recombination rate approaches a maximum as the energy level of the recombination centre approaches midgap [161]. Thus the most effective recombination centres are those located near the middle of the bandgap, and in this case one can easily prove that:

\[
n_t = n_i \sqrt{N_C / N_V}
\]

(5.79)

\[
p_t = n_i \sqrt{N_V / N_C}
\]

The minority carrier lifetimes have a strong dependence on the total doping density, i.e., the more impurity is added to the semiconductor crystal, the more defects are introduced [279]. The following relations are often used in the device simulation programs to model the doping dependence of carrier lifetimes [17,170,279]:

\[
\tau_n,\text{SRH} = \frac{\tau_{n0,\text{SRH}}}{1 + \frac{N_D + N_A}{N_n,\text{SRH}}}
\]

(5.80a)

\[
\tau_p,\text{SRH} = \frac{\tau_{p0,\text{SRH}}}{1 + \frac{N_D + N_A}{N_p,\text{SRH}}}
\]

(5.80b)

The values of the parameters \( \tau_{n(p)0,\text{SRH}} \) and \( N_{n(p),\text{SRH}} \) are given for various materials in subsection 5.5.4. In general, \( \tau_{n(p),\text{SRH}} \) are very much dependent on the quality of the material and the growth method/parameters, and it is almost impossible to assign a unique value to the above SRH lifetime parameters. But, usually materials that benefit from a mature growth technology (such as GaAs and Si) and are free from natural defects such as DX centres have larger values of \( \tau_{n(p)0,\text{SRH}} \) and \( N_{n(p),\text{SRH}} \).

Considering all three R-G mechanisms discussed above, the net recombination rate can be written as:

\[
R = R_{\text{rad}} + R_{\text{Aug}} + R_{\text{SRH}}
\]

(5.81)

Sometimes the parameters of various mechanisms of R-G are obtained by curve fitting to the data for "effective" lifetime versus doping density [111,337-339]. In order to elaborate this technique, consider an n-type material under low-level injection and full ionisation of donors.
conditions \( (p_0 \ll \Delta n = \Delta p \ll n_0, \quad n = n_0 + \Delta n = n_0 = N_D \rightarrow p = p_0 + \Delta p = \Delta p \) and \( np >> n_0 p_0 = n_0^2 \gamma_n \gamma_p \)). Under these conditions, the recombination rate equations can be simplified to:

\[
R_{\text{rad}} \approx B_{\text{rad}} N_D \Delta p
\]

\[
R_{\text{Aug}} = C_{n, \text{Aug}} N_D^2 \Delta p
\]

\[
R_{\text{SRH}} = \frac{\Delta p}{\tau_{p, \text{SRH}}}
\]

Then defining the effective lifetime of minority carriers as:

\[
\tau_{p, \text{eff}} \equiv \frac{\Delta p}{R}
\]  

(5.82)

\( \tau_{p, \text{eff}} \) will be written as:

\[
\frac{1}{\tau_{p, \text{eff}}} = \frac{1}{\tau_{p, \text{SRH}}} + B_{\text{rad}} N_D + C_{n, \text{Aug}} N_D^2 = \frac{1}{\tau_{p,0, \text{SRH}}} + \left( \frac{1}{\tau_{p,0, \text{SRH}}} N_D + C_{n, \text{Aug}} N_D^2 \right) N_D^{\frac{1}{2}} + C_{n, \text{Aug}} N_D^2
\]

(5.83)

which is a parabolic variation of \( (\tau_{p, \text{eff}})^{-1} \) with \( N_D \). The fact that the doping dependence of SRH lifetime leads to an additional linear term in (5.83) needs a particular attention. From the fitting to the experimental data point of view, one can assume a constant SRH lifetime with doping and attribute all the linear term of \( (\tau_{p, \text{eff}})^{-1} \) versus \( N_D \) to the radiative constant, \( B_{\text{rad}} \). However, this latter assumption leads to an unphysical behaviour of \( B_{\text{rad}} \) in some materials. For instance, as will be shown in the next subsection, \( B_{\text{rad}} \) and \( C_{n, \text{Aug}} \) obtained from (5.75) and (5.77) in Al\(_{1-x}\)Ga\(_x\)As \((x < 0.35)\) are both smaller than those in GaAs, but the measured lifetimes of AlGaAs are much shorter than those of GaAs (see e.g. \[92,111\]). The short lifetimes in AlGaAs are known to be due to the existence of a considerable amount of defects (and perhaps DX centres) \[92\], and thus, it is unphysical to attribute it to a larger \( B_{\text{rad}} \) in this material. As shown in subsection 5.5.4, the observed short lifetimes in Al\(_{1-x}\)Ga\(_x\)As are modelled using a sharply decreasing \( \tau_{p(0), \text{SRH}} \) and \( N_p(0), \text{SRH} \) with Al mole fraction, \( x \).

### 5.5.4. Fitting Parameters for Recombination Lifetime

Only materials such as GaAs, In\(_{0.53}\)Ga\(_{0.47}\)As, and In\(_{0.72}\)Ga\(_{0.28}\)As\(_{0.60}\)P\(_{0.40}\) (used for \( \lambda = 1.3 \mu \text{m} \) lasers) are extensively studied in terms of their recombination parameters. Even a well-studied material such as InP is lacking data for recombination parameters. In the particular case of InP, lack of a confining layer lattice-matched to InP in a double heterostructure (DH) system is responsible for the scarcity of lifetime data \[186\], since majority of the techniques for characterising recombination parameters are relying on a DH system with the material under study as active layer \[287\]. Calculation of \( B_{\text{rad}} \) and \( C_{n(p), \text{Aug}} \) has also been carried out for several materials \[287,335,340\], but the results in majority of cases have shown a significant
discrepancy from available experimental data.

<table>
<thead>
<tr>
<th>Material</th>
<th>$f_1$</th>
<th>$f_2$</th>
<th>$f_3$</th>
<th>$g$</th>
<th>$\tau_{n,SRH}$ (µs)</th>
<th>$\tau_{p,SRH}$ (µs)</th>
<th>$N_{n,SRH}$ (cm$^{-3}$)</th>
<th>$N_{p,SRH}$ (cm$^{-3}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GaAs</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>10</td>
<td>5</td>
<td>2×10$^{16}$</td>
<td>2×10$^{16}$</td>
</tr>
<tr>
<td>AlAs</td>
<td>0.75</td>
<td>0.30</td>
<td>0.30</td>
<td>0.6</td>
<td>0.01</td>
<td>0.02</td>
<td>1×10$^{16}$</td>
<td>1×10$^{16}$</td>
</tr>
<tr>
<td>InAs</td>
<td>0.21</td>
<td>7.14</td>
<td>9.90</td>
<td>1.0</td>
<td>0.5</td>
<td>0.5</td>
<td>2×10$^{16}$</td>
<td>2×10$^{16}$</td>
</tr>
<tr>
<td>InP</td>
<td>0.60</td>
<td>0.27</td>
<td>0.27</td>
<td>1.0</td>
<td>0.3</td>
<td>5</td>
<td>1×10$^{15}$</td>
<td>3×10$^{15}$</td>
</tr>
<tr>
<td>GaP</td>
<td>1.39</td>
<td>0.30</td>
<td>0.30</td>
<td>0.6</td>
<td>0.2</td>
<td>0.2</td>
<td>1×10$^{16}$</td>
<td>1×10$^{16}$</td>
</tr>
<tr>
<td>In$<em>{0.53}$Ga$</em>{0.47}$As</td>
<td>0.96</td>
<td>4.26</td>
<td>5.72</td>
<td>—</td>
<td>10</td>
<td>10</td>
<td>2.3×10$^{15}$</td>
<td>2.3×10$^{15}$</td>
</tr>
<tr>
<td>In$<em>{0.49}$Ga$</em>{0.51}$P</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.3</td>
<td>0.5</td>
<td>2.5×10$^{16}$</td>
<td>1.5×10$^{16}$</td>
</tr>
<tr>
<td>Al$<em>{0.2}$Ga$</em>{0.8}$As</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.01</td>
<td>0.02</td>
<td>1×10$^{16}$</td>
<td>1×10$^{16}$</td>
</tr>
</tbody>
</table>

We start the parameter fitting by finding an appropriate $B_{rad}$(GaAs) and $f_1(x,y)$, and then testing if (5.75) generates reasonable data for other materials and as a function of temperature. $B_{rad}$(GaAs,300K) is experimentally observed to be in the range (0.67-2.0)×10$^{-10}$ cm$^3$/s [28,191,337-338,341-343]. A value of 1.0×10$^{-10}$ cm$^3$/s adopted in the present work seems to agree very well with the recent experimental lifetime data. $f_1(x,y)$ was also calculated based on the known values of band parameters for the binaries (see subsection 5.5.1), and these are summarised together with other fitting recombination parameters in Table 5.11. The value of $B_{rad}$(InAs,300K) obtained in this way (2.1×10$^{-11}$ cm$^3$/s) is the same as the one given in [177]. No experimental data for $B_{rad}$(InP,300K) is reported in the literature, but since the band parameters of InP and GaAs are almost similar, one expects to observe almost similar values for $B_{rad}$ in these two materials [341]. The values of $B_{rad}$ in AlAs and GaP, as expected, will be extremely small (<10$^{-21}$ cm$^3$/s). The value of $f_1(x,y)$ in In$_{1-x}$Al$_x$As, Al$_x$Ga$_{1-x}$As, and In$_{1-x}$Ga$_x$P is obtained by linear interpolation between those of the binaries. $B_{rad}$(In$_{0.49}$Ga$_{0.51}$P,300K) = 1.0×10$^{-10}$ cm$^3$/s thus obtained is in excellent agreement to (1.0±0.3)×10$^{-10}$ measured by Strauss et al. [330]. $B_{rad}$(In$_{0.53}$Ga$_{0.47}$As, 300K) is reported by various authors in the range (0.4-1.43)×10$^{-10}$ cm$^3$/s [218,339,344-345]. Zielinski et al. [218] have measured the radiative constant of In$_{0.53}$Ga$_{0.47}$As in the temperature range 77 ≤ T ≤ 340K, and it is their value at room temperature that is used in the present work. Moreover, the temperature dependent radiative constant of In$_{0.53}$Ga$_{0.47}$As calculated using (5.75) is in very good agreement to those measured in [218]. Recently Ahrenkiesel et al. [339] fitted an equation of the form (5.83) to the variation of experimental ($\tau_{p(n),eff}^{-1}$) data in In$_{0.53}$Ga$_{0.47}$As with $N_{D(A)}$. They observed that lifetimes of both electrons and holes are comparable and their variation can be explained by $C_{n,Aug} = C_{p,Aug} = 8.1×10^{-29}$ cm$^6$/s, and $B_{eff} = 1.43×10^{-10}$ cm$^3$/s as the linear coefficient in Eq. (5.83). This value of $B_{eff}$ is adopted in the present work and later $N_{n(p),SRH}$ are adjusted such that the doping dependence of the SRH
lifetime compensates the difference between \( B_{\text{rad}} = 0.96 \times 10^{-10} \text{ cm}^3/\text{s} \) and the above \( B_{\text{eff}} \). Finally, \( f_1(\text{In}_{1-x}\text{Ga}_x\text{As}) \) can be quadratically interpolated between the values for GaAs, InAs, and \( \text{In}_{0.53}\text{Ga}_{0.47}\text{As} \). As to the value of \( f_1 \) in \( \text{InGaAsP} \), the following interpolation may be used:

\[
f_1(\text{In}_{1-x}\text{Ga}_x\text{As}_y\text{P}_{1-y}) = y \cdot f_1(\text{In}_{1-x}\text{Ga}_x\text{As}) + (1-y) \cdot f_1(\text{In}_{1-x}\text{Ga}_x\text{P}) \quad (5.84)
\]

Table 5.12 compares the recombination coefficients of the above mentioned materials with those reported in the literature at \( T = 300\text{K} \). The first step towards finding appropriate parameters for the Auger coefficients in Eq. (5.77) is defining a reasonable value for \( N \).

Figure 5.34 shows \( C_p,\text{Aug}(\text{GaAs}) \) as a function of temperature. The measured data points are mainly from [340], with few others from [28,337-338,343]. Also shown in this figure are the calculated Auger coefficients of Bardyszewski and Yevick [335], which seem to have reasonable agreement with experimental data. \( C_p,\text{Aug}(\text{GaAs},300\text{K}) \) is reported in the range \((1-20) \times 10^{-30} \text{ cm}^6/\text{s} \). In this work, a room temperature value of \( 3 \times 10^{-30} \text{ cm}^6/\text{s} \) is adopted. Additionally, it is shown in Fig. 5.34 that using \( N = 25 \) results in a reasonable agreement to the experimental/calculated Auger coefficients in the temperature range 100-700K. Assuming \( f_2(\text{GaAs}) = 1.0 \), results in a fitting value of \( C_2 = 2.70 \times 10^{-29} \text{ cm}^6/\text{s} \). Also, as mentioned earlier, \( g(x,y) \) assumes values of unity and 0.6 for direct and indirect bandgap binaries, respectively, to take the different temperature sensitivity of the direct and phonon-assisted Auger processes into account. \( g, f_2, \) and \( f_3 \) for ternary and quaternary compounds are assumed to be linearly interpolated between corresponding values of the binaries. Therefore, the only remaining task to fully characterise \( C_p,\text{Aug} \) in Eq. (5.77a) is to find \( f_2(x,y) \) for \( \text{InAs}, \text{InP}, \text{GaP}, \) and \( \text{AlAs} \).

The experimental Auger coefficients of \( \text{In}_{0.53}\text{Ga}_{0.47}\text{As} \) reported in various works [192,339, 344-346] show a large discrepancy. A recent work by Ahrenkiel et al. [339] suggests that for \( \text{In}_{0.53}\text{Ga}_{0.47}\text{As} \) at 300K, \( C_n,\text{Aug} = C_p,\text{Aug} = 8.1 \times 10^{-29} \text{ cm}^6/\text{s} \). However, the above value of \( C_p,\text{Aug} \) seems to underestimate the gains of \( \text{InP}/\text{InGaAs} \) HBTs fabricated in our laboratory and others reported in the literature. Additionally, the above value of \( C_p,\text{Aug} \) produces a large negative temperature coefficient for current gain of \( \text{InGaAs}-\text{based} \) Npn HBTs which is far from those reported in the literature [347-349]. Therefore, the lower value of \( C_p,\text{Aug} = 3.6 \times 10^{-29} \text{ cm}^6/\text{s} \) reported in [192] is adopted in the present work. The above Auger coefficient results in \( f_2(\text{In}_{0.53}\text{Ga}_{0.47}\text{As}) = 4.26 \), and hence, \( f_2(\text{InAs}) = 7.14 \) can be obtained from linear interpolation. Room temperature Auger coefficients of \( \sim 1 \times 10^{-30} \) and \( 0.9 \times 10^{-30} \text{ cm}^6/\text{s} \) are reported in [335].

![Figure 5.34](image-url)
are given in [184] for GaP and InP, respectively, which lead to $f_2(GaP) = 0.30$ and $f_2(InP) = 0.27$. Due to the unavailability of reliable data for AlAs, $f_2(AlAs) = f_2(GaP)$ is assumed.

Table 5.12: Comparison between the recombination parameters obtained in this work and those reported in the literature for various materials ($T = 300K$).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Material</th>
<th>This work</th>
<th>Reported range</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B_{rad}(cm^3/s)$</td>
<td>GaAs</td>
<td>$1.0 \times 10^{-10}$</td>
<td>$(0.67-2) \times 10^{-10}$</td>
<td>[28,191,337-338,341-343]</td>
</tr>
<tr>
<td></td>
<td>InAs</td>
<td>$2.1 \times 10^{-11}$</td>
<td>$2.1 \times 10^{-11}$</td>
<td>[177]</td>
</tr>
<tr>
<td></td>
<td>In$<em>{0.53}$Ga$</em>{0.47}$As</td>
<td>$1.0 \times 10^{-10}$</td>
<td>$(1.0 \pm 0.3) \times 10^{-10}$</td>
<td>[330]</td>
</tr>
<tr>
<td></td>
<td>In$<em>{0.77}$Ga$</em>{0.28}$As$<em>{0.60}$P$</em>{0.40}$</td>
<td>$0.96 \times 10^{-10}$</td>
<td>$(0.4-1.43) \times 10^{-10}$</td>
<td>[218,339,344-345]</td>
</tr>
<tr>
<td></td>
<td>In$<em>{0.58}$Ga$</em>{0.42}$As$<em>{0.90}$P$</em>{0.10}$</td>
<td>$0.77 \times 10^{-10}$</td>
<td>$(0.3-1.5) \times 10^{-10}$</td>
<td>[287,333,345,350]</td>
</tr>
<tr>
<td>$C_{n, Aug}(cm^3/s)$</td>
<td>GaAs</td>
<td>$3.0 \times 10^{-30}$</td>
<td>$(1-20) \times 10^{-30}$</td>
<td>[28,335,337,340,343]</td>
</tr>
<tr>
<td></td>
<td>In$<em>{0.53}$Ga$</em>{0.47}$As</td>
<td>$3.6 \times 10^{-29}$</td>
<td>$3.6 \times 10^{-29}$, $8.1 \times 10^{-29}$</td>
<td>[192,339]</td>
</tr>
<tr>
<td></td>
<td>In$<em>{0.72}$Ga$</em>{0.28}$As$<em>{0.60}$P$</em>{0.40}$</td>
<td>$2.23 \times 10^{-29}$</td>
<td>$1 \times 10^{-29}$, $5 \times 10^{-29}$</td>
<td>[287]</td>
</tr>
<tr>
<td>$C_{n, Aug}(cm^3/s)$</td>
<td>In$<em>{0.53}$Ga$</em>{0.47}$As</td>
<td>$8.1 \times 10^{-29}$</td>
<td>$8.1 \times 10^{-29}$, $9 \times 10^{-29}$</td>
<td>[192,339,344]</td>
</tr>
<tr>
<td></td>
<td>GaAs</td>
<td>$5 \times 10^{-30}$</td>
<td>$5 \times 10^{-30}$</td>
<td>[343]</td>
</tr>
<tr>
<td></td>
<td>In$<em>{0.49}$Ga$</em>{0.51}$P</td>
<td>$1.24 \times 10^{-30}$</td>
<td>$&lt;3 \times 10^{-30}$</td>
<td>[330]</td>
</tr>
</tbody>
</table>

Calculation of $C_{n, Aug}$ in InGaAsP lattice-matched to InP has shown that $C_{n, Aug}$ is only slightly larger than $C_{p, Aug}$ for the entire range of composition (i.e., from In$_{0.53}$Ga$_{0.47}$As to InP) [287]. Also, $C_{n, Aug}$(GaAs,300K) = $5 \times 10^{-30}$ cm$^3$/s is obtained in [343], which is again larger than the value for p-GaAs. Setting $f_3$(GaAs) = 1.0 results in $C_3 = 4.52 \times 10^{-29}$ cm$^3$/s. For In$_{0.53}$Ga$_{0.47}$As, $C_{n, Aug} = 8.1 \times 10^{-29}$ cm$^3$/s as given by Ahrenkiel et al. [339] results in $f_3$(In$_{0.53}$Ga$_{0.47}$As) = 5.72, and $f_3$(InAs) is adjusted accordingly. The values of $f_3$ in the binaries together with all other fitting parameters required to model the recombination mechanisms are summarised in Table 5.11. A comparison of reported experimental recombination parameters with those obtained in this work (Table 5.12) supports the feasibility of the models used here.

Parameters of the defect related SRH recombination are very much dependent on the material quality and the growth conditions. They may vary by orders of magnitude in the same material grown under different conditions, and therefore, it is almost impossible to assign a unique value to these parameters. Nevertheless, useful trends can be observed in various materials. For example, the maximum effective lifetimes ($\tau_{n(p)0,SRH}$) observed in various materials depend on the maturity of that particular material and its growth technology. For well studied materials such as GaAs, In$_{0.53}$Ga$_{0.47}$As, and n-InP, $\tau_{n(p)0,SRH}$ is large, on the order of a few microseconds. But, materials such as p-InP, which has much less application in semiconductor devices, or Al$_x$Ga$_{1-x}$As, which suffers from Al-related defects and DX centres, have much smaller SRH lifetimes.

The procedure used in this work to select the SRH recombination parameters for various materials is as follows. $\tau_{n(p)0,SRH}$ for a certain material is almost equal to the maximum lifetime observed to date for that material. Then $N_{n(p),SRH}$ is selected based on the experimentally observed low-injection “effective” lifetimes (see Eq. (5.82)) reported in the
literature, in order to bring reasonable agreement between those data and Eq. (5.83) with parameters listed in Table 5.11. For this purpose the data in [28,92,111,184,186,191,204,248, 331-332,339,341,343,346] are used. In particular, $N_{n(p)0,SRH}$ for $In_{0.55}Ga_{0.47}As$ is set such that the linear coefficient of Eq. (5.83) is equal to the value of $1.43 \times 10^{10}$ cm$^3$/s fitted to the effective lifetime data in [339]. Also it is observed that the lifetime in $Al_xGa_{1-x}As$ drops significantly as the Al mole fraction increases from 0.0 to about 0.2, but afterwards remains almost constant [92,204,346]. This can be modelled by a sharply decreasing $\tau_{n(p)0,SRH}$ and $N_{n(p),SRH}$ with increasing $x$ up to 0.2, and constant values afterwards.

The values of the SRH recombination parameters are given in Table 5.11 for the five binaries as well as $In_{0.55}Ga_{0.47}As$, $In_{0.49}Ga_{0.51}P$, and $Al_{0.2}Ga_{0.8}As$. Logarithms of these parameters will then be linearly interpolated between the given values in Table 5.11 in order to obtain them in the entire range of ternary composition. For example, $\tau_{n0,SRH}(In_{1-x}Ga_xAs)$ is determined as:

$$
\log(\tau_{n0,SRH}(In_{1-x}Ga_xAs)) = \begin{cases} 
\left(1 - \frac{x}{0.47}\right) \cdot \log(\tau_{n0,SRH}(InAs)) + \frac{x}{0.47} \cdot \log(\tau_{n0,SRH}(In_{0.55}Ga_{0.47}As)) & \text{for } 0 \leq x \leq 0.47 \\
\frac{x - 0.47}{0.53} \cdot \log(\tau_{n0,SRH}(GaAs)) + \frac{1 - x}{0.53} \cdot \log(\tau_{n0,SRH}(In_{0.55}Ga_{0.47}As)) & \text{for } 0.47 \leq x \leq 1
\end{cases}
$$

(5.85)

As to the SRH parameters in $InGaAsP$, logarithm of these can be interpolated linearly between those of $In_{1-x}Ga_xAs$ and $In_{1-x}Ga_xP$, similar to (5.84). Finally, the SRH related parameters are assumed to be temperature independent, as the results by Mogensen et al. [332] also imply.

Fig. 5.35 summarises the effective lifetimes (Eq. (5.83)) obtained for some materials of interest as a function of total doping concentration. Graded $p^+ - Al_xGa_{1-x}As$ ($x = 0.11 \rightarrow 0.21$) with $p = 1 - 2 \times 10^{19}$ cm$^{-3}$ has been used as the base layer of Npn HBTs [111], but the authors measured electron lifetimes almost an order of magnitude smaller than a $p$-GaAs base which significantly degraded the DC gain of their devices. This is consistent with the data in Fig. 5.35. Also it can be seen that the hole lifetime of $In_{0.49}Ga_{0.51}P$ in the doping range of $10^{17} - 10^{18}$ cm$^{-3}$ is almost half as much as hole lifetime in GaAs, but more than an order of magnitude larger than that of $Al_{0.3}Ga_{0.7}As$. It can also be observed that the lifetimes of minority carriers in heavily-doped $In_{0.55}Ga_{0.47}As$ sharply decrease at high doping density, mainly due to the presence of a large Auger coefficient. Therefore, doping of the base in InP/InGaAs HBTs should not be increased beyond $4 - 5 \times 10^{19}$ cm$^{-3}$, otherwise the DC gain will be seriously degraded.
Figure 5.35- Variation of the low-injection effective carrier lifetime with doping concentration in various materials. Materials commonly used as either the base or the emitter of HBTs are shown in this figure.

5.5.5. Interface Recombination

It is well known that the quality of the B-E heterointerface seriously degrades the current gain of HBTs at low to medium current density range, especially in the case of graded HBTs [52]. Any lattice mismatch or interruption during the growth of the B-E heterojunction may create dangling bonds and trapping states which significantly reduce the carrier lifetime inside the top layer (usually emitter) and degrade the current gain. Therefore, it is important to include a mechanism of recombination at an interface between two layers.

The approach used in this thesis in order to include the effect of interface recombination is to modify the SRH lifetime of carriers within a specific width on top of an interface. Then, either the thickness of the interfacial layer or the effective lifetime inside this region can be used as a fitting parameter [52]. Here it is found that fixing the thickness and using the effective lifetime as the fitting parameter is an easier choice. An appropriate thickness for the interfacial layer is found to be ~100Å. As an example, simulation of a simple InGaP/GaAs SHBT is shown in Fig. 5.36. For this simulation, the width of the interfacial region at the B-E heterointerface is fixed at 100Å, but the effective lifetime in this region is changed from 0.1ps to 3ps. A case where interface recombination is completely ignored is also shown as reference. It can be seen that the interface recombination only affects the low to medium range of base current, and it has no effect on I_C. The medium base current ideality factor changes from 1.1 to 2.2 as the effective lifetime is reduced to 0.1ps.
5.6. Bandgap Narrowing (BGN)

As the doping concentration in semiconductor increases, the strict periodicity of the lattice is disturbed by the existence of the impurity atoms, and various heavy doping effects occur which may include impurity band widening, band tailing, electron-electron interaction, and screening effects [162,204]. But according to the rigid band model [162], all these effects can be accounted for by a single parameter, the bandgap narrowing $\Delta E_{g}^{\text{BGN}}$, which is considered a function of doping concentration.

Since the base, cap layer, and sub-collector of HBTs are normally doped well into the degenerate level, BGN is expected to have a pronounced effect on device behaviour. In contrast to $n^+p^+n$ Si BJTs, bandgap narrowing actually enhances the performance of $N^+n$ HBTs. Heavy doping inside the base of HBTs should increase the bandgap difference between the base and emitter, and therefore, enhance the emitter injection efficiency [343]. Additionally, BGN inside the base reduces the electron injection barrier, hence the turn-on voltage. This is also beneficial, since it reduces the power dissipation [337]. Not only the total BGN inside the base of HBTs is important, but the distribution of BGN between the conduction and valence band also has implications on device performance [351]. In abrupt HBTs, BGN can affect the band alignment at the B-E heterointerface, which itself influences the tunnelling and thermionic-emission current across this junction.

In order to determine the magnitude of BGN and its distribution between conduction and valence bands, a modified version of Jain and Roulston model [352] is used in the present work. This model has been successfully applied to Si, Ge, p-GaAs, and p-GaSb, and presumably can be also applied (with some modifications) to other III-V materials [352-353].

According to [352], there are four major contributions to the BGN: (1) shift $\Delta E_{x,\text{maj}}$ of the majority band edge due to exchange interaction; (2) shift $\Delta E_{\text{cor},\text{mino}}$ of the minority band edge due to carrier-carrier interaction, also designated as the correlation energy; (3) shift $\Delta E_{i,\text{maj}}$ of the majority band edge due to carrier-impurity interactions; and (4) shift $\Delta E_{i,\text{mino}}$ of the minority band edge due to carrier-impurity interaction. All shifts are into the bandgap and the total BGN is obtained by numerically adding all the shifts.
\[ \Delta E_{g}^{\text{bgn}} = \Delta E_{x,\text{maj}} + \Delta E_{\text{cor,mino}} + \Delta E_{i,\text{maj}} + \Delta E_{i,\text{mino}} \]  

(5.86)

Following [352], various energies will be normalised by the effective Rydberg energy, \( R_y \), and the doping concentration will be expressed in terms of the dimensionless parameter \( r_s \):

\[ R_y = \frac{13.6m_{\text{maj}}}{e_{s}^2} \text{ (eV)} \]  

(5.87)

\[ r_s = \frac{1}{a_B} \sqrt[3]{\frac{3}{4\pi N}} \]  

(5.88)

where \( a_B \) is the effective Bohr radius given by:

\[ a_B = \frac{0.53e_s}{m_{\text{maj}}} \times 10^{-8} \text{ (cm)} \]  

(5.89)

Here \( m_{\text{maj}} \) is the density-of-states (DOS) effective mass of majority carriers divided by the free electron mass. The final expression for BGN applicable to all semiconductors reads:

\[ \frac{\Delta E_{g}^{\text{bgn}}}{R_y} = \frac{1.83\Lambda}{N_{b}^{1/3}r_s} + \frac{1.57}{N_{b}^{1/3}r_s^{3/4}} + \frac{0.95}{\Delta E_{i,\text{maj}}} + \frac{m_{\text{mino}}}{\Delta E_{\text{cor,mino}}} \frac{1.57C_{\text{bgn}}(x,y)}{\Delta E_{i,\text{mino}}} \]  

(5.90)

where \( N_{b} \) \((N_{b,\text{mino}})\) is the number of equivalent valleys in the majority (minority) band, and \( \Lambda \) is a parameter which is fitted to be 1 for n- and 0.75 for p-type III-V compounds [353]. The only differences between this work and [352] are in the last term of (5.90):

1. For n-type materials, the normal DOS effective mass is used as \( m_{\text{mino}} \) rather than the average of the light and heavy hole masses used in [352]. Also \( N_b \) of the minority band is used in the last term on the RHS of (5.90) instead of \( N_b \) of the majority band used in [352]. The latter assumption seems more reasonable, because it appears that the interaction between impurities and minority carriers is not related to the number of equivalent valleys in the majority band. After all, the above two assumptions give almost the same results as those in [352] for n- and p-type GaAs.

2. To account for the BGN near the crossover region of materials such as AlGaAs and InGaP, one has to consider the three different conduction band minima with three different values of \( N_b \) (i.e., 1 for the \( \Gamma \)-band, 4 for the L-band, and 3 for the X-band). Therefore, the amount of BGN in the conduction band will be different for these three minima. The appropriate formulae for the amount of BGN in the valence band and the three conduction bands are summarised in Table 5.13. Bear in mind that evaluation of BGN at the crossover point of an n-type material like AlGaAs using the formulation of Jain and Roulston (i.e., using \( N_{b,\text{maj}} \) in the expression for \( \Delta E_{i,\text{mino}} \)) leads to a paradox; it is not clear that \( \Delta E_{i,\text{mino}} \) should be evaluated using \( N_b = 1 \) (for \( \Gamma \)-band) or 3 (for X-band).
A detailed analysis of photoluminescence spectra of n-InP by Sieg and Ringel [354] has revealed that the original form of the theoretical equation (5.90) as given in [352] (i.e., without the fitting parameter $C_{bgn}$) largely overestimates the experimental data. Similar disagreement can be observed in other reported experimental data in the case of n-GaAs [355-356], n-InP [357], n-In$_{0.49}$Ga$_{0.51}$P [260], and n-In$_{0.72}$Ga$_{0.28}$As [358]. Sieg and Ringel [354] showed that the disagreement is mainly due to the last term in Eq. (5.90) which largely overestimates the hole-donor interaction component of BGN because of a large $m_p^*/m_n^*$ ratio in III-V compounds. Some possible explanation for this disagreement is also proposed in [354]. Therefore, in this thesis a material dependent fitting parameter $C_{bgn}$ was multiplied by the last term in (5.90) in order to bring agreement to the experimental data.

Table 5.13- The amount of BGN in the valence band and the three conduction bands in n- and p-type materials.

<table>
<thead>
<tr>
<th>$\Delta E_{gV}$/Ry</th>
<th>$n$-type material</th>
<th>p-type material</th>
</tr>
</thead>
<tbody>
<tr>
<td>$2$</td>
<td>$0.95 \frac{m_p^*}{r_s} + \frac{1.57C_{bgn}(x,y)}{2 \times r_s^{3/2}}$</td>
<td>$1.83 \times 0.75 + \frac{1.57}{2 \times r_s^{3/2}}$</td>
</tr>
<tr>
<td>$1$</td>
<td>$1.83 \times 1 + \frac{1.57}{1 \times r_s^{3/2}}$</td>
<td>$0.95 \frac{m_n^*}{r_s^{3/2}} + \frac{1.57}{1 \times r_s^{3/2}}$</td>
</tr>
<tr>
<td>$3$</td>
<td>$1.83 \times 1 + \frac{1.57}{3 \times r_s^{3/2}}$</td>
<td>$0.95 \frac{m_n^*}{r_s^{3/2}} + \frac{1.57}{3 \times r_s^{3/2}}$</td>
</tr>
<tr>
<td>$4$</td>
<td>$1.83 \times 1 + \frac{1.57}{4 \times r_s^{3/2}}$</td>
<td>$0.95 \frac{m_n^*}{r_s^{3/2}} + \frac{1.57}{4 \times r_s^{3/2}}$</td>
</tr>
</tbody>
</table>

For $p$-type materials $C_{bgn}$ is set equal to unity. Values of $C_{bgn}$ equal to 0.2, 0.25, 0.25, 1.0, and 0.13 can best fit the experimental data for GaAs, InP, In$_{0.49}$Ga$_{0.51}$P, Al$_{0.3}$Ga$_{0.7}$As, and In$_{0.72}$Ga$_{0.28}$As, respectively (see Figs. 5.37-5.39). The values of the parameters for other binary and ternary compounds can be found using appropriate linear interpolations. The only exception is Al$_x$Ga$_{1-x}$As. To avoid $C_{bgn}$ values larger than unity, a linear interpolation between $C_{bgn}$(GaAs) and $C_{bgn}$(Al$_{0.3}$Ga$_{0.7}$As) is used for $C_{bgn}$(Al$_x$Ga$_{1-x}$As) in the case of $0 \leq x \leq 0.3$, while for $0.3 \leq x \leq 1$ $C_{bgn}$(Al$_x$Ga$_{1-x}$As) = 1 is assumed.
Chapter 5 Physical Parameters in III-V Semiconductors

Note that $N_0$ is assumed to be 2 for the valence band due to the presence of the light and heavy hole subbands. Also note that the BGN modifies both the bandgap and the electron affinity; bandgap will be reduced by the sum of BGN in the valence and the conduction bands, and electron affinity will be increased by the amount of BGN in the conduction band. The modifications can be formulated as follows:

\[
E_{gv} = E_{gv0} - \Delta E_{gv} - \Delta E_{gV}
\]

(5.91)

\[
E_g = \min\{E_{g\Gamma}, E_{gX}, E_{gL}\}
\]

(5.92)

\[
\chi = \chi_0 + (E_g - E_{g0}) - \Delta E_{gV}
\]

(5.93)

where the parameters with subscript ‘0’ are those of an undoped material, and $\nu$ stands for $\Gamma$, L, or X. Figs. 5.37-5.39 compare the experimental data for BGN in various III-V compounds with those obtained in this work using Eq. (5.90). Also shown in Fig. 5.37 is the theoretical prediction of BGN in n-GaAs given by Jain and Roulston's original formula [352]. It can be seen that Jain and Roulston's formula largely overestimates the experimental BGN in n-GaAs, while this work's modified version closely represents the measured data.

![Figure 5.38- Comparison between the measured data for BGN in n-InP [354,357] with those calculated in the present work.](image)

![Figure 5.39- Comparison between the measured data for BGN in n-In$_{0.49}$Ga$_{0.51}$P [260], n-Al$_{0.3}$Ga$_{0.7}$As [360], and p-Al$_{0.33}$Ga$_{0.67}$As [361] with those calculated in the present work.](image)

5.7. Ionised Doping Concentration

For modelling semiconductor devices operating at room temperature, it is common practice to employ a 100% ionisation approximation, regardless of the impurity doping level. However, at low temperatures or medium range of doping concentrations, this assumption is inaccurate.

The incomplete ionisation of impurities can be taken into consideration by the following formulas [161]:

116
where \( g_D \) and \( g_A \) are the ground state degeneracy factors of the donor and acceptor impurity levels, \( E_D \) and \( E_A \), respectively. We can rewrite (5.94) and (5.95) in terms of the ionisation energies, \( (E_C - E_D) \) and \( (E_A - E_V) \). For example, (5.94) can be written as:

\[
N_D^+ = \frac{N_D}{1 + g_D \cdot \exp \left( \frac{E_m - E_C}{kT} \right) \cdot \exp \left( \frac{E_C - E_D}{kT} \right)}
\]

Now using (4.11), we will have:

\[
N_D^+ = \frac{N_D}{1 + g_D \cdot \exp \left( F_{1/2} \left( \frac{n}{N_C} \right) \right) \cdot \exp \left( \frac{E_C - E_D}{kT} \right)}
\]

Following the same trend as in the case of donors yields:

\[
N_A^- = \frac{N_A}{1 + g_A \cdot \exp \left( F_{1/2} \left( \frac{p}{N_V} \right) \right) \cdot \exp \left( \frac{E_A - E_V}{kT} \right)}
\]

In (5.96) and (5.97), \( F_{1/2}^{-1} \) is the inverse function of the Fermi-Dirac integral of order \( 1/2 \). An analytical approximation for this function is given in Appendix C. The degeneracy factor for donors, \( g_D \), is two, whereas that for acceptors, \( g_A \), is four [161].

The ionisation energies are dependent on both the host semiconductor and dopant species. For instance, the Si (donor) ionisation energy in AlGaAs rises from a few meV in GaAs to a maximum (around 100meV) near the composition \( x = 0.4 \) at which a crossover from direct to indirect energy gap occurs [362]:

\[
E_D = \min \{E_{\Gamma 0} - 0.009, E_{V0} + 0.54x + 1.57\} \quad \text{for Si-doped Al}_{x}\text{Ga}_{1-x}\text{As}
\]
and \( N_A^- = N_A \) will be assumed in the majority of the simulations performed in this work (see the following paragraph), detailed formulation of activation energies in various materials used in the simulation code will be omitted here.

Fig. 5.40 shows, as an example, the ionisation ratio in Si-doped GaAs and \( Al_{0.3}Ga_{0.7}As \) as a function of total doping concentration. An uncompensated, charge neutral material (i.e., \( n = N_D^+ \)) is assumed for this figure. The ionisation ratios in this figure are surprisingly low for highly-doped materials at room temperature. This is in complete contrast to the accepted fact that highly-doped materials are normally fully ionised, even at low temperatures [161: p.133] [354]. When a semiconductor becomes degenerate, formation of dopant bands and band tailing inside the semiconductor causes the dopant ionisation relationships (5.94) and (5.95) to fail [161]. In such circumstances, it is common practice to assume that the dopant sites are totally ionised. Since according to Fig. 5.40, lowly-doped materials are also fully ionised, and degeneracy of n-type III-V compounds occurs around doping level of \(~10^{17}\ \text{cm}^{-3}\), only for \( N_D \) in the range \( 10^{16}-10^{17}\ \text{cm}^{-3} \) incomplete (minimum \(~70\%\) at room temperature) ionisation of donors may occur\(^\ddagger\). Additionally, majority of the layer structures specified by the growers rely on electrical methods (such as capacitance-voltage method) for the determination of doping densities. These methods directly measure the ionised, rather than total, doping densities. Therefore, incomplete ionisation of dopants is not considered in the majority of the simulations used in the present work, and it is assumed that \( N_D^+ = N_D \) and \( N_A^- = N_A \).

\[^\ddagger\]\ For p-type materials a similar conclusion can be drawn that incomplete activation of acceptors occurs in a doping range just before the limit for degeneracy.
CHAPTER 6. NUMERICAL METHODS

The system of partial differential equations which form the basic semiconductor equations together with appropriate boundary conditions and models for physical parameters have been investigated in the previous chapters. This system can not be solved explicitly in general. Therefore, the solution must be calculated by means of numerical approaches. We shall consider in this chapter such solution procedures for the Eqs. (4.6)-(4.8).

Any numerical approach for the solution of such a system consists essentially of three tasks. First, the domain, i.e., the simulation geometry of the device, has to be partitioned into a finite number of sub-domains, in which the solution can be approximated easily with a desired accuracy. Secondly, the differential equations have to be approximated in each of the sub-domains by algebraic equations, which involve only values of the continuous dependant variables at discrete points in the domain and knowledge of the structure of the chosen functions which approximate the dependant variables within each of the sub-domains. In that way, one obtains a fairly large system of, in general non-linear, algebraic equations with unknowns comprised of approximations of the continuous dependant variables at discrete points. The solution of this system is the final third task to be carried out. In general, only iterative methods are applicable for the solution of systems of non-linear algebraic equations. The most important method, without any doubt, is Newton’s method together with some modifications. All the above topics together with some discussions on normalisation of the equations and calculation of device terminal characteristics are addressed in this chapter.

6.1. Normalisation of the Variables and Parameters

Since the dependant variables \( (\psi, n, p) \) in the basic equations are of greatly different orders of magnitude and show strongly different behaviours in regions with small and large space charge, the first step towards the structural analysis of the basic equations has to be appropriate scaling. The most often used normalisation factors are the electron charge \( q \) for charge, and the thermal voltage \( V_T = kT/q \) for potential. As to the normalisation factor for carrier and doping concentration, the common choice is the intrinsic carrier concentration, \( n_i \) [364]. While it has a proper value at room temperature, the intrinsic carrier concentration changes exponentially with temperature causing extremely large values for the normalised doping concentrations at low temperatures. To scale by the maximum doping density as recommended by Markowich [365] is also not feasible, since the scaled \( n_i \) would be extremely small. A way out of this dilemma is to use the concentration \( n_{\text{norm}} \) defined by [366]:

\[
n_{\text{norm}} = \sqrt{4.8 \times 10^{22} n_{ir}} \tag{6.1}
\]

The magic concentration in the above equation is the number of silicon atoms \( /\text{cm}^3 \). It serves as an absolute upper limit for the maximum possible concentration of any type. \( n_{ir} \) in the above equation is the intrinsic carrier concentration at the reference point. For GaAs at
Chapter 6 Numerical Methods

\( T = 300 \text{K}, \quad n_{\text{norm}} = 3.22 \times 10^{14} \text{ cm}^{-3} \) will be achieved, while at \( T = 77 \text{K}, \quad n_{\text{norm}} = 2.38 \times 10^{-5} \text{ cm}^{-3} \). This scaling equilibrates the relevant concentrations and thus is optimal for avoiding fatal floating point underflow and overflow problems.

Applying the above mentioned scaling factors to the Poisson's equation results in:

\[
\frac{d}{dz} \left( \frac{e_r}{e_0} \frac{d\psi}{dz} \right) = q \cdot n_{\text{norm}} \cdot L_{\text{norm}}^2 \cdot \left( \hat{n} - \hat{p} - \hat{N}_D^+ + \hat{N}_A^- \right)
\]

(6.2)

where the sign "^\prime" above the variables in equation (6.2) means that they are normalised, and \( V_{\text{norm}} \) and \( L_{\text{norm}} \) are the normalising factors for electrostatic potential (equal to \( V_T \) in the present case) and length, respectively. If we define:

\[
L_{\text{norm}} = \sqrt{ \frac{e_0 \cdot V_{\text{norm}}}{q \cdot n_{\text{norm}}} }
\]

(6.3)

**Table 6.1- Normalisation factors for variables used in the basic equations.**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Examples</th>
<th>Normalisation Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Charge</td>
<td>( Q_B, q )</td>
<td>( Q_{\text{norm}} = q )</td>
</tr>
<tr>
<td>Potential, Energy (in eV)</td>
<td>( \psi, V, \phi_n, E_g )</td>
<td>( V_{\text{norm}} = V_T = kT/q )</td>
</tr>
<tr>
<td>Concentration</td>
<td>( n, p, N_D, N_D )</td>
<td>( n_{\text{norm}} = \sqrt{4.8 \times 10^{22} n_{ir}} )</td>
</tr>
<tr>
<td>Length</td>
<td>( z )</td>
<td>( L_{\text{norm}} = \sqrt{e_0 \cdot V_{\text{norm}} / q \cdot n_{\text{norm}}} )</td>
</tr>
<tr>
<td>Electric Field</td>
<td>( E, E_n, E_p )</td>
<td>( E_{\text{norm}} = V_{\text{norm}} / L_{\text{norm}} )</td>
</tr>
<tr>
<td>Mobility</td>
<td>( \mu_n, \mu_p )</td>
<td>( \mu_{\text{norm}} = 1/L_{\text{norm}} )</td>
</tr>
<tr>
<td>Time</td>
<td>( t, \tau_n, \tau_p )</td>
<td>( t_{\text{norm}} = L_{\text{norm}}^2 / \mu_{\text{norm}} \cdot V_{\text{norm}} )</td>
</tr>
<tr>
<td>Recombination Rate</td>
<td>( R )</td>
<td>( R_{\text{norm}} = n_{\text{norm}} / t_{\text{norm}} )</td>
</tr>
<tr>
<td>Current Density</td>
<td>( J_n, J_p )</td>
<td>( J_{\text{norm}} = q \cdot \mu_{\text{norm}} \cdot n_{\text{norm}} \cdot E_{\text{norm}} )</td>
</tr>
</tbody>
</table>

then the constant on the right-hand-side of (6.2) equates to unity. The normalisation factors for other variables can be derived by the above basic normalisation factors and the semiconductor equations themselves. The details are listed in Table 6.1. Note that for the mobility normalisation factor we have assumed that the diffusion constants are scaled using \( D_{\text{norm}} = 1 \text{ cm}^2 /\text{s} \).

The normalised form of the essential equations that will be used at other parts of this chapter are summarised as follows:

\[
\frac{dJ_n}{dz} = R \quad \text{electron continuity equation} \quad (6.4)
\]
\[
\frac{dJ_p}{dz} = -R \quad \text{hole continuity equation} \quad (6.5)
\]
\[
\frac{d}{dz} \left( e_r \frac{d\psi}{dz} \right) = n - p + N_A^- - N_D^+ \quad \text{Poisson's equation} \quad (6.6)
\]
6.2. Discretisation of the Equations

The first step in solving differential equations numerically involves the discretisation of both the definition domain of the equations and the equations themselves. We name the former process formation of the grid space (or mesh generation), and the latter, discretisation of the differential equations.

6.2.1. Mesh Generation

For the one-dimensional case, the grid space formation is rather straightforward. A non-uniform mesh spacing is usually adopted in the $z$ direction with coarser spacing in the regions where the dependant variables change gradually and finer spacing in the regions of drastic change of variables. The algorithm of grid generation is as follows: the physical junctions and the interface boundaries are first allocated on the grids, the mesh spacing next to those grid

$$J_n = V_T \cdot \mu_n \frac{dn}{dz} - n \cdot \mu_n \frac{d\psi_n}{dz} \quad \text{electron current density (drift-diffusion)}$$  \hspace{0.5cm} (6.7)

$$J_p = -V_T \cdot \mu_p \frac{dp}{dz} - p \cdot \mu_p \frac{d\psi_p}{dz} \quad \text{hole current density (drift-diffusion)}$$  \hspace{0.5cm} (6.8)

$$\psi_n = \psi + (\chi - \chi_r) + V_T \cdot \ln\left(\frac{N_C}{N_{Cr}}\right) \quad \text{effective potential for electrons}$$  \hspace{0.5cm} (6.9)

$$\psi_p = \psi + (\chi - \chi_r) + (E_g - E_{gr}) - V_T \cdot \ln\left(\frac{N_V}{N_{Vr}}\right) \quad \text{effective potential for holes}$$  \hspace{0.5cm} (6.10)

Figure 6.1- Doping profile of the sample HBT, also showing the grid spacing.
points is chosen to be a small number and the mesh spacing increases for the grid away from the junction or interface according to a geometrical series with ratio larger than one (i.e., \( h_{i+1} = r \cdot h_i \), \( r > 1 \)). In the present software a default minimum mesh spacing of \( h_{\text{min}} = 20\text{Å} \) and a ratio of \( r = 2 \) is used. Also an upper limit is set for the mesh spacing which has a default value of \( h_{\text{max}} = 500\text{Å} \). Figure 6.1 shows the doping profile of a sample HBT structure, as specified in Table 6.2, together with the mesh spacing generated by the program.

### Table 6.2- Layer specification of a sample AlGaAs/GaAs HBT.

<table>
<thead>
<tr>
<th>Layer</th>
<th>Material</th>
<th>Al Mole Fraction</th>
<th>Thickness (Å)</th>
<th>Doping ((\text{cm}^{-3}))</th>
<th>Dopant (C)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(n^+)-GaAs</td>
<td>-</td>
<td>2800</td>
<td>(5 \times 10^{18})</td>
<td>Si</td>
<td>Cap Layer</td>
</tr>
<tr>
<td>2</td>
<td>(n)-AlGaAs</td>
<td>0.3</td>
<td>800</td>
<td>(4 \times 10^{17})</td>
<td>Si</td>
<td>Emitter</td>
</tr>
<tr>
<td>3</td>
<td>(p^+)-GaAs</td>
<td>-</td>
<td>1000</td>
<td>(4 \times 10^{19})</td>
<td>C</td>
<td>Base</td>
</tr>
<tr>
<td>4</td>
<td>(n^-)-GaAs</td>
<td>-</td>
<td>10000</td>
<td>(3 \times 10^{16})</td>
<td>Si</td>
<td>Collector</td>
</tr>
<tr>
<td>5</td>
<td>(n^+)-GaAs</td>
<td>-</td>
<td>6000</td>
<td>(5 \times 10^{18})</td>
<td>Si</td>
<td>Sub-Collector</td>
</tr>
</tbody>
</table>

#### 6.2.2. Discretisation Scheme

The discretisation scheme for the differential equations uses the finite difference method, that is, the differential operators are replaced by difference operators. For the non-uniform grid space, however, the conventional central difference operator is not sufficient to achieve the necessary accuracy. Moreover, the current expression is quite different for tunnelling and thermionic-emission currents compared to the drift and diffusion current as given in (6.7) and (6.8). We, therefore, use a so-called “box” method [279] for discretisation. The basic idea is not to discretise the differential equation directly, but to first transform the differential equation into integral form, and then to express the relevant quantities at the middle of the grid interval and the integral in terms of basic variables at the grid points.

Integrating (6.4) to (6.6) from \(z_{i-1/2}\) to \(z_{i+1/2}\) yields:

\[
\varepsilon_{i+1/2} \frac{d\psi}{dz}_{i+1/2} - \varepsilon_{i-1/2} \frac{d\psi}{dz}_{i-1/2} = \int_{z_{i-1/2}}^{z_{i+1/2}} \left( n - p + N_A^- - N_D^+ \right) dz \tag{6.11}
\]

\[
J_{n,i+1/2} - J_{n,i-1/2} = \int_{z_{i-1/2}}^{z_{i+1/2}} R \cdot dz \tag{6.12}
\]

\[
J_{p,i+1/2} - J_{p,i-1/2} = \int_{z_{i-1/2}}^{z_{i+1/2}} R \cdot dz \tag{6.13}
\]

Further discretisation involves the evaluation of \( \frac{d\psi}{dz}_{i\pm1/2} \), \( J_{n,i\pm1/2} \), and \( J_{p,i\pm1/2} \) and the
numerical integration of the integrals on the right-hand-side of (6.11)-(6.13). The quantities at
the midpoint of the grid spacing can be evaluated with higher accuracy than those at the other
points, in terms of quantities at the grid points, for the same computation effort. This is
because we can take advantage of the inherent symmetry. For example, the simple expression
\[
\frac{d\psi}{dz}\bigg|_{i+1/2} = \frac{\psi_{i+1} - \psi_i}{z_{i+1} - z_i}
\]
has the accuracy of order two. Defining the mesh spacing as:
\[h_i = z_{i+1} - z_i, i = 1, 2, ..., N - 1\] (6.14)
where \(N\) is the total number of mesh points, yields:
\[
\frac{d\psi}{dz}\bigg|_{i+1/2} = \frac{\psi_{i+1} - \psi_i}{h_i} \tag{6.15}
\]
\[
\frac{d\psi}{dz}\bigg|_{i-1/2} = \frac{\psi_i - \psi_{i-1}}{h_{i-1}} \tag{6.16}
\]
Now the integrals of Eqs. (6.11)-(6.13) should be evaluated. Using the truncated Taylor series
\[
f(z) = f(z_i) + (z - z_i) \cdot f'(z_i) + \frac{(z - z_i)^2}{2} \cdot f''(z_i) \tag{6.17}
\]
it is a straightforward task to prove that the approximation
\[
\int_{z_{i-1/2}}^{z_{i+1/2}} f(z) \cdot dz = \left(a_i \cdot f_{i+1} + b_i \cdot f_i + c_i \cdot f_{i-1}\right) \cdot \frac{z_{i+1} - z_{i-1}}{2} \tag{6.18}
\]
with appropriate values of \(a_i, b_i,\) and \(c_i\) results in an accuracy of order four, where \(f_i\) is the
function value of \(f(z)\) at \(z = z_i\). If the ratio of adjacent grid spacings is denoted as:
\[r_i = \frac{h_{i-1}}{h_i}, i = 2, 3, ..., N - 1\] (6.19)
then the coefficients \(a_i, b_i,\) and \(c_i\) can be expressed as:
\[
\begin{align*}
a_i &= \frac{1}{12} \left[ \frac{1}{1 + r_i \cdot (1 - 2r_i)} \right] \\
b_i &= 1 - a_i - c_i \\
c_i &= \frac{1}{12} \left[ \frac{1}{1 + \frac{r_i - 2}{r_i \cdot (1 + r_i)}} \right]
\end{align*}
\] (6.20)
Also
\[
\int_{z_i}^{z_{i+1}} f(z) \cdot \frac{dz}{2} = \frac{1}{4} \left( 3f_i + f_{i+1} \right) \cdot \frac{h_i}{2}
\]  
(6.21)

has accuracy of order three.

To complete the discretisation of the entire integral equations, \( J_n \) and \( J_p \) at the midpoints should be replaced by appropriate expressions. First, we consider the drift-diffusion current (formulated by (6.7) and (6.8)). In the derivation of \( \frac{d\psi}{dz}_{i+1/2} \) we have already assumed that this derivative (i.e., electric field) is constant between two adjacent mesh points. Now, \( J_n, J_p, \) \( \frac{d\psi_n}{dz}, \) and \( \frac{d\psi_p}{dz} \) are additionally assumed to be constant between two adjacent mesh points. Therefore, for the electron concentration in the interval \([z_i, z_{i+1}]\) we have:

\[
V_T \cdot \mu_n \cdot \frac{dn}{dz} - \mu_n \cdot \frac{d\psi_n}{dz} \cdot n = J_{n,i+1/2} \quad \text{for} \quad n \in [z_i, z_{i+1}]
\]  
(6.22)

This ordinary differential equation should be solved together with the two boundary conditions: \( n(z_i) = n_i, \) \( n(z_{i+1}) = n_{i+1}. \) The solution reads [279]:

\[
n(z \in [z_i, z_{i+1}]) = [1 - g_i(z, \psi_n)] \cdot n_i + g_i(z, \psi_n) \cdot n_{i+1}
\]  
(6.23)

with the growth function \( g_i(z, \psi) \) defined as:

\[
g_i(z, \psi) = \frac{1 - \exp \left( \frac{\psi_{i+1} - \psi_i \cdot z - z_i}{V_T \cdot h_i} \right)}{1 - \exp \left( \frac{\psi_{i+1} - \psi_i}{V_T} \right)}
\]  
(6.24)

and

\[
J_{n,i+1/2} = \frac{V_T \cdot \mu_n |_{i+1/2}}{h_i} \cdot \left[ B \left( \frac{\psi_{n,i+1} - \psi_{n,i}}{V_T} \right) \cdot n_{i+1} - B \left( \frac{\psi_{n,i} - \psi_{n,i+1}}{V_T} \right) \cdot n_i \right]
\]  
(6.25)

Fully analogously one obtains the following expressions for holes:

\[
p(z \in [z_i, z_{i+1}]) = [1 - g_i(z, -\psi_p)] \cdot p_i + g_i(z, -\psi_p) \cdot p_{i+1}
\]  
(6.26)

\[
J_{p,i+1/2} = \frac{V_T \cdot \mu_p |_{i+1/2}}{h_i} \cdot \left[ B \left( \frac{\psi_{p,i+1} - \psi_{p,i}}{V_T} \right) \cdot p_i - B \left( \frac{\psi_{p,i} - \psi_{p,i+1}}{V_T} \right) \cdot p_{i+1} \right]
\]  
(6.27)

In the above equations:

\[
B(x) = \frac{x}{\exp(x) - 1}
\]  
(6.28)

is the Bernoulli function. The above mentioned approximations for the midpoint current
densities have been first suggested by Scharfetter and Gummel [367].

**Note 1.** The mid-interval values of the carrier mobility can be obtained by a simple interpolation. It might be more physically justified to assume that the relaxation times, which are proportional to the reciprocals of the carrier mobilities, are linear functions between neighbouring mesh points [279], which leads to the following interpolation scheme:

\[
\mu_{n,i+1/2} = \frac{2}{\frac{1}{\mu_{n,i}} + \frac{1}{\mu_{n,i+1}}}
\]

(6.29)

**Note 2.** The Bernoulli function has to be implemented very carefully for numerical computations in order to avoid potential underflow and overflow traps. A very effective method of implementation for the Bernoulli function has been suggested in [279].

So far we have considered the drift-diffusion component of current density. However, as mentioned earlier in section (4.1), this approach is not applicable at the heterojunction interfaces. Instead, a TFE boundary condition will be applied at the heterointerfaces in series with drift-diffusion current formulation in the bulk. In order to elaborate this, consider the discretisation of the continuity equation for electrons around the heterojunction interface. Using the mesh numbers in Fig. 6.2, discretised continuity equations for electrons can be written as:

\[
\begin{align*}
J_{n,i+1/2} - J_{n,i-1/2} &= \int_{z_i-1/2}^{z_i+1/2} R \cdot dz \\
J_{n,i+3/2} - J_{n,i+1/2} &= \int_{z_i+1/2}^{z_{i+1}} R \cdot dz
\end{align*}
\]

(6.30)

for mesh points \(i\) and \((i+1)\). In these equations, \(J_{n,i-1/2}\) and \(J_{n,i+3/2}\) can be formulated using (6.25). However, \(J_{n,i+1/2}\) represents the electron current across the heterointerface and can not be expressed in a form similar to (6.25). Therefore, another treatment for it is necessary.

In the “pure drift-diffusion (pure DD)” model for heterojunctions, one does not give an expression for \(J_{n,i+1/2}\) explicitly but uses an equation without it. Eliminating \(J_{n,i+1/2}\) in (6.30), one obtains the following equation:
\[ J_{n,i+3/2} - J_{n,i-1/2} = \int_{z_{i-1/2}}^{z_{i+3/2}} R \cdot dz \]  

(6.31)

To be sure, this represents current continuity across the heterojunction interface. However, now we are one equation short. Then, in the pure DD model for heterojunctions, it is assumed that the electron quasi-Fermi level is continuous at the heterojunction interface [53]:

\[ E_{fn,i} = E_{fn,i+1} \]  

(6.32)

As is evident from the above treatment, current transport mechanisms across the heterojunction interface are not taken into account explicitly in the pure DD model. Furthermore, (6.32) is only an assumption, and it is not a general condition. At abrupt heterojunctions where the conduction band discontinuity is large, the electron quasi-Fermi level is split [165,368]. Therefore, one has to use another approach. In this case, an expression for \( J_{n,i+1/2} \) based on TFE current formulation (i.e., equation (4.27)) will be employed. Discretising (4.27) and (4.31) results in:

\[ J_{n,i+1/2} = -A^*T^2 \cdot (1 + \delta) \left[ \frac{n_i}{N_{C,i} \cdot \gamma_{n,i}} \cdot \exp \left( \frac{E_{C,i} - E_{C,max}}{kT} \right) - \frac{n_{i+1}}{N_{C,i+1} \cdot \gamma_{n,i+1}} \cdot \exp \left( \frac{E_{C,i+1} - E_{C,max}}{kT} \right) \right] \]  

(6.33)

\[ J_{p,i+1/2} = A^*T^2 \left[ \frac{p_i}{N_{V,i} \cdot \gamma_{p,i}} \cdot \exp \left( \frac{E_{V,min} - E_{V,i}}{kT} \right) - \frac{p_{i+1}}{N_{V,i+1} \cdot \gamma_{p,i+1}} \cdot \exp \left( \frac{E_{V,min} - E_{V,i+1}}{kT} \right) \right] \]  

(6.34)

### 6.2.3. Discretisation of the Boundary Conditions

So far we have derived the discrete equations at all inner points of a given mesh. In the following we shall discuss the discretisation of boundary conditions. The discretisation of Dirichlet-type boundary conditions (e.g., (4.35), (4.37), and (4.40)) are rather straightforward. For example, (4.35) can be discretised as:

\[ n_j = \sqrt{\left( N_{D,j}^+ - N_{A,j}^- \right)^2 + 4n_{n,j}^2 \cdot \gamma_{n,j}} + \left( N_{D,j}^+ - N_{A,j}^- \right) \]  

(6.35)

\[ p_j = \sqrt{\left( N_{D,j}^+ - N_{A,j}^- \right)^2 + 4n_{p,j}^2 \cdot \gamma_{n,j} - \left( N_{D,j}^+ - N_{A,j}^- \right)} \]  

(6.36)

However, from the numerical calculation point of view, (6.36) should not be evaluated
directly because of inherent problems with cancellation errors. One should preferably use:

\[ p_j = \frac{n_{i,j}^2 \cdot \gamma_{n,j}}{n_j} \]  

(6.37)

Also in (6.35) and (6.37) the value of \( \gamma_{n,j} \) from previous step of iteration might be used. For the boundary condition (4.36) with a finite surface recombination velocity for minority carriers (\( S_p \)), one needs to replace the appropriate equation for the hole current density:

\[ S_p \cdot (p_j - p_0) = -J_{p,j+1/2} = \frac{V_T \cdot n_{i,j+1/2}}{h_j} \cdot \left[ B \left( \frac{\psi_{p,j} - \psi_{p,j+1}}{V_T} \right) p_{j+1} - B \left( \frac{\psi_{p,j+1} - \psi_{p,j}}{V_T} \right) p_j \right] \]  

(6.38)

(6.38) is written for the top side ohmic contact (i.e., \( j = 1 \)). For the bottom side ohmic contact (\( j = N \)), one has to use \( J_{p,j-1/2} \) on the right-hand-side. The case of current controlled base ohmic contact (Eq. (4.43)) is very similar to the above case, and we shall not repeat the formulation here. However, discretisation of a voltage-controlled ohmic contact with non-zero series resistance (Eqs. (4.39) and (4.42)) will be treated somehow different. The reason is that using Eqs. (6.25) and (6.27) directly to compute the current density for the majority carriers is prone to large cancellation errors [170] (also see section 6.6.1). Therefore, it is better to use the calculated current densities from previous iteration step. Thus, treating the current densities as known values, Eqs. (4.39) and (4.42) will be discretised in exactly the same manner as Dirichlet boundary conditions.

6.3. Solution Strategies

The main result obtained in the preceding section is that discretisation of the basic semiconductor equations yields a large system of non-linear algebraic equations with the values of the dependent variables of the differential equations at discrete points as unknowns. For the consideration in this section we adopt the following nomenclature for the system of discretised equation:

\[ F(w) = 0 \]  

(6.39a)

with

\[ F = \begin{bmatrix} f_\psi(w) \\ f_n(w) \\ f_p(w) \end{bmatrix} \quad \text{and} \quad w = \begin{bmatrix} \psi \\ n \\ p \end{bmatrix} \]  

(6.39b)

\( F \) is a vector function of rank three which itself consists of the vector functions \( f_\psi, f_n, \) and \( f_p \). These vector functions correspond to the discrete approximations for the Poisson equation and the continuity equations, respectively. The vector of unknowns \( w \) is also comprised of three vectors formed by the values of the electrostatic potential \( \psi \), electron concentration \( n \), and
hole concentration $p$ at discrete points of the grid. Thus $3N$ non-linear equations have to be solved, where, e.g., in the two-dimensional case the number of grid points $N$ can be as large as several thousands. The solution of this large system is the most time-consuming task during a numerical device simulation. Therefore, it is of great importance to choose an efficient numerical algorithm. Most simply, one treats each of the differential equations separately by decoupling the equations and solving the three systems successively. First, Poisson’s equation is solved assuming known carrier densities. Next comes each of the continuity equations with $\psi$ given from the first step. This sequence is iteratively repeated until self-consistent values of the desired accuracy for all unknown variables are obtained. Fig. 6.3 shows the flow diagram of this successive or decoupled method, which is due to Gummel [369]. It is advantageous for multi-dimensional simulations, as it saves storage and converges quite well as long as the coupling of the three equations is only weak. Moreover, the result of the first cycle of this iteration is rather close to the exact solution if the first guess of the majority carrier density is accurate enough [370].

The relatively simple implementation of the successive method has to be paid by its possible slow convergence if the three equations are strongly coupled. Considering the steady-state analysis of a bipolar device, Gummel’s algorithm converges well for low and moderate injection levels. But in the conditions of high level injection, both carriers become strongly coupled via the electrostatic potential in order to maintain quasi-neutrality and convergence slows down rapidly [370]. This convergence problem can be overcome mathematically by solving the three systems of equations simultaneously rather than alternatingly. This simultaneous method has the advantage that the mutual coupling between all equations is taken into account by a quadratically converging overall Newton iteration (see next section for details). Although the simultaneous Newton method is advantageous from a purely mathematical point of view, it is more involved with regard to program structure and storage requirements. Hence, comparisons between both methods depend not only on the device and its operating conditions, but also on the algorithms implemented for solving the different linearised systems [279]. Due to the relatively small storage requirements in one-dimensional simulations, the simultaneous method is adopted in the present case.

6.4. Linearisation of the Equations

6.4.1. Newton and Newton-like Methods

In general, only iterative methods are applicable for the solution of systems of non-linear
algebraic equations. The most important method, without any doubt, is Newton's method together with some modifications. In this section we will briefly review these methods and their properties.

Using the same nomenclature as in section (6.3), the Newton iteration method is defined as:

\[ M(w^k) \cdot (w^{k+1} - w^k) = -F(w^k) \quad k = 0, 1, 2, \ldots \]  

(6.40)

where the superscript \( k \) is the iteration step and the operator \( M(w) \) is defined as:

\[ M(w) = F'(w) \]  

(6.41)

Here \( F' \) is the Jacobian matrix, with its \((i,j)\)-th element comprised of the partial derivative of the \( i \)-th equation with respect to the \( j \)-th variable. It can be shown [279] that the Newton method mathematically guarantees the convergence of the iteration to the accurate solution \( w^* \), provided that the initial guess \( w^0 \) is sufficiently close to \( w^* \). For the solution of (6.40) it is necessary to solve a system of linear algebraic equations. This problem will be the topic of section 6.5.

In the following we consider modification of the Jacobian matrix which gives a linear operator \( M(w) \) in such a manner that the corresponding modified Newton method (6.40) exhibits improved convergence properties for an initial guess which is not sufficiently close to the solution \( w^* \). The main problem associated with the classical Newton method is the tendency to overestimate the length of the actual correction step for the iterate \([279]\). This phenomenon is frequently termed *overshoot*. It may also happen that the initial guess with which the classical Newton method is started does not lie in the region of attraction. Newton's method will then not converge at all. The modification to the classical Newton method is therefore also intended to enlarge the region of convergence for the initial guess.

One of the best established modifications to avoid overshoot of the classical Newton method is given by:

\[ M(w^k) = \frac{1}{t_k} \cdot F'(w^k) \]  

(6.42)

where \( t_k \) is a properly chosen parameter in the interval \([0,1]\) which satisfies the following:

\[ \left\| F'(w^k)^{-1} \cdot F\left(w^k - t_k \cdot F'(w^k)^{-1} \cdot F(w^k)\right) \right\| < \left\| F'(w^k)^{-1} \cdot F(w^k) \right\| \]  

(6.43)

Here the operator ‘\( \| \| \)’ is a vector norm. Condition (6.43) guarantees that the correction of the \( k \)-th iterate is an improved approximation to the solution [279]. This condition can be easily evaluated only if the inverse function of the Jacobian matrix is available. However, calculation of the inverse function can be very much time-consuming and must be avoided if possible. In section 6.5 it is shown that if one uses the LU decomposition method, instead of Gaussian elimination, for solving the system of linear equations, there will be no need to
calculate the inverse matrix in order to evaluate (6.43).

The value of \( t_k \) is a question of trial and error. Frequently one chooses the following sequence:

\[
 t_k = \frac{1}{2^i} \quad i = 0, 1, 2, \ldots
\]  
(6.44)

It is obvious that the largest value of \( t_k \) should be taken with which (6.43) is satisfied. Sufficiently close to the solution, (6.43) will be satisfied with \( t_k = 1 \) so that the convergence properties of the classical Newton method are retained in its limit. A final remark should be made on termination criteria for modified Newton methods. Usually one can use:

\[
 \left\| w^{k+1} - w^k \right\| < \epsilon \cdot \left\| w^{k+1} \right\|
\]  
(6.45)

\( \epsilon \) is a properly chosen relative accuracy parameter. However, since the orders of magnitude of the electrostatic potential and carrier densities are completely different, in the present program relative accuracy of each vector of dependent variables is separately examined:

\[
\begin{align*}
\left\| \psi^{k+1} - \psi^k \right\| &< \epsilon_\psi \cdot \left\| \psi^{k+1} \right\| \\
\left\| n^{k+1} - n^k \right\| &< \epsilon_n \\
\left\| p^{k+1} - p^k \right\| &< \epsilon_p
\end{align*}
\]  
(6.46)

A suitable set of values for the relative accuracy parameters in our case are found to be:

\[
\epsilon_\psi = 10^{-4} \cdot N, \quad \epsilon_n = \epsilon_p = 10^{-3} \cdot N
\]

(6.46) alone may be inadequate, since it can be satisfied far too early in the case of a strongly damped iteration scheme. In this case, one can check directly the residual of the non-linear system as an additional criterion:

\[
\left\| F(w^{k+1}) \right\| < \delta_F
\]  
(6.47)

\( \delta_F \) is a properly chosen absolute accuracy. It has to be defined in consistency with the scaling of the system of equations. For the scaling factors given in Table 6.1 a value of \( \delta_F = 2 \times 10^{-4} \cdot N \) is found to be appropriate. In order to terminate the modified Newton iteration, all the conditions (6.43), (6.46) and (6.47) should be met simultaneously.

### 6.4.2. Matrix Form of the Equations

In order to obtain the dependant variables \( \psi, n, \) and \( p \) at discrete grid points, the linearised system of equations (6.40) should be solved repeatedly until convergence is achieved. However, since the variable at each grid point is only affected by the variables at that grid and
the two adjacent grids, from the numerical calculation viewpoint, it is quite advantageous to use another ordering method for the vector of unknowns \( w \) rather than the one used in (6.39b). The optimum ordering is based on mesh numbers instead of dependent variables, i.e., \( w = [\psi_1, n_1, p_1, \psi_2, n_2, p_2, \ldots, \psi_N, n_N, p_N]^T \). The final matrix form of (6.40) will be in the form (6.48).

\[
\begin{bmatrix}
\frac{\partial f_{\psi_1}}{\partial \psi_1} & \frac{\partial f_{\psi_1}}{\partial n_1} & \frac{\partial f_{\psi_1}}{\partial p_1} & \cdots & \frac{\partial f_{\psi_1}}{\partial \psi_N} & \frac{\partial f_{\psi_1}}{\partial n_N} & \frac{\partial f_{\psi_1}}{\partial p_N} \\
\frac{\partial f_{n_1}}{\partial \psi_1} & \frac{\partial f_{n_1}}{\partial n_1} & \frac{\partial f_{n_1}}{\partial p_1} & \cdots & \frac{\partial f_{n_1}}{\partial \psi_N} & \frac{\partial f_{n_1}}{\partial n_N} & \frac{\partial f_{n_1}}{\partial p_N} \\
\frac{\partial f_{p_1}}{\partial \psi_1} & \frac{\partial f_{p_1}}{\partial n_1} & \frac{\partial f_{p_1}}{\partial p_1} & \cdots & \frac{\partial f_{p_1}}{\partial \psi_N} & \frac{\partial f_{p_1}}{\partial n_N} & \frac{\partial f_{p_1}}{\partial p_N} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
\frac{\partial f_{\psi_{N-1}}}{\partial \psi_1} & \frac{\partial f_{\psi_{N-1}}}{\partial n_1} & \frac{\partial f_{\psi_{N-1}}}{\partial p_1} & \cdots & \frac{\partial f_{\psi_{N-1}}}{\partial \psi_N} & \frac{\partial f_{\psi_{N-1}}}{\partial n_N} & \frac{\partial f_{\psi_{N-1}}}{\partial p_N} \\
\frac{\partial f_{n_{N-1}}}{\partial \psi_1} & \frac{\partial f_{n_{N-1}}}{\partial n_1} & \frac{\partial f_{n_{N-1}}}{\partial p_1} & \cdots & \frac{\partial f_{n_{N-1}}}{\partial \psi_N} & \frac{\partial f_{n_{N-1}}}{\partial n_N} & \frac{\partial f_{n_{N-1}}}{\partial p_N} \\
\frac{\partial f_{p_{N-1}}}{\partial \psi_1} & \frac{\partial f_{p_{N-1}}}{\partial n_1} & \frac{\partial f_{p_{N-1}}}{\partial p_1} & \cdots & \frac{\partial f_{p_{N-1}}}{\partial \psi_N} & \frac{\partial f_{p_{N-1}}}{\partial n_N} & \frac{\partial f_{p_{N-1}}}{\partial p_N} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
\frac{\partial f_{\psi_{N_1}}}{\partial \psi_1} & \frac{\partial f_{\psi_{N_1}}}{\partial n_1} & \frac{\partial f_{\psi_{N_1}}}{\partial p_1} & \cdots & \frac{\partial f_{\psi_{N_1}}}{\partial \psi_N} & \frac{\partial f_{\psi_{N_1}}}{\partial n_N} & \frac{\partial f_{\psi_{N_1}}}{\partial p_N} \\
\frac{\partial f_{n_{N_1}}}{\partial \psi_1} & \frac{\partial f_{n_{N_1}}}{\partial n_1} & \frac{\partial f_{n_{N_1}}}{\partial p_1} & \cdots & \frac{\partial f_{n_{N_1}}}{\partial \psi_N} & \frac{\partial f_{n_{N_1}}}{\partial n_N} & \frac{\partial f_{n_{N_1}}}{\partial p_N} \\
\frac{\partial f_{p_{N_1}}}{\partial \psi_1} & \frac{\partial f_{p_{N_1}}}{\partial n_1} & \frac{\partial f_{p_{N_1}}}{\partial p_1} & \cdots & \frac{\partial f_{p_{N_1}}}{\partial \psi_N} & \frac{\partial f_{p_{N_1}}}{\partial n_N} & \frac{\partial f_{p_{N_1}}}{\partial p_N} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
\end{bmatrix}
\]

As can be seen from (6.48), the Jacobian matrix \( F'(w^k) \) is a band matrix with bandwidth of ten. There are very efficient algorithms for solving systems of equations comprised of a band matrix as the matrix of coefficients, in terms of computation time and storage requirements. Since using direct solution methods (e.g. Gaussian elimination and LU decomposition) for a linear system of equation does not introduce any fill-in outside the bandwidth of a band matrix, no process needs to be done on the zero elements outside the bandwidth. Therefore,
the computational effort is reduced significantly. Moreover, there is no need to store the zero elements and storage of 11 vectors of rank 3N will be required rather than a 3N × 3N matrix.

For the evaluation of the Jacobian matrix, the derivatives of \( \mathbf{F} \) with respect to \( \psi, n, \) and \( p \) must be explicitly calculated in order to guarantee the convergence of the iteration. In particular, \( \gamma_n \) and \( \gamma_p \) are considered as functions of \( n \) and \( p \), respectively; their derivatives can be evaluated explicitly as follows:

\[
\frac{\partial \gamma_n}{\partial n} = \gamma_n \left[ 1 - \frac{n}{N_C} \right] \left( F^{-1/2} \left( F_{1/2}^{-1} \left( \frac{n}{N_C} \right) \right) \right) \tag{6.49}
\]

\[
\frac{\partial \gamma_p}{\partial p} = \gamma_p \left[ 1 - \frac{p}{N_V} \right] \left( F^{-1/2} \left( F_{1/2}^{-1} \left( \frac{p}{N_V} \right) \right) \right) \tag{6.50}
\]

Also from (6.9) and (6.10) one can write:

\[
\frac{\partial \psi_n}{\partial n} = V_T \cdot \frac{\partial \gamma_n}{\partial n} = V_T \left[ 1 - \frac{1}{N_C \cdot F^{-1/2} \left( F_{1/2}^{-1} \left( \frac{n}{N_C} \right) \right)} \right] \tag{6.51}
\]

\[
\frac{\partial \psi_p}{\partial p} = \frac{-V_T}{\gamma_p} \cdot \frac{\partial \gamma_p}{\partial p} = -V_T \left[ 1 - \frac{1}{P \cdot N_V \cdot F^{-1/2} \left( F_{1/2}^{-1} \left( \frac{p}{N_V} \right) \right)} \right] \tag{6.52}
\]

Derivatives of the recombination rate with respect to \( \psi, n, \) and \( p \) should also be calculated explicitly for the evaluation of the Jacobian matrix. As an example of the final discretised form of the basic equations the electron continuity equation at a point just before a heterointerface is presented here:

\[
F_{n_i} = \left( n_i \cdot \frac{\mu_{n,i-1/2}}{h_{i-1}} \cdot B \left( \frac{\psi_{n,i} - \psi_{n,i-1}}{V_T} \right) \right) \cdot n_{i-1} \cdot \frac{V_T \cdot \mu_{n,i-1/2}}{h_{i-1}} \cdot B \left( \frac{\psi_{n,i-1} - \psi_{n,i}}{V_T} \right)

+ A \cdot T^2 \cdot (1 + \delta_j) \cdot \exp \left( \frac{E_{C,i} - E_{C,max}}{kT} \right) \cdot \exp \left( F_{1/2}^{-1} \left( \frac{n_i}{N_{C,i}} \right) \right)

- A \cdot T^2 \cdot (1 + \delta_j) \cdot \exp \left( \frac{E_{C,i+1} - E_{C,max}}{kT} \right) \cdot \exp \left( F_{1/2}^{-1} \left( \frac{n_{i+1}}{N_{C,i+1}} \right) \right)

+ \frac{h_i + h_{i-1}}{2} \cdot (a_i \cdot R_{i+1} + b_i \cdot R_i + c_i \cdot R_{i-1}) = 0 \tag{6.53}
\]

where, in fact, the ratio \( \left[ n_i / N_{C,i} \cdot \gamma_{n,i} \right] \) is replaced by \( \exp \left( F_{1/2}^{-1} \left( \frac{n_i}{N_{C,i}} \right) \right) \). The partial derivative of the above equation with respect to \( n_i \) may be written as:
\[
\frac{\partial F_{n_i}}{\partial n_i} = \mu_{n,i-\frac{1}{2}} n_i \cdot B \left( \frac{\psi_{n,i} - \psi_{n,i-1}}{V_T} \right) + n_{i-1} \cdot B \left( \frac{\psi_{n,i-1} - \psi_{n,i}}{V_T} \right) \frac{\partial \psi_{n,i}}{\partial n_i} + \\
\frac{V_T \cdot \mu_{n,i-\frac{1}{2}}}{h_{i-1}} B \left( \frac{\psi_{n,i} - \psi_{n,i-1}}{V_T} \right) + \frac{h_i + h_{i-1}}{2} b_i \cdot \frac{\partial R_i}{\partial n_i} + \\
A^* \cdot T^2 \cdot (1 + \delta_j) \cdot \exp \left( \frac{E_{C,i} - E_{C,\text{max}}}{kT} \right) \cdot \exp \left( F_{i/2} \left( n_i / N_{C,i} \right) \right) \frac{N_{C,i} \cdot F_{i/2}^{-1} \left( n_i / N_{C,i} \right)}{N_{C,i} \cdot F_{i/2}^{-1} \left( n_i / N_{C,i} \right)}
\]

Here \( B' \) is the derivative of the Bernoulli function which again must be evaluated with special care.

### 6.4.3. The Initial Guess Problem

One of the key steps in using the Newton (or Newton-like) method is finding an appropriate initial guess. The initial guess is usually taken as the solution at the previous bias or the solution at the equilibrium, if the solution for the first bias is sought. At thermal equilibrium, due to the fact that the system Fermi level is flat and known (assumed to be zero in the present case, see (4.24)), the three variables at each grid point is reduced to only one, i.e., \( \psi \); the solution of Poisson's equation alone is enough to obtain the solution for the system. The initial guess to the Poisson's equation at equilibrium is obtained by assuming that charge quasi-neutrality exists everywhere in the system, thus \( \psi \) can be solved analytically using (4.33)-(4.35) and

\[
\begin{align*}
\psi &= V_T \cdot \ln \left( \frac{n_0}{(\omega / \omega_r) \cdot \gamma_n \cdot n_i} \right) \\
\psi &= V_T \cdot \ln \left( \frac{\rho_0 \cdot (\omega / \omega_r)}{\gamma_p \cdot n_i} \right)
\end{align*}
\]

The Jacobian matrix for the Poisson's equation is a three diagonal \( N \times N \) matrix. The derivative of the Poisson's equation with respect to potential should include the terms corresponding to the partial derivatives with respect to \( n \) and \( p \), i.e.,

\[
\frac{df_{\psi_i}}{d\psi_j} = \frac{\partial f_{\psi_i}}{\partial \psi_j} + \frac{\partial f_{\psi_i}}{\partial n_j} \frac{\partial n_j}{d\psi_j} + \frac{\partial f_{\psi_i}}{\partial p_j} \frac{\partial p_j}{d\psi_j}, \quad j = i-1, i, i+1
\]

A different approach for initial guess can be applied which allows one to maintain the advantages of a successive procedure, assures convergence, and reduces the computation time in some circumstances. This approach is called STEPSOLVING, since it reaches the desired solution for an arbitrary bias point in a few steps of a hierarchical nature. Further details about this method can be found in [370].
6.5. Solution of the Linear System of Equations

For the solution of non-linear equations representing the discretised semiconductor equations it is required to solve repeatedly a linear system of algebraic equations. Since the number of equations in this system is generally large, their solution is a major problem in itself. Methods of solution belong essentially to either the class of direct methods or the class of iterative methods. If the matrix of coefficients can be entered into the fast storage unit of the computer, then the direct methods are quicker and more accurate than iterative methods [371].

Direct methods solve the system of equations in a known number of arithmetic operations, and errors in the solution arise entirely from rounding errors introduced during the computation. Basically, these direct methods are elimination methods of which the best known examples are the systematic Gaussian elimination method and the triangular decomposition method. The latter method factorises the matrix $A$ of the equation $A \cdot x = b$ into $A = L \cdot U$, where $L$ and $U$ are lower and upper triangular matrices, respectively. In this method, once the decomposition has been determined, the solution is calculated from $L \cdot (U \cdot x) = b$ by putting $U \cdot x = y$ and then solving $L \cdot y = b$ for $y$ by forward substitution and $U \cdot x = y$ for $x$ by backward substitution.

Although Gaussian elimination and LU decomposition are mathematically equivalent, the overall errors of the LU method can be made smaller than those of the Gaussian elimination method [371]. More importantly, if the RHS vector, $b$, in the LU method is replaced by another vector, the decomposition process for the matrix $A$ (which is the most time-consuming part of the solution procedure) does not need to be repeated; only forward and backward substitution should be redone to find the new vector of unknowns. This is in fact happening in the evaluation of the condition (6.43) for the modified Newton method, as mentioned earlier in section 6.4.1. The RHS of (6.43) is indeed $\|w^{k+1} - w^k\|$ as already calculated by solving (6.40). Considering (6.40) and (6.42), the LHS of (6.43) can also be written as $\|F'(w^k)^{-1} \cdot F(w^{k+1})\|$. If one defines:

$$F'(w^k)^{-1} \cdot F(w^{k+1}) = v$$

then $v$ can be found by solving

$$F'(w^k) \cdot v = F(w^{k+1})$$

(6.57)

which is very similar to (6.40) only with a different RHS vector. Therefore, the LU decomposition of the Jacobian matrix does not need to be repeated, which is saving a lot of computation time. The latter is the main reason why LU decomposition is adopted in the present work. In our case, since the Jacobian matrix is a band matrix with bandwidth of ten, both $L$ and $U$ are band matrices with bandwidth of five [372].
6.6. Terminal Characteristics

6.6.1. Calculation of Terminal Currents

The direct evaluation of current density from Eqs. (4.15) and (4.16) may be very inaccurate due to cancellation errors in evaluating the derivative of large numbers (say the majority carrier concentration). Fundamentally this phenomenon stems from the relatively large residue in the final solution of the non-linear system even when double precision arithmetic is used. In order to avoid the above problem, the present work evaluates the current density directly only for the minority carriers at the terminal. The calculation of the current density at the next interval for the same carriers is then done by applying the continuity principle and using the known recombination rate at every grid point. The process is repeated until the next terminal, where those carriers become majority, is reached. If \( M \) is the index of the base contact grid point, the current densities \( J_{p,1+1/2}, J_{p,N-1/2}, \) and \( J_{n,M \pm 1/2} \) are first evaluated by Eqs. (4.15) and (4.16), because all of them are minority carrier currents at those nodes. The remaining current densities are then computed by integration as follows:

\[
J_{n,M-1/2} - J_{n,1+1/2} = \int_{z_{1+1/2}}^{z_{M-1/2}} \mathbf{R} \cdot dz = \sum_{i=2}^{M-1} \int_{z_{i-1/2}}^{z_{i+1/2}} \mathbf{R} \cdot dz
\]

\[
J_{n,1+1/2} = J_{n,M-1/2} - \sum_{i=2}^{M-1} \left[ (a_i \cdot R_{i+1} + b_i \cdot R_i + c_i \cdot R_{i-1}) \cdot \frac{h_{i-1} + h_i}{2} \right]
\]

(6.58)

Similarly, one can show that:

\[
J_{n,N-1/2} = J_{n,M+1/2} + \sum_{i=M+1}^{N-1} \left[ (a_i \cdot R_{i+1} + b_i \cdot R_i + c_i \cdot R_{i-1}) \cdot \frac{h_{i-1} + h_i}{2} \right]
\]

(6.59)

The terminal currents are then given by:

\[
I_E = -A_E \cdot (J_{n,1+1/2} + J_{p,1+1/2})
\]

(6.60)

\[
I_C = -A_E \cdot (J_{n,N-1/2} + J_{p,N-1/2})
\]

(6.61)

\[
I_B = I_E - I_C
\]

(6.62)

In the case of current-controlled base ohmic contact, the base current is known. Thus one can use (6.60) to find \( I_E \) and then: \( I_C = I_E - I_B \).

In order to demonstrate the inaccuracy of the terminal current calculation directly from (4.15) and (4.16), the base current in a common-emitter output characteristics simulation is calculated by the following two additional methods (remember that in this case the base current is constant and known, 100\(\mu\)A in the present simulation):

(i) \( I_E \) and \( I_C \) are found using the above method from (6.60) and (6.61), and then \( I_B \) is
calculated from (6.62).

(ii) $I_B$ is calculated from $-A_E \left( J_{n,M-1/2} + J_{p,M-1/2} - J_{n,M+1/2} - J_{p,M+1/2} \right)$, where $J_{p,M \pm 1/2}$ are evaluated directly from (4.16).

The relative error for the calculated $I_B$ using the above two methods are plotted as a function of $V_{CE}$ in figure 6.4. While the relative error of the first method is always smaller than 0.01%, that of the second method can be as large as 100%! This demonstrates the accuracy of the present current calculation scheme and incapability of the direct evaluation method.

A very accurate approach to compute the terminal current in two- and three-dimensional simulations is suggested in [373] and modified for use in bipolar device modelling by the author [374] which, unfortunately, is not applicable to 1-D case.

It is important to notice that one can not use the hole continuity equation between the two mesh mid-points $(M+1/2)$ and $(M-1/2)$, since a non-zero base current is flowing out of the system. In other words:

$$J_{p,M-1/2} - J_{p,M+1/2} \neq \int_{Z_{M-1/2}}^{Z_{M+1/2}} R \cdot dz$$

$$J_{n,M+1/2} + J_{p,M+1/2} \neq J_{n,M-1/2} + J_{p,M-1/2}$$

Therefore, calculations of $J_{n,M+1/2}$ and $J_{n,M-1/2}$ should start from $J_{n,M-1/2}$ and $J_{n,M+1/2}$, respectively.

### 6.6.2. Small-Signal Equivalent Circuit Parameters

This subsection states the extraction of device parameters from the known distribution of $\psi$, $n$, and $p$. The parameters discussed here are the common-emitter current gain, junction capacitance, sheet resistance, delay times, and the cut-off frequency, which, in turn, can be used to extract the equivalent circuit parameters necessary for circuit simulators like SPICE.

The DC or large-signal current gain $\beta_{dc}$ is simply defined as the ratio of collector current to the base current. The AC or small-signal current gain $\beta_{ac}$ is defined as:

$$\beta_{ac} = \frac{\Delta I_C}{\Delta I_B} \bigg|_{V_{CE}=\text{cte}}$$

(6.63)
Chapter 6

Numerical Methods

The sheet resistance of any layer with the same type of net impurity is calculated from:

\[
R_{sh} = \frac{1}{\int_{Z_{j1}}^{Z_{j2}} q \cdot (p \cdot \mu_{p,LF} + n \cdot \mu_{n,LF}) \, dz}
\]

(6.64)

where \( R_{sh} \) is in unit of \( \Omega/\text{square} \) and \( Z_{j1} \) and \( Z_{j2} \) are the positions of the physical junctions determined from doping profile. Calculation of sheet resistance may be useful in finding the base and sub-collector series resistances. It is important to use the low-field mobilities in the above equation, since the electric field in the \( z \)-direction is not going to modify the sheet resistance which is determined by drift of the carriers parallel to the surface.

The cut-off frequency \( f_T \) is defined as the frequency at which \( \beta_{ac} \) falls to unity. It is related to the emitter-to-collector delay time \( \tau_{EC} \) by the following equation (see Chap. 10):

\[
f_T = \frac{1}{2\pi \cdot \tau_{EC}}
\]

(6.65)

\( \tau_{EC} \) can, in turn, be calculated at certain bias conditions from [50,375]:

\[
\tau_{EC} = \left. \frac{\Delta Q}{\Delta J_C} \right|_{V_{CE}=\text{cte}}
\]

(6.66)

where \( \Delta Q \) is the incremental charge of one type (n or p) in the entire region of the transistor due to the incremental current density \( \Delta J_C \). Emitter charging time (\( \tau_E \)), base transit time (\( \tau_B \)), and collector delay time (\( \tau_C \)) can be found from [126]:

\[
\tau_E = \left. \frac{\Delta Q_E}{\Delta J_C} \right|_{V_{CE}=\text{cte}} \quad \tau_B = \left. \frac{\Delta Q_B}{\Delta J_C} \right|_{V_{CE}=\text{cte}} \quad \tau_C = \left. \frac{\Delta Q_C}{\Delta J_C} \right|_{V_{CE}=\text{cte}}
\]

(6.67)

where \( \Delta Q_E, \Delta Q_B, \) and \( \Delta Q_C \) are the increments in electron charge density inside emitter, base, and collector, respectively. In figure 6.5, the numerically calculated values of base transit time from (6.67) are plotted against the base width, \( W_B \), for an InGaP/GaAs SHBT with base doping of \( 2 \times 10^{19} \text{ cm}^{-3} \). Also shown in figure 6.5 is the plot of \( \tau_B \) calculated from the following analytical equation [121,376]:

\[
\tau_B = \frac{W_B^2}{2\mu_{nB} \cdot V_T} + \frac{W_B}{v_{n,\text{sat}}}
\]

(6.68)

with \( \mu_{nB} = 1320 \text{ cm}^2/\text{V} \cdot \text{s} \) and \( v_{n,\text{sat}} = 1 \times 10^7 \text{ cm/s} \) as given in chapter 5 for GaAs base doped to \( 2 \times 10^{19} \text{ cm}^{-3} \). It can be seen that Eq. (6.68) gives an excellent fit to numerically calculated base transit times down to the base thickness of 100\( \AA \).

The capacitance of a p-n junction is calculated as the ratio of the incremental charge of one
type of carrier in the entire region of the p-n junction to the incremental voltage applied to it. Since the base region of an Npn HBT is highly doped, calculation of hole charge increments inside the base is prone to large errors. Therefore, calculation based on electron charges produces more accurate capacitance results for Npn HBTs. For example, the base-emitter capacitance of the transistor will be evaluated as:

\[ C_{BE} = \frac{\Delta Q_{nBE}}{\Delta V_{BE}} \bigg|_{V_{CB}=\text{cte}} \]  

For a forward biased B-E junction, this capacitance includes the dielectric junction capacitance \( C_{jBE} \) as well as the so-called diffusion capacitance \( C_{dBE} \), which are approximately partitioned as follows:

\[ C_{jBE} = \frac{\Delta Q_{nE}}{\Delta V_{BE}} \bigg|_{V_{CB}=\text{cte}} \quad C_{dBE} = \frac{\Delta Q_{nB}}{\Delta V_{BE}} \bigg|_{V_{CB}=\text{cte}} \]  

Some further comments are necessary as to the calculation of the integral in the denominator of (6.64) and the incremental charges in (6.66)-(6.69). The incremental electron charge inside the layer \( L \) of the device may be expressed as:

\[ Q_n(L) = q \cdot \int_{z_i}^{z_{i+1}} n \cdot dz = q \cdot \sum_{i=K+1}^{K+N_{\text{layer}}(L)-1} \int_{z_i}^{z_{i+1}} n \cdot dz \]  

where \( N_{\text{layer}}(L) \) is the number of mesh points in layer \( L \), and \( K \) is the total number of mesh points in the previous layers:

\[ K = \sum_{\ell=1}^{L-1} N_{\text{layer}}(\ell) \]

Also for calculating the denominator of (6.64) one needs to evaluate the following integral:

\[ \int_{z_{j1}}^{z_{j2}} n \cdot \mu_{n,LF} \cdot dz = \sum_{i=K+1}^{K+N_{\text{layer}}(L)-1} \int_{z_i}^{z_{i+1}} n \cdot \mu_{n,LF} \cdot dz = \sum_{i=K+1}^{K+N_{\text{layer}}(L)-1} \int_{z_i}^{z_{i+1}} n \cdot \mu_{n,LF,i+1/2} \cdot dz \]  

where a constant mobility between the two neighbouring mesh points is assumed. Both of the
above problems reduce to the calculation of the integral of carrier densities between two neighbouring mesh points. This integral can not be evaluated using conventional integration approaches such as trapezoidal method, since these methods assume a linear variation of \( n \) inside the interval \([z_j, z_{j+i}]\), which is in contrast to the exponential variation obtained in (6.23). Instead, a direct integration of (6.23) must be employed. Simple mathematical derivation yields:

\[
\int_{z_i}^{z_{i+1}} n \cdot dz = h_i \cdot \left[ n_i \cdot \left( n_{i+1} - n_i \right) \cdot G \left( \frac{\Psi_{n,i+1} - \Psi_{n,i}}{V_T} \right) \right]
\]  

(6.72)

where the function \( G(u) \) is defined by:

\[
G(u) = \frac{\exp(u) - u - 1}{u \cdot [\exp(u) - 1]}
\]  

(6.73)

Special care must be taken in the implementation of \( G(u) \) to avoid potential overflow and underflow traps due to the calculation of exponential terms for large values of \( u \) or values close to zero. One possible evaluation scheme is as follows:

\[
G(u) = \begin{cases} 
\frac{u + 1}{u} & u < -U_1 \\
\frac{u}{u + 3} & -U_2 < u < U_2 \\
\frac{u^2 + 3u + 6}{u} & u > U_1 \\
\frac{1}{u \cdot [\exp(u) - 1]} & \text{otherwise}
\end{cases}
\]  

(6.74)

\( U_1 \) and \( U_2 \) are properly chosen large and small positive numbers, respectively. For the condition \( \Psi_{n,i+1} = \Psi_{n,i} \), (6.72) simply reduces to the trapezoidal integration:

\[
\int_{z_i}^{z_{i+1}} n \cdot dz = h_i \cdot \left( \frac{n_{i+1} + n_i}{2} \right)
\]

Following similar guidelines one can easily obtain:

\[
\int_{z_i}^{z_{i+1}} p \cdot dz = h_i \cdot \left[ p_i \cdot \left( p_{i+1} - p_i \right) \cdot G \left( \frac{\Psi_{p,i} - \Psi_{p,i+1}}{V_T} \right) \right]
\]  

(6.75)

The above formulae must not be used for the calculation of integral of carrier densities across a heterojunction interface, since (6.23) and (6.26) are derived based on the drift-diffusion approach which is not valid at a heterointerface.
In figure 6.6 the calculated base sheet resistance of an HBT with 800Å and 1000Å GaAs base is plotted against the base doping concentration at thermal equilibrium. This plot may be used to interpret the base TLM results in terms of the base doping density. The calculated base sheet resistance showed a slight dependence on bias which may be ignored.

6.7. Flowchart of the Completed Program

To summarise this chapter a flow diagram of the developed numerical code for the simulation of HBTs is presented in Fig. 6.7. The program starts by reading the input parameters, including layer structure, temperature, and biasing conditions. Then a non-uniform mesh is generated and the values of the physical parameters are calculated at the discretised mesh points. A very initial solution is formed by assuming thermal equilibrium and charge neutrality in the entire device, followed by normalisation of equations, parameters, and variables. Then the thermal equilibrium solution is calculated to serve as an initial guess. The biasing condition is modified according to the required characteristics. To solve the equations under the new condition, the Jacobian matrix of coefficients is first formed, and then solved by the LU technique to find the correction vector $\Delta w(k)$. Then, the RHS of the condition (6.43) is evaluated, the variables are modified, and all the other parameters are refreshed. The condition (6.43) is verified by evaluating the LHS using the modified set of variables. If this condition is not met, the modified Newton inner loop is iterated until the condition is satisfied. Otherwise, the convergence criteria (6.46)-(6.47) are verified. If these conditions are met AND none of the modified carrier densities at discretised mesh points are negative, the terminal parameters are calculated and the bias is modified until the last point of the characteristics is calculated. Otherwise, a new Jacobian matrix is formed using the upgraded set of variables and parameters, and the Newton outer iteration loop is repeated until convergence is achieved.
Figure 6.7- Flow chart of the developed numerical code for the simulation of HBTs.
CHAPTER 7. NUMERICAL SIMULATION RESULTS

Basic semiconductor equations, the methods to solve them numerically, and the physical parameters involved in them were discussed in full details in chapters 4-6. Based on those discussions, a 1-D numerical simulation code for modelling the physical behaviour of III-V HBTs was developed by the author of the present thesis. In this chapter some of the results obtained from this numerical simulation package are presented. Other results of numerical simulation are given in other chapters of this work wherever necessary. In section 7.1, the high field mobility model for the forward-biased B-E junction, which is a matter of controversy among published literature, is discussed in detail. Then, importance of thermionic-emission and tunnelling to the I-V characteristics of abrupt B-E HBTs is highlighted in a second example in section 7.2. Section 7.3 deals with the degradation of device DC characteristics due to the traps introduced at the B-E interface of various types of HBTs. It is demonstrated that an abrupt B-E HBT, and in particular an InGaP/GaAs HBT, performs superior to the graded AlGaAs/GaAs HBTs as far as B-E interface recombination is concerned. Temperature dependence of current gain in various types of HBTs is studied thoroughly in section 7.4. The contribution of each component of the base current to the temperature dependence of current gain is discussed separately. To the best of author's knowledge, this study is the most comprehensive amongst those published earlier. Next, the DC and high frequency performance of graded base GaAs HBTs is optimised using numerical simulation in section 7.5. The graded base of such devices can be realised using either Al\textsubscript{x}Ga\textsubscript{1-x}As or In\textsubscript{1-x}Ga\textsubscript{x}As, and it is shown that only lightly grading the base is necessary to improve the performance of abrupt emitter InGaP/GaAs and graded emitter AlGaAs/GaAs HBTs. Section 7.6 deals with the capacitance calculation for both forward- and reverse-biased junctions. Finally, the theory of Kirk effect is illustrated in section 7.7 using numerical simulation results. It is shown that the occurrence of Kirk effect in III-V HBTs does not necessarily lead to a sharp current gain fall-off as in Si BJTs.

7.1. Driving Force for Velocity Saturation

It is well-known that the velocity of electrons saturates at the reverse-biased B-C junction. This has been modelled by the electron velocity-field relationship as given by (5.67). But, it should be noted that the use of the field-dependent electron mobility also limits the velocity of electrons to the saturation velocity in the forward-biased B-E junction, where the electrons may not be heated as in the bulk case since they diffuse against the field across the emitter depletion region. This problem has been pointed out by various authors [56,206,328], and some solutions to avoid it are suggested. In order to investigate the effect of using different electron velocity versus field characteristics on the device terminal behaviour, four possible approaches are considered here:

Model 0: When the carriers are diffusing against the field, i.e., \( J_n \cdot E_n < 0 \) or \( J_p \cdot E_p < 0 \), use
the low-field mobility. This approach is used in early versions of device simulation software MINIMOS [281].

**Model 1**: Employ the normal velocity saturation formulae of (5.67) and (5.69) in the entire device, irrespective of the fact that carriers may move against the field force in some regions.

**Model 2**: When the electrons are diffusing against the field, use the normal velocity saturation formula of (5.67), but with the saturation velocity $v_{n,sat}$ replaced by the thermal velocity of electrons defined as $v_{th} = \sqrt{3kT/m_n^*}$. Some physical justification for this choice of saturation velocity is given in [56,377].

**Model 3**: Instead of the effective electric fields $E_n$ and $E_p$ in (5.67) and (5.69), use effective driving forces, $F_n$ and $F_p$, defined as [206,328]:

$$J_n = q\mu_n F_n \quad J_p = q\mu_p F_p$$

The above definitions together with (4.9)-(4.10), (4.15)-(4.16), (5.68), and (5.70) result in:

$$F_n = \frac{1}{q} \frac{dE_{fn}}{dz} = E_n + \frac{V_T}{n} \frac{dn}{dz}$$

$$F_p = \frac{1}{q} \frac{dE_{fp}}{dz} = E_p - \frac{V_T}{p} \frac{dp}{dz}$$

### Table 7.1- Layer structure of the InGaP/GaAs DHBT wafer no. EPI 1-3514A.

<table>
<thead>
<tr>
<th>Material</th>
<th>Thickness (Å)</th>
<th>Doping (cm$^{-3}$)</th>
<th>Dopant</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n^+$-GaAs</td>
<td>2800</td>
<td>$4\times10^{18}$</td>
<td>Si</td>
<td>Cap layer</td>
</tr>
<tr>
<td>$n^+$-In$<em>{0.49}$Ga$</em>{0.51}$P</td>
<td>200</td>
<td>$2\times10^{18}$</td>
<td>Si</td>
<td>Cap layer</td>
</tr>
<tr>
<td>n-In$<em>{0.49}$Ga$</em>{0.51}$P</td>
<td>1000</td>
<td>$3\times10^{17}$</td>
<td>Si</td>
<td>Emitter</td>
</tr>
<tr>
<td>p$^-$-GaAs</td>
<td>1000</td>
<td>$3\times10^{19}$</td>
<td>C</td>
<td>Base</td>
</tr>
<tr>
<td>n$^+$-GaAs</td>
<td>200</td>
<td>$1\times10^{16}$</td>
<td>Si</td>
<td>Spacer</td>
</tr>
<tr>
<td>n-In$<em>{0.49}$Ga$</em>{0.51}$P</td>
<td>4800</td>
<td>5-7$\times10^{16}$</td>
<td>Si</td>
<td>Collector</td>
</tr>
<tr>
<td>n$^+$-In$<em>{0.49}$Ga$</em>{0.51}$P</td>
<td>1000</td>
<td>$2\times10^{18}$</td>
<td>Si</td>
<td>Sub-collector</td>
</tr>
<tr>
<td>n$^+$-GaAs</td>
<td>7000</td>
<td>$4\times10^{18}$</td>
<td>Si</td>
<td>Sub-collector</td>
</tr>
<tr>
<td>S.I. GaAs</td>
<td>635μm</td>
<td>undoped</td>
<td>—</td>
<td>Substrate</td>
</tr>
</tbody>
</table>

The above models have been applied for the simulation of an InGaP/GaAs DHBT, as specified in Table 7.1, and the simulated Gummel plots are compared with the measured data in Fig. 7.1. The device has an emitter area of $8\times10$ μm$^2$. The emitter, collector, and base series resistances of the device are measured, using techniques described in chapter 9, as 16, 14, and 20Ω, respectively. Since the above resistances are partly due to bulk resistances of the materials within the domain of simulation and are already accounted for, slightly smaller resistances are added externally for the purpose of fitting to the experimental data. $R_E = 14Ω$, $R_C = 10Ω$, and $R_B = 20Ω$ are found to give the best fit to the high current region of the experimental data. No other fitting parameter is used for the adjustment of collector current. An excellent fit to the experimental data was observed when Model 0 for the velocity-field
characteristics is adopted. On the other hand, when the normal v-E formulae of (5.67) and (5.69) are used in the entire device (i.e., Model 1), collector current is largely underestimated. Furthermore, Model 1 cannot correctly reproduce the collector ideality factor of the experimental data (~1.1). Model 3 only slightly improves the situation, but Model 2 produces results which are almost similar to those of Model 0. The above trend was found in almost all the simulations performed in the present work; Models 1 and 3 underestimate the experimental data by sometimes more than an order of magnitude, while Model 0 produces the closest results to the measured data. Therefore, it was concluded that using the low-field mobility of electrons and holes for the case of a forward-biased junction, where carriers move against the electric field force, is the most appropriate approach, and will be employed hereafter in this work.

The above underestimation of $I_c$ when Models 1 and 3 are used can be described as follows. According to the analytical formulation of collector current by Datta et al. [121], the electron component of $J_c$ can be expressed in terms of four effective velocities as:

$$J_{nC} = \frac{q_n n_0 (z_{PE}) \exp(qV_{BE} / kT)}{S_{in} + \frac{1}{S_{en}} + \frac{1}{S_{bn}} + \frac{1}{v_{n,sat}}}$$  \hspace{1cm} (7.3)
where $n_0(Z_{PE})$ is the thermal equilibrium electron concentration at the emitter end of the quasi-neutral base, $v_{n,\text{sat}}$ corresponds to the saturation velocity of electrons in the B-C space charge region (assuming the device is biased in forward active mode), $S_{in}$ is the velocity of electrons limited by the thermionic-field-emission at the B-E heterointerface, $S_{en}$ is the electron drift-diffusion velocity in the emitter SCR, and $S_{bn}$ is the effective hole velocity associated with drift-diffusion across the quasi-neutral base region. From (7.3) it is apparent that the smallest of the four effective electron velocities determines the collector current density $J_{nc}$. In the case of hole transport in Pnp HBTs, a similar formulation can be derived. Since the mobility of holes is low, the drift-diffusion velocity of holes inside the emitter SCR is small and comparable to the thermionic-field-emission velocity. Indeed, it has been shown in [121] that the drift-diffusion of holes across the emitter SCR and thermionic-field-emission of holes across the B-E heterointerface are equal in importance, and form the dominant mechanisms limiting the collector current of Pnp HBTs. This is in contrast to the case for electron injection into the base of III-V Npn HBTs where the velocity $S_{in}$ is the smallest, and the contribution of drift-diffusion across the emitter SCR is negligible due to the large mobility of electrons. However, if Model 1 (or Model 3) is used for velocity saturation of electrons, the effective velocity of electrons inside the high-field emitter SCR becomes extremely small, and hence, collector current will be mainly limited by the transport of electrons inside the emitter SCR rather than B-E heterojunction. Since transport across the emitter SCR is due to drift-diffusion, this component of current will have an ideality factor very close to unity, which can not reproduce the ideality factor of the measured $I_C$. Also the measured data will be underestimated. This situation is clearly the same as the cases for Models I and 3 in Fig. 7.1.

While fitting the simulated collector current to the measured data is achieved by adjusting only the series resistances (mainly $R_E$), the measured base current of the device was not exactly reproduced unless recombination parameters were adjusted. This situation was predictable as mentioned in subsection 5.5.4, since the recombination lifetimes are very much dependent on the quality of the materials, and hence significantly change for different growth parameters and process conditions. In order to fit the simulated $I_B$ to the measured data, an effective interface lifetime of $\tau_{\text{int}} = 0.4\,\text{ps}$ is used in a region $\sim 100\,\text{Å}$ thick above the B-E heterointerface (see subsection 5.5.5), and then all the recombination rates in the entire device were scaled by a factor of 0.4 (i.e., lifetimes were multiplied by 2.5). The above interface lifetime may seem unrealistically low, but it should be mentioned that the interface recombination rate in this work includes the effects of other 2-D or 3-D recombination components (such as extrinsic base surface recombination) which may have ideality factors somewhat larger than unity. In order to include such effects more accurately and physically more meaningful, two- or three-dimensional simulations are obviously required.

A constant current base boundary condition can be used to calculate the output characteristics of the above DHBT. As demonstrated in Fig. 7.2, a very reasonable agreement can be observed between the simulated and measured data. However, few points have to be considered...
when simulating the $I_C-V_{CE}$ characteristics of HBTs. First, due to the 1-D nature of the present device simulator code, the difference in the B-C and B-E junction areas can not be taken into account. Therefore, the calculated offset voltage of the device will be underestimated, which sometimes results in negative offset voltages for DHBTs. Secondly, the observed small positive slope of $I_C$ in the forward active region for relatively low current levels is due to the rather weak base-width modulation present in the $p^+$-base of HBTs. At higher current levels, device will be heated and the gain drops as $V_{CE}$ increases. Therefore, the small positive slope of $I_C$ will be overtaken by the negative differential resistance (NDR) due to self-heating. Obviously, to obtain a good fit at higher current levels, one has to consider the variation of lattice temperature with dissipated power in the simulation.

7.2. Thermionic-Field Emission (TFE)

In chapter 4 an expression of thermionic-field emission current formulated based on the WKB approximation was used as a boundary condition at an abrupt heterointerface which eventually limits the current transport over the barrier while maintaining the current continuity. In this section, it will be demonstrated that the role of tunnelling and thermionic-emission in the overall current transport is quite significant. An abrupt InP/InGaAs HBT as specified in Table 7.2 is considered as an example in this section. The fabricated HBT had a circular geometry with B-E junction diameter of 16 μm ($A_{BE} = 2 \times 10^{-6}$ cm$^2$).

Table 7.2- Layer structure of an InP/InGaAs SHBT used as example in this section.

<table>
<thead>
<tr>
<th>Material</th>
<th>Thickness (Å)</th>
<th>Doping (cm$^{-3}$)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n^+<em>\text{In}</em>{0.53}\text{Ga}_{0.47}\text{As}$</td>
<td>1000</td>
<td>$2 \times 10^{19}$</td>
<td>Cap Layer</td>
</tr>
<tr>
<td>$n^+_\text{InP}$</td>
<td>500</td>
<td>$2 \times 10^{18}$</td>
<td>Cap Layer</td>
</tr>
<tr>
<td>$n^-\text{InP}$</td>
<td>1500</td>
<td>$5 \times 10^{17}$</td>
<td>Emitter</td>
</tr>
<tr>
<td>$n^-\text{In}<em>{0.53}\text{Ga}</em>{0.47}\text{As}$</td>
<td>20</td>
<td>$2 \times 10^{16}$</td>
<td>Spacer</td>
</tr>
<tr>
<td>$p^-\text{In}<em>{0.53}\text{Ga}</em>{0.47}\text{As}$</td>
<td>1000</td>
<td>$1.3 \times 10^{19}$</td>
<td>Base</td>
</tr>
<tr>
<td>$n^-\text{In}<em>{0.53}\text{Ga}</em>{0.47}\text{As}$</td>
<td>4000</td>
<td>$2 \times 10^{16}$</td>
<td>Collector</td>
</tr>
<tr>
<td>$n^+<em>\text{In}</em>{0.53}\text{Ga}_{0.47}\text{As}$</td>
<td>3000</td>
<td>$2 \times 10^{19}$</td>
<td>Sub-collector</td>
</tr>
</tbody>
</table>

Fig. 7.3 compares the measured collector current of the InP/InGaAs SHBT with the calculated
currents under three different conditions; first with the full thermionic-field emission (TFE) approach, second with the thermionic-emission (TE) approach neglecting tunnelling at the abrupt B-E heterointerface, and third using the conventional drift-diffusion (DD) approach. In the latter case, both thermionic-emission and tunnelling are ignored at the B-E heterojunction, and instead a continuous quasi-Fermi level is assumed at this interface [53]. The only fitting parameter for the calculation was emitter series resistance, which was set to $6\,\Omega$ for a good fit to the high current region of $I_C$.

As can be seen in Fig. 7.3, the TFE model predicts the closest result to the measured data. The ideality factor for the collector current as calculated by the TFE model is identical to the measured $\eta_C$ of 1.12. A perfect fitting of the TFE model to the measured data can be obtained by using $\Delta E_C(\text{In}_{0.53}\text{Ga}_{0.47}\text{As/InP})$ only 20meV smaller than the default value of 225meV, which would still lie within the often reported range of (185-265)meV (see section 5.3.2.c for more details). On the other hand, the TE model underestimates the measured data at low to moderate current range by more than a factor of 10. Additionally, the ideality factor for the TE model is smaller than that of the experimental data.

In abrupt HBTs, collector ideality factors in the range 1.1-1.2 are usually observed as opposed to $\eta_C = 1$ for graded HBTs. This difference can be explained by the partitioning of the applied bias between emitter and base and its effect on the tunnelling-emission process. If the tunnelling process is neglected, the conduction band barrier height as a function of $V_{BE}$
determines the electron injection by thermionic-emission over the barrier. Using the depletion approximation, the voltage drop in the emitter region is given by [47]:

\[ V_{BE1} = \frac{\varepsilon_B N_{AB}}{\varepsilon_E N_{DE} + \varepsilon_B N_{AB}} \cdot V_{BE} \]  

(7.4)

where \( \varepsilon_B \) and \( \varepsilon_E \) are the base and emitter relative dielectric constants, respectively. The above relation yields the following ideality factor for the collector current:

\[ \eta_C = 1 + \frac{\varepsilon_E N_{DE}}{\varepsilon_B N_{AB}} \]  

(7.5)

which is equal to 1.04 for the present device. Presence of a thin spacer layer between emitter and base regions further reduces the percentage of the voltage drop before the conduction band spike \( V_{BE1} \), and increases \( \eta_C \) to 1.09. When the tunnelling effect is taken into account, the ideality factor increases even further to 1.12 due to the bias dependence of the tunnelling factor, \( \delta \). Variation of the tunnelling factor with bias is also shown in Fig. 7.3, where a significant reduction of \( \delta \) at high currents is observed, similar to the variation reported in [25]. This reduction can be explained by the fact that at large \( V_{BE} \)'s, the potential barrier for the injecting electrons from emitter to base reduces, and more electrons will have the chance to go over the barrier rather than tunnel through it. Consequently, the TE current will be enhanced by a larger factor at lower biases, which effectively results in a larger ideality factor for the TFE model.

![Figure 7.4- The band diagram of an InP/InGaAs SHBT simulated using three different models of thermionic-field emission (TFE), thermionic-emission only (TE), and drift-diffusion (DD). Of particular interest is the quasi-Fermi level discontinuity observed at the B-E heterointerface.](image)

The conventional drift-diffusion model assumes a continuous quasi-Fermi level across an abrupt heterointerface. But as shown in Fig. 7.3, this model results in calculated collector
currents almost two orders of magnitude smaller than the experimental data at low biases. Fig. 7.4 shows the simulated band structure of the device under study for various approaches used in this section. It can be seen that dominance of thermionic-emission at the B-E heterojunction forces a discontinuity in the electron quasi-Fermi level, \( \Delta E_{\text{fn}} \), as also described in [53,169]. Inclusion of tunnelling mechanism reduces \( \Delta E_{\text{fn}} \) to some extent, but using the DD model strangely results in a large \( \Delta E_{\text{fn}} \). A closer inspection of the numeric data used for the graph in Fig. 7.4 reveals that the quasi-Fermi level discontinuity for the DD model is across the two points prior to the heterointerface. In fact, forcing a continuous quasi-Fermi level across the interface has resulted in a larger and unphysical \( \Delta E_{\text{fn}} \) between two points on the emitter side. This results in a significant underestimation of the simulated \( I_C \), and demonstrates that an appropriate TFE model must be employed for any realistic simulation of electron devices including abrupt heterointerfaces. Indeed, it has been shown in [169] that the contribution of thermionic-emission and tunnelling is even more important for the simulation of n-N heterojunctions and at lower temperatures.

7.3. Interface Recombination: Abrupt versus Graded HBTs

The conduction band spike at the B-E heterojunction is known to limit the injection efficiency of abrupt HBTs, especially in the case of AlGaAs/GaAs HBTs where majority of the bandgap difference exists in the conduction band. Therefore, in such structures grading of the Al mole fraction is usually employed to remove this spike and improve the injection efficiency. However, grading of the emitter region next to the base increases the recombination rate inside the emitter SCR, which eventually limits the current gain of graded HBTs. Additionally, the potential spike at the B-E junction of abrupt HBTs may inject electrons into the base region with a substantial kinetic energy, and hence with a very high velocity. Consequently, the resultant velocity overshoot inside the base reduces the base transit time and improves the high frequency characteristics of abrupt HBTs as compared to graded devices. In this section, the effect of interfacial recombination at the B-E heterojunction on the DC current gain (\( \beta \)) of abrupt and graded HBTs is studied using numerical simulation. It will be shown that the DC current gain of compositionally graded AlGaAs/GaAs HBTs is critically affected by the presence of the B-E heterojunction traps, whereas the dependence is much less pronounced for abrupt HBTs.

Three structures, as specified in Table 7.3, are used as examples in this section: a graded AlGaAs/GaAs, an abrupt AlGaAs/GaAs, and an abrupt InGaP/GaAs HBT. The base and collector layers of all three devices are similar; only the material and design of emitter layers are different. To understand the role of traps in the B-E heterojunction region, a thin layer with a thickness of \( t_r \) and very short lifetime of 20ps is assumed just above the base layer. The calculated current gains as a function of \( t_r \) at several current density levels are shown in Fig. 7.5. Emitter, collector, and base series resistances of 5, 5, and 20\( \Omega \), respectively, were used in the calculation, and the active area of the device was assumed as 100 \( \mu \text{m}^2 \). It is seen that for
the graded AlGaAs structure $\beta$ is reduced drastically as $t_r$ increases. The interface layer is likely to be related to the necessity of a growth interruption step between the heavily doped base and the AlGaAs emitter, during which oxygen and other impurity species are adsorbed on the surface and become efficient traps when Al is subsequently introduced. The sensitive dependence of $\beta$ on the interface quality may explain the difficulty in achieving current gain reproducibility for high performance AlGaAs/GaAs HBTs. The current gain of the abrupt AlGaAs/GaAs HBT is less sensitive to interfacial recombination than its graded counterpart, which is explained as follows. The recombination rate is proportional to the intrinsic carrier density, $n_i$, inside the SCR, which extends mainly on the emitter side. Compositional grading near this region leads to a larger $n_i$ and, thus, more recombination. In an abrupt structure, on the contrary, $n_i$ is small over the entire emitter region. This can also be explained in terms of hole back injection from base to emitter. The abrupt hole barrier ($\Delta E_V$) of the abrupt emitter suppresses the hole back injection into the emitter SCR more effectively, even though its height is only about 38% of that of the graded emitter case ($\sim \Delta E_g$) [47].

Table 7.3- Layer structure of the three HBTs used in the study of interface recombination effect on DC current gain.

<table>
<thead>
<tr>
<th>Graded AlGaAs/GaAs</th>
<th>Abrupt AlGaAs/GaAs</th>
<th>Abrupt InGaP/GaAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000Å, $n^+$-GaAs, $4 \times 10^{18}$ cm$^{-3}$</td>
<td>1200Å, $n^+$-Al$<em>{0.3}$Ga$</em>{0.7}$As, $4 \times 10^{17}$ cm$^{-3}$</td>
<td>1200Å, $n^+$-In$<em>{0.49}$Ga$</em>{0.51}$P, $4 \times 10^{17}$ cm$^{-3}$</td>
</tr>
<tr>
<td>300Å, $n^+$-Al$<em>x$Ga$</em>{1-x}$As ($x = 0.3 \rightarrow 0.0$), $2 \times 10^{18}$ cm$^{-3}$</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>900Å, $n^+$-Al$<em>{0.7}$Ga$</em>{0.3}$As, $4 \times 10^{17}$ cm$^{-3}$</td>
<td>1200Å, $n^+$-Al$<em>{0.3}$Ga$</em>{0.7}$As, $4 \times 10^{17}$ cm$^{-3}$</td>
<td>—</td>
</tr>
<tr>
<td>300Å, $n^+$-Al$<em>x$Ga$</em>{1-x}$As ($x = 0.0 \rightarrow 0.3$), $4 \times 10^{17}$ cm$^{-3}$</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>800Å, $p^+$-GaAs, $1 \times 10^{19}$ cm$^{-3}$</td>
<td>4000Å, $n^+$-GaAs, $3 \times 10^{16}$ cm$^{-3}$</td>
<td></td>
</tr>
<tr>
<td>4000Å, $n^+$-GaAs, $3 \times 10^{16}$ cm$^{-3}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000Å, $n^+$-GaAs, $4 \times 10^{18}$ cm$^{-3}$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Finally, the larger $\Delta E_V$ of the InGaP/GaAs interface as compared to its AlGaAs/GaAs counterpart is mainly responsible for the much weaker dependence of $\beta$ on interfacial recombination in former type of HBTs, especially at high current levels. Additionally, the
bandgap of In\textsubscript{0.49}Ga\textsubscript{0.51}P is larger than that of Al\textsubscript{0.3}Ga\textsubscript{0.7}As, which results in smaller \( n_i \) in InGaP emitters. It should also be pointed out that although the interface recombination lifetimes of both InGaP and AlGaAs emitters were assumed similar, AlGaAs is much more prone to oxygen related traps and defects (such as DX centre) than InGaP. Hence, the interface recombination velocity of InGaP/GaAs heterojunction is much smaller than that of AlGaAs/GaAs heterointerface. This together with the larger \( \Delta E_V \) and other advantages of InGaP/GaAs (as outlined in section 3.1) makes InGaP HBTs much more favourable structures than both graded and abrupt AlGaAs/GaAs HBTs.

7.4. Temperature Dependence of Current Gain: HEBT versus HBT

When HBTs operate at high current densities, a phenomenon commonly observed is negative differential resistance (NDR) or a decreasing current gain (\( \beta \)) with increasing \( V_{CE} \). The mechanisms responsible are generally attributed to a variation in current gain with junction temperature and are the same as those giving rise to the variation of gain with ambient temperature. In general, non-uniform power dissipation in a multi-finger bipolar transistor will produce a thermal gradient across the fingers of the device. However, if the current gain was independent of temperature, there would not be a feedback mechanism to produce a catastrophic effect. In reality, however, when the temperature coefficient of the gain is positive, thermal runaway will take place as in power silicon BJTs. On the other hand, a negative temperature coefficient gives rise to gain depression, thus the \textit{collapse} of the collector current in the I-V plot of the HBT [378-379]. Therefore, achieving an HBT with a temperature insensitive \( \beta \) would be an ideal target both for stability and linearity of power amplifiers.

Gao \textit{et al.} [380] previously analysed the temperature dependence of \( \beta \) and concluded that reverse hole injection is responsible for NDR in HBTs. Others [381-382] have produced models which attribute reducing electron mobility in the base with increasing temperature as responsible for NDR. Experimental results in [30] suggested that reverse hole injection was important at high temperature in AlGaAs/GaAs HBTs but not in InGaP/GaAs devices, which is due to the larger value of \( \Delta E_V \) in the latter case. Later experimental results showed that current gain of some InGaP/GaAs HBTs may significantly vary with temperature [31,155], which has been attributed to the difference in the valence band discontinuities of ordered and disordered InGaP/GaAs heterojunctions [31,44]. Kren \textit{et al.} [383] have shown that SCR and surface recombination may be important for the reducing gain with increasing temperature in AlGaAs/GaAs devices. Recently, Yang \textit{et al.} [155-156] demonstrated that heterostructure-emitter bipolar transistors (HEBTs) have current gains much less sensitive to temperature than HBTs. In fact, \( \beta(\text{HEBT}) \) was observed to increase with temperature at low collector current densities, similar to that of Si BJTs. Yang \textit{et al.} speculated that the increase of gain in HEBTs is due to the dominance of the hole back injection from base to emitter by thermionic-emission, which is believed to have a different temperature dependence than the conventional...
diffusion current. A detailed analytical study of temperature dependence of $\beta$ for various types of HBTs was carried out by Ng et al. [24] in which the authors discussed each of the possible base current components and assessed its relative importance to temperature dependence. In this section, the main conclusions of the latter article are summarised and some of its results are demonstrated using numerical simulation. Additionally, temperature dependence of $\beta$(HEBT) is compared against $\beta$(HBT) and it is shown that the observed behaviour of $\beta$(HEBT) by Yang et al. results from the recombination of electron-hole pairs inside the narrow bandgap emitter (spacer) region of these devices.

### 7.4.1. Analysis of Temperature Dependence of $\beta$

Following [24], the collector current of an HBT in forward active mode can be written as:

$$I_C(T) = CC(T) \cdot \exp \left( \frac{qV_{BE}}{\eta_C kT} - \frac{E_1}{kT} \right)$$  \hspace{1cm} (7.6)

where for an abrupt emitter-base with thermionic-emission and tunnelling $E_1 = E_{gB} + \Delta E_{fn}$ and

$$CC(T) = \frac{(qA_{BE}D_{nB}N_{CB}N_{VB})}{(W_B N_{AB})}$$

where $\eta_C$ is the collector ideality factor, $E_1$ is the effective barrier to collector current flow, $\Delta E_{fn}$ is the electron quasi-Fermi level splitting, and other symbols have their usual meanings. Note that since $\Delta E_{fn}$ is voltage dependent, it will contribute to the measured collector current ideality factor (see section 7.2 for more details). For a graded HBT and assuming that neutral base diffusion dominates the transport current: $E_1 = E_{gB}$, $\eta_C = 1$, and

$$CC(T) = \frac{(qA_{BE}D_{nB}N_{CB}N_{VB})}{(W_B N_{AB})}$$

For both abrupt and graded HBTs, the collector current can be expressed as a product of a pre-exponential part which changes slowly with $T$ and an exponential part which can change much more rapidly with $T$.

The main components which make up the base current are space-charge region recombination ($I_{B,SCR}$), interfacial recombination at the emitter-base interface ($I_{B,int}$), extrinsic base surface recombination ($I_{B,BS}$), reverse hole injection from base to emitter ($I_{B,p}$), and base bulk recombination ($I_{B,bulk}$). These components can also be expressed as a product of two factors:

$$I_B(T) = BB(T) \cdot \exp \left( \frac{qV_{BE}}{\eta_B kT} - \frac{E_2}{kT} \right)$$ \hspace{1cm} (7.7)

where $\eta_B$ is the base ideality factor and $E_2$ is the effective barrier to base current flow. Defining $\beta_i(T) = I_C(T)/I_{B,i}(T)$, where subscript "i" refers to individual components of the base current, we have:

$$\beta_i(T) = \frac{CC_0}{BB_0} \left( \frac{T}{T_0} \right)^{x-y} \exp \left( \frac{qV_{BE}}{\eta kT} - \frac{E}{kT} \right)$$ \hspace{1cm} (7.8)

where $1/\eta = 1/\eta_C - 1/\eta_B$, $E = E_1 - E_2$, and the temperature dependence of the pre-exponential factors can be expressed as:
Chapter 7 Numerical Simulation Results

\[ CC(T) = C C_0 (T / T_0)^x \quad \text{and} \quad BB(T) = B B_0 (T / T_0)^y \]  

(7.9)

where the subscript "0" denotes the value of the parameter at 300K. Differentiating \( \beta_i(T) \) with respect to \( T \), one obtains:

\[
\frac{\partial \beta_i}{\partial T} = \beta_i \left[ \frac{x - y}{T} + \frac{q}{\eta KT} \left( \frac{\partial V_{BE}}{\partial T} - \frac{V_{BE}}{T} \right) - \frac{1}{kT} \left( \frac{\partial E}{\partial T} - \frac{E}{T} \right) \right]
\]  

(7.10)

For a constant base current as in the common-emitter output characteristics, one has the additional condition \( \partial I_B / \partial T = 0 \) and it follows [24]:

\[
\left( \frac{\partial \beta_i}{\partial T} \right)_{I_B = \text{cte}} = \beta_i \left[ \frac{x - \eta B}{\eta C} \frac{y}{T} + \frac{E_1 - \eta B}{\eta C} \frac{E_2}{T} \right] + \frac{1}{kT} \left( \frac{\eta B \partial E_2}{\eta C \partial T} - \frac{\partial E_1}{\partial T} \right)
\]  

(7.11)

Similarly, one can show that for a constant \( I_C \):

\[
\left( \frac{\partial \beta_i}{\partial T} \right)_{I_C = \text{cte}} = \beta_i \left[ \frac{n_C}{n_B} \frac{x - y}{T} + \frac{n_C}{n_B} \frac{E_1 - E_2}{kT^2} + \frac{1}{kT} \left( \frac{\partial E_2}{\partial T} - \frac{n_C \partial E_1}{n_B \partial T} \right) \right] = \frac{n_C}{n_B} \left( \frac{\partial \beta_i}{\partial T} \right)_{I_B = \text{cte}}
\]  

(7.12)

Since \( n_C \leq n_B \), temperature sensitivity of \( \beta \) under constant \( I_B \) condition is always greater than that under constant \( I_C \) condition. (7.11) and (7.12) are key equations in the present analysis. The first term inside the square brackets of these equations represents the difference in temperature sensitivity of the pre-exponential terms in the collector and base currents, the second term relates to the difference in temperature sensitivity due to the activation barriers to the collector (\( E_1 \)) and base (\( E_2 \)) current flow, and the third term the difference in the variation with temperature of \( E_1 \) and \( E_2 \) (usually associated with bandgap changes with temperature). In SHBTs, \( E_2 \) for SCR current in the emitter and reverse hole injection is related to the emitter bandgap whilst the base side components of the base current have \( E_1 = E_2 \). Thus \( E_2 > E_1 \) in the cases of SCR and reverse hole current and (7.11) and (7.12) can be negative when the second term dominates. For homojunction transistors, when significant bandgap shrinkage in the heavily doped emitter is taken into account, \( E_2 < E_1 \) and the commonly observed positive gain dependence of Si BJTs is predicted. In the following, Eq. (7.12) will be applied to each of the base current components individually, and various forms of temperature dependence observed in various HBTs will be explained.

7.4.2. Discussion of Various Base Current Components

7.4.2.a. SCR Recombination

For SCR recombination on the emitter side \( E_2 = E_{BE} / 2, \eta_B = 2 \) and \( BB(T) = BB_{SCR}(T) = \frac{[qA_{BE} W_{eff,E} (N_{CE,N_{VE}})^{0.5}]}{2\tau_{SCR,E}} \). Here \( W_{eff,E} \) and \( \tau_{SCR,E} \) are the effective recombination
region width and SCR recombination lifetime, respectively. An analytical expression for SCR recombination on the base side is more complex [33] and therefore difficult to define in the context of (7.12). At low current densities, SCR recombination on the emitter side is usually much larger than on the base side. At high currents, other base side recombination currents become more important. Therefore, SCR recombination on the base side can be neglected without introducing much error in the analysis [33]. For abrupt HBTs, \( E_1 = E_g B \) and \( \eta_B E_2 / \eta_C = E_g E \). Thus, the second term in (7.11) and (7.12) is negative and dominates the temperature dependence of \( \beta_{SCR} \). Therefore, the gain of HBTs at low currents, where SCR recombination is dominant, is reducing with temperature. Variation of \( \beta \) with \( I_C \) at various temperatures is simulated for abrupt InGaP/GaAs, AlGaAs/GaAs and graded AlGaAs/GaAs HBTs specified in Table 7.3, as well as abrupt InP/InGaAs HBT as specified in Table 7.2, and the results are shown in Fig. 7.6. The parameters used for the calculation of the first three structures are exactly the ones used for the simulations in section 7.3, apart from \( \tau_{int} \) which is set to 50ps here. As to the parameters for the calculation of gain in InP/InGaAs HBT, emitter, collector, and base series resistances of 6, 5, and 40Ω, respectively, are assumed together with \( A_{BE} = 2.0 \times 10^{-6} \text{ cm}^2 \) and \( \tau_{int} = 50 \text{ps} \). It can be seen that at low to medium currents, where SCR recombination dominates, \( \beta \) reduces significantly with increasing \( T \) for both InGaP and AlGaAs abrupt HBTs. The crossover behaviour of gain for graded AlGaAs/GaAs HBT is due to the recombination inside the narrow bandgap graded region and will be thoroughly explained in subsection 7.4.3. It must be mentioned that in the present work B-E interface recombination is treated as part of the emitter SCR recombination current component, since a region \( \sim 100 \AA \) thick is assumed to be affected by the presence of interface dislocations.

**7.4.2.b. Extrinsic Base Surface Recombination**

The extrinsic base surface recombination current can be a dominant component of base current even up to the highest injection levels. If this recombination process is limited by the supply of electrons from the emitter into the base because of the high surface recombination velocity of p-type base materials (such as p-GaAs), then this component can be written as \( I_{BS} = \xi I_C \) where \( \xi \) is a temperature independent parameter proportional to the emitter perimeter to area ratio [24]. In the present one-dimensional simulation we are forced to neglect this component of base current, but it is not expected to produce much error in the temperature dependence of \( \beta \), especially for large geometry HBTs.

**7.4.2.c. Reverse Hole Injection**

The reverse hole current for a short emitter has \( E_2 = E_g E \), \( \eta_B = 1 \), \( BB(T) = BB_p(T) = qA_{BE} D_{pE} N_{CE} N_{VE} / N_{DE} W_E \) where \( W_E \) is the emitter width. Note that the hole quasi-Fermi level \( E_{fp} \) is assumed to be continuous across the SCR since there is no band spike in the valence band. The second term in (7.11) will dominate the temperature dependence of this mechanism for most systems where \( E_1 - E_2 = -\Delta E_V \) for an abrupt junction and \( E_1 - E_2 = -\Delta E_g \) for a graded junction [20,30]. In abrupt HBTs, the \( \Delta E_C / \Delta E_V \) partitioning
ratio for different material systems, even for the same emitter and base bandgaps, will determine the relative importance of reverse hole current compared to injected electron current. For abrupt AlGaAs/GaAs or AlInAs/InGaAs HBTs, the conduction band discontinuity suppresses the electron current, causing an increase in the relative importance of the reverse hole current. On the other hand, the very small $\Delta E_C/\Delta E_V$ ratio in InGaP/GaAs HBTs significantly suppresses the reverse hole injection current. Reverse hole injection usually has a minor role in limiting $\beta$ at 300K, but its rate of change with temperature is much faster than the dominant base bulk recombination (see next subsection). Although the exact conditions when reverse hole injection is important relative to other mechanisms are difficult to define, numerical simulations show that reverse hole injection could be significant as the temperature is raised in abrupt AlGaAs/GaAs devices but not in InGaP/GaAs at high current densities, as also evidenced by the results of [30].
collector junction leakage effects tend to dominate, there is little evidence in the literature of NDR at high currents. Indeed the experimental results on InP/InGaAs HBTs in [348-349] at low power show an increasing gain with ambient temperature near 300K, contrary to the AlGaAs/GaAs devices. The numerical simulation shown in Fig. 7.6(d) also suggests that $\beta$ increases with temperature for $100K < T < 200K$, almost saturates for $200K < T < 300K$, and then slowly reduces for $T > 300K$, which is very similar to the trends experimentally observed in [348-349]. This behaviour is fully explained in the next subsection when temperature dependence of radiative and Auger base bulk recombination is considered. Where AlInAs, which has a large $\Delta E_C (0.52eV)$ and a relatively small $\Delta E_V (0.18eV)$, is used as the emitter with an InGaAs base, slight reduction of $\beta$ with temperature is experimentally observed [347]. However, as will be shown in the next subsection, not all the temperature dependence of $\beta$ at high current densities can be attributed to the back injection of holes.

A graded B-E junction will increase the electron flow relative to the hole injection by reducing the conduction band discontinuity. In this case, the bandgap difference between the emitter and base alone is important in determining the injection efficiency. According to (7.11), a graded junction will increase the temperature dependence by increasing the magnitude of the second term. However, the relative importance of the reverse hole current is much less compared with an abrupt junction because its magnitude is reduced by a further factor of $\exp(-\Delta E_C / kT)$ for the same electron current [20]. In this case, other base current components will contribute to the temperature dependence as also shown in Fig. 7.6.

### 7.4.2.d. Base Bulk Recombination

For base bulk recombination the analysis is much simpler and more general than that leading up to (7.11)-(7.12) and can easily be treated from $I_B = I_{bulk} = I_C (1 - \alpha_T) / \alpha_T$ where $\alpha_T$ is the base transport factor. Differentiating $I_B$ with respect to temperature and setting $\partial I_B / \partial T = 0$ as before and expressing in terms of current gain one gets:

$$\frac{\partial \beta_{bulk}}{\partial T} = \frac{1}{(1 - \alpha_T)^2} \frac{\partial \alpha_T}{\partial T}$$  \hspace{1cm} (7.13)

The condition for NDR is then simply determined by the temperature dependence of the base transport factor, $\alpha_T = 1 - W_B^2 / 2D_{nB} \tau_{nB}$ [316: p.141], which depends on the electron minority lifetime and diffusion constant. There is a change in the base bulk recombination mechanism with base doping and temperature. As discussed in chapter 5 and shown in Fig. 5.35, when the base doping is lower than $\sim 10^{19} \text{ cm}^{-3}$ at room temperature, base bulk recombination is dominated by the radiative recombination mechanism. On the other hand, when the base doping exceeds this, base bulk recombination begins to be dominated by the Auger recombination mechanism. The radiative recombination lifetime was shown to have a positive temperature dependence $\tau_{rad} \propto T^{1.5}$ (see subsection 5.5.1). Also $D_{nB} \propto \mu_{nB} T \propto T^m$ with $0 \leq m \leq 1$, since the mobility of majority electrons (which is assumed for the simulations
in the present work) inside the highly doped base region is almost constant with temperature and that of minority electrons is approximately proportional to $1/T$ (see subsection 5.4.2 and [384]). Therefore, one expects $\partial \alpha_T / \partial T > 0$ for lower range of base doping or temperatures below 300K. Conversely, Auger recombination has a lifetime which shows a negative temperature dependence, leading to $\partial \alpha_T / \partial T < 0$ and NDR in the output characteristics. Consequently, if the base bulk recombination is the dominant component of base current, one expects to observe a peak in the $\beta$ versus T characteristics at a temperature which is dependent on the base doping and bandgap.

Figure 7.7(a) shows the calculated gain at $J_C = 1 \times 10^4 \text{ A/cm}^2$ at various temperatures normalised to the gain at 300K for the HBTs analysed in Fig. 7.6. It can be seen that the gain of InGaP/GaAs and InP/InGaAs HBTs are relatively less sensitive to T than their AlGaAs counterparts. This is especially important for high temperature and/or high power applications. Variation of gain in graded AlGaAs device is due to the SCR recombination inside the graded region. But, the gains of abrupt devices at high current densities show a peak at temperatures in the range 200-400K. This behaviour of gain versus temperature is due to the competition between the positive temperature coefficient of radiative recombination inside the base bulk region and the negative coefficient due to the two components of Auger recombination inside the base bulk and back injection of holes (see also [385]). For larger Auger coefficients or smaller $\Delta E_y$'s, the peak in $\beta$ occurs at lower temperatures. The gain of InP/InGaAs HBT shows a reduction at temperatures above 250K and one may expect to observe NDR in the output characteristics of such devices. However, the large B-C leakage current (due to generation of carriers inside the narrow bandgap InGaAs base and collector regions) usually dominates, and a positive dynamic resistance is experimentally observed in the $I_C-V_{CE}$ characteristics of InP/InGaAs HBTs.
In order to compare the relative importance of Auger recombination inside the base and back injection of holes, the calculated gains of abrupt AlGaAs, InGaP, and InP HBTs with and without a temperature dependent Auger coefficient are shown in Fig. 7.7(b). For the case of a constant Auger coefficient, the 300K value of this parameter is used in the entire range of temperature. It can be seen that the negative temperature coefficient of abrupt AlGaAs/GaAs HBT is entirely due to the back injection of holes [30], since the simulated gains with and without a temperature dependent $C_{p,Aug}$ are almost identical. On the other hand, it can be seen that the negative temperature coefficient of gain in InP/InGaAs HBT is mainly due to the temperature dependence of Auger coefficient. A slight reduction of gain in the latter device due to the back injection of holes can only be observed at $T$ above 550K. In the case of InGaP/GaAs HBTs, a combination of both base current components seems to be responsible for the negative slope of gain at higher temperatures, with the back injection of holes mainly dominating the negative slope for $T > 450K$.

A final point has to be remarked as to the variation of gain with $T$ experimentally observed in InGaP/GaAs HBTs, such as those reported in [30-31,155]. Contrasting difference can be observed in the $\beta$ vs. $T$ characteristics for InGaP/GaAs HBTs in the above references. The result of [155] and that of [31] for semi-ordered InGaP show significant reduction of gain at high temperatures similar to those observed in Fig. 7.6(a). On the other hand, [30] shows an almost constant $\beta$ for InGaP/GaAs HBTs. The experimentally observed constant gain may be due to one of the following:

- As stated in [31], HBTs having ordered InGaP as emitter material present an almost constant $\beta(T)$, while those composed of disordered InGaP show a significant reduction of gain with temperature, similar to that of abrupt AlGaAs/GaAs HBTs. This has been attributed to the larger $\Delta E_V$ for ordered InGaP/GaAs heterojunctions compared to that of disordered InGaP/GaAs (see subsections 5.3.1.f and 5.3.2.d for more details), which suppresses the back injection of holes in the former devices. Indeed, when a larger value of $\Delta E_V$ was used in the simulation, a less sensitive gain at higher temperatures was observed.
- Existence of a large density of traps inside the base may affect the $\beta(T)$ characteristics of HBTs. Enhanced SRH recombination due to the traps produces a temperature insensitive lifetime for base bulk recombination, and consequently, results in a flat gain as a function of temperature. This might well be the reason for a reasonably stable gain in DHBTs, where an interruption may be necessary during the growth of GaAs base on top of InGaP collector (see subsection 9.2.6 for more details).
- If significant extrinsic base surface recombination dominates the base current of HBTs, a current gain less sensitive to temperature may be obtained, since this component of base current is proportional to $I_C$ with a proportionality factor independent of $T$ [24].

7.4.3. HEBT versus HBT

The heterostructure-emitter bipolar transistor (HEBT) with the separation of electron injection
and hole confinement is proposed to take advantage of both the heterojunction and homojunction. As discussed in chapter 3, HEBT can effectively eliminate the emitter potential spike and offers a low offset voltage. However, if the narrow bandgap emitter region is too thick, recombination inside this region significantly deteriorates the gain [31]. Therefore, in most of the recent works [155-156] thickness of this region is reduced to a level that makes this design no more than a conventional HBT with a thin “spacer” layer between the base and emitter. As far as temperature dependence of current gain is concerned, it has recently been shown that using a HEBT design results in a current gain which is much less sensitive to the variation of temperature than an HBT design [155-156]. In [156] it was shown that the current gain of an AlGaAs/GaAs HEBT reduces only by 10% when the substrate temperature is increased from 25 to 175°C, while β of a similar AlGaAs/GaAs HBT was decreased by more than 60% in the same temperature range. Similar comparison was made in [155] between InGaP/GaAs HEBTs and HBTs. It was shown that while β(InGaP HBT) has a large negative temperature coefficient, β(InGaP HEBT) first shows some increase and then reduces with increasing temperature with the maximum variation of ±8% at high current densities. In the above two works the authors attributed the better thermal stability of HEBTs to the dominance of hole back injection by thermionic-emission, which, to their belief, has a different temperature dependence than the conventional diffusion current. However, it will be shown in the present work that simulation of current gain including the effects of thermionic-emission and tunnelling at all heterojunctions of the device does not reproduce their experimental data. Instead, when some extra recombination centres were introduced inside the spacer layer, simulated current gains closely resemble the data in the above works.

The layer structure of the HEBT used for this analysis consists of a 5000Å GaAs sub-collector layer (n = \(3 \times 10^{18} \text{ cm}^{-3}\)), a 5000Å GaAs collector layer (n = \(3 \times 10^{16} \text{ cm}^{-3}\)), a 1000Å GaAs base (n = \(4 \times 10^{19} \text{ cm}^{-3}\)), a 150Å GaAs emitter (spacer) layer (n = \(1 \times 10^{17} \text{ cm}^{-3}\)), a 500Å InGaP confinement (emitter) layer (n = \(3 \times 10^{17} \text{ cm}^{-3}\)), and a 1000Å GaAs emitter cap layer (n = \(4 \times 10^{18} \text{ cm}^{-3}\)). The layer structure for the HBT is similar to that of the HEBT only without the spacer layer. An active B-E area of \(4 \times 10^{-5} \text{ cm}^2\) (similar to that in [155]), and emitter, collector, and base series resistances of 2, 5, and 20Ω, respectively, were used for the simulation. The results of the HEBT simulation are shown in Fig. 7.8(a). Note that all the recombination currents are scaled with a factor of 0.2 to bring agreement between the simulated and measured maximum current gains. As can be observed, the gain of the HEBT reduces with increasing T in the entire range of \(I_C\). Therefore, the thermionic-emission theory can not explain the β vs. T characteristics of HEBTs. On the other hand, when a region \(~120Å\) thick right above the base is defined as interfacial layer with an effective recombination lifetime of \(\tau_{int} = 250\text{ps}\), the simulated current gain plots (shown in Fig. 7.8(b)) closely resemble the experimental data in [155].
According to the formulation of temperature dependence of current gain as in Eq. (7.11), the above behaviour of $\beta(I_C,T)$ can be explained as follows. When the emitter SCR recombination occurs inside the wide bandgap materials such as InGaP or AlGaAs, it results in an activation energy $E_2 = E_{gB}/\eta_C$, which is larger than $E_1 = E_{gB}$. However, when SCR recombination occurs inside the narrow bandgap GaAs spacer, $E_2 = E_{g,\text{spac}}/2$ and $E_2/\eta_C$ becomes smaller than $E_1$. (Bear in mind that the actual value of $\eta_B$ is somewhat smaller than 2.) Therefore, the second term in (7.11) is positive, which leads to a positive temperature coefficient for $\beta$ at low to medium current densities where SCR recombination inside the spacer region is dominant.

Figure 7.9 shows the simulated current gains of the above InGaP/GaAs HEBT and HBT as a function of temperature at $I_C$ levels of 0.01, 1, and 10 mA. These particular current levels are chosen to be the same as those used in Fig. 2 of [155]. As can be seen, the gain of HBT continuously reduces with temperature at all current levels. But the gain of the HEBT first increases with temperature due to the dominant SCR recombination inside the spacer region. At high collector currents or high temperatures, the negative temperature coefficient of Auger recombination inside the base and/or back injection of holes overturns the positive
coefficient of SCR recombination, and the gain sarts to decrease with temperature. This behaviour is exactly similar to that experimentally observed by Yang et al. [155].

7.4.4. Conclusions

In this section, the contribution of various components of base current to the temperature dependence of current gain in HBTs was discussed in details. From this analysis, the difference in the base and collector current activation energies (or barriers to the current flow) and the temperature dependence of the pre-exponential terms were shown to be important in determining the temperature dependence of gain. Since NDR occurs at high current densities, the base current components relevant to this mechanism are reverse hole injection and base bulk recombination. There are generally several base current mechanisms with different weighting operating in a device at one time and this accounts for the wide variation in reported temperature dependent studies. From this study, however, the following observations can be made:

- The gain of HBTs in the low current region is governed by SCR recombination in the emitter and is subject to large reduction with increasing temperature despite the fact that the recombination lifetime is generally thought to increase with temperature.
- Extrinsic base surface recombination in GaAs is unlikely to cause significant temperature dependence of gain, even in unpassivated structures where it can dominate the base current, because this process is electron supply limited, having the same temperature dependence as the collector current.
- At high current densities, reverse hole injection can contribute to limiting the gain and is the main cause of NDR in abrupt AlGaAs/GaAs HBTs, while contributing to the NDR of InGaP/GaAs HBTs to some smaller extent. However, the gain of InP/InGaAs HBTs is mainly limited by the Auger recombination inside the base bulk region, which has a negative temperature coefficient but usually does not lead to NDR because of the significant B-C leakage current at high $V_{CE}$ in these devices.
- At high base doping levels employed in advanced HBTs ($> 1 \times 10^{19} \text{ cm}^{-3}$), in devices whose gain is limited by base bulk recombination, the negative temperature dependence of the dominant Auger recombination lifetime can result in NDR.
- In Si BJTs, bandgap narrowing inside the emitter results in a smaller activation energy for the holes back injecting to the emitter than electrons injecting to the base. Consequently, a positive temperature coefficient of current gain is observed in such devices, which may lead to thermal runaway in power transistors.
- When a narrow bandgap spacer layer is inserted between the emitter and base regions (as in HEBTs), SCR recombination inside this region may result in a positive temperature coefficient for low to medium current densities where this component of base current is dominant.
7.5. Graded Base HBTs

Recent improvements in growth techniques have made bandgap engineered base layers, which enhance HBT performance, a manufacturable technique. Early results showed that bipolar transistors utilising a strained Si$_x$Ge$_{1-x}$ base [138] and AlGaAs/GaAs HBTs utilising an Al$_x$Ga$_{1-x}$As graded base [111,132] both display reduced base transit times and increased current gains. An alternative method for bandgap engineering the base of GaAs HBTs is to use strained In$_{1-x}$Ga$_x$As to compositionally grade the base from narrow bandgap In$_{1-x}$Ga$_x$As at the collector to wider bandgap lattice-matched GaAs at the emitter [14,93,133-134].

In addition to producing a field which enhances electron transport through the base, the In$_{1-x}$Ga$_x$As grade shows promise for several reasons. First, unlike an AlGaAs grade, the In$_{1-x}$Ga$_x$As grade does not alter the B-E band alignment, maintaining the same large valence band discontinuity as in an HBT with a non-graded base, and the same conduction band discontinuity which might be necessary for a near-ballistic electron transport across the base. Second, compared to AlGaAs graded base device, the majority carrier mobility, and hence the extrinsic base resistance, is improved in two ways: the hole mobility in InGaAs is higher than AlGaAs (see chapter 5), and the strained layer causes valence band splitting, reducing the hole effective mass and further increasing the mobility [192,386]. Third, the minority electron mobility inside the base is significantly larger in InGaAs than in AlGaAs. Fourth, the indium co-doping in the carbon-doped base may improve device reliability over a carbon-doped base without indium [70]. Finally, presence of Al inside the base of AlGaAs-graded HBTs results in a significant reduction of electron lifetime, thus a degraded gain [111], while lifetime of the base layers is not significantly degraded by adding a small percentage of indium. However, there are some problems associated with the use of InGaAs inside the base of GaAs HBTs. The existence of a large In mole fraction at the collector end of the base results in a conduction band discontinuity at the B-C heterojunction, which may degrade the base transit time, current gain, and cutoff frequency of the device (similar to the case of abrupt B-C heterojunction in DHBTs; see chapter 11 for more details). Also the mechanical strain developed because of the lattice mismatch between In$_{1-x}$Ga$_x$As and GaAs substrate may seriously deteriorate the material quality of the base and emitter layers. Additionally, when carbon is used as the dopant species inside the base layer, the incorporation of carbon dopants exponentially decreases with In mole fraction [133]. The first two of the above problems may be avoided by using a small indium content inside the base, and possibly using a spacer layer between the base and collector. The last issue also can be resolved by step adjusting the V/III ratio and CCl$_4$ dopant source flow rate throughout the growth of the base in order to maintain a constant carbon concentration [133]. Indeed by using the latter method, Hartmann et al. [134] demonstrated graded InGaAs base HBTs with small indium contents of 3-5% which showed nearly 75% improvement in current gain and over 20% improvement in cutoff frequency compared to a similar device with a standard non-graded base design.
In this section, first the DC and high frequency parameters of InGaP/GaAs HBTs with a graded In_{1-x}Ga_xAs base are studied as a function of In mole fraction at the collector end of the base using numerical simulation. Then, the performance of an AlGaAs/GaAs HBT with graded Al_xGa_{1-x}As inside the base is investigated and compared against the graded In_{1-x}Ga_xAs structure. Ideally, one should also take the effects of strain on material properties of In_{1-x}Ga_xAs grown lattice mismatched on GaAs into account, as in [387]. Therefore, quantitative analysis performed in this section is not accurate, and needs to be combined with some qualitative interpretation.

### 7.5.1. Graded In_{1-x}Ga_xAs Base

The layer structure of the graded base HBT used in this section is shown in the first column of Table 7.4. The indium content of the In_{1-x}Ga_xAs base is linearly graded from (1-Xc) at the collector end of the base to 0 (i.e., GaAs) at the emitter end. Values of Xc equal to 0.80, 0.85, 0.90, 0.93, 0.95, 0.97, and 1.00 (i.e., standard non-graded base), corresponding to indium contents (In_C) of 0.20, 0.15, 0.10, 0.07, 0.05, 0.03, and 0.00, respectively, are considered. A B-E area of 4×10 μm², similar to the ones experimentally tested in [134], and emitter, collector, and base series resistances of 10, 5, and 20Ω, respectively, are used for the numerical analysis. Variation of gain for this HBT versus In_C is shown with dashed lines in Fig. 7.10. It can be seen that first the gain slightly improves for In_C = 0.03, but then significantly degrades for stronger gradings. This is mainly due to the electron blocking effect of the conduction band barrier formed at the B-C heterojunction (see chapter 11 for more details), which is also confirmed by the V_CB dependence of the gain for this structure, again shown in Fig. 7.10. To remove this undesirable effect, a 200Å In_{1-Xc}Ga_{Xc}As (n = 1×10¹⁷ cm⁻³) spacer layer is inserted between the base and collector of all of the graded InGaAs HBT structures shown in Table 7.4. The spacer, as shown in Fig. 7.11, lowers the potential spike at the B-C heterojunction and allows the majority of electrons to jump over it and get collected. Once the spacer is included in the calculation, the gain of the graded base HBTs rises to almost twice as large as the gain of the standard HBT for indium compositions in the range 5-10%, before again starting to reduce for higher indium mole fractions.

<table>
<thead>
<tr>
<th>Graded InGaAs Base HBT</th>
<th>Graded AlGaAs Base HBT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000Å GaAs Cap  n = 5×10¹⁸ cm⁻³</td>
<td>1000Å GaAs Cap  n = 5×10¹⁸ cm⁻³</td>
</tr>
<tr>
<td>200Å In_{0.49}Ga_{0.51}P Cap  n = 2×10¹⁸ cm⁻³</td>
<td>300Å Al_{x}Ga_{1-x}As Cap  n = 2×10¹⁸ cm⁻³</td>
</tr>
<tr>
<td>(x = 0.3 → 0.0)</td>
<td></td>
</tr>
<tr>
<td>800Å In_{0.49}Ga_{0.51}P Emitter n = 5×10¹⁷ cm⁻³</td>
<td>1000Å Al_{0.3}Ga_{0.7}As Emitter n = 5×10¹⁷ cm⁻³</td>
</tr>
<tr>
<td>700Å In_{1-x}Ga_xAs Base p = 4×10¹⁹ cm⁻³</td>
<td>700Å Al_{x}Ga_{1-x}As Base p = 4×10¹⁹ cm⁻³</td>
</tr>
<tr>
<td>(x = Xc → 1)</td>
<td>(x = 0 → Xe)</td>
</tr>
<tr>
<td>5000Å GaAs Collector n = 1×10¹⁷ cm⁻³</td>
<td>5000Å GaAs Collector n = 1×10¹⁷ cm⁻³</td>
</tr>
<tr>
<td>5000Å GaAs Sub-Collector n = 5×10¹⁸ cm⁻³</td>
<td>5000Å GaAs Sub-Collector n = 5×10¹⁸ cm⁻³</td>
</tr>
</tbody>
</table>

Table 7.4- Layer structures of the two HBTs used for the study of graded base designs.
Chapter 7 Numerical Simulation Results

1.0

Figure 7.10 - Variation of the gain versus In mole fraction at the collector end of the graded base region for InGaP/(In)GaAs HBTs with (solid line) and without (dashed line) the 200Å spacer layer between the base and collector regions.

Figure 7.11 - Band structure of the graded base InGaP/(In)GaAs HBT (solid lines) with 200Å In$_0.2$Ga$_{0.8}$As spacer inserted between the base and collector regions for $V_{BE} = 1.0$V and $V_{CB} = 0.0$V. The dashed lines show the band structure of a standard non-graded base HBT.

Table 7.5 summarises the calculated values of electron drift field inside the base $E_n$, DC current gain, base sheet resistance $R_{shb}$, peak cutoff frequency $f_T$, base transit time $\tau_B$, and the total collector delay time (comprised of the collector depletion region delay time $\tau_C$ and the collector charging time $\tau_{CC}$) for various indium contents at the collector end of the base. The base sheet resistance is not affected significantly by adding indium to the base. But the base transit time is continuously reduced due to the presence of larger drift fields inside the base, until the drift velocity is nearly saturated for very large electric fields. It is important to mention that the delay associated with the conduction band spike at the B-C heterojunction is not included inside $\tau_B$. Instead, it is calculated as part of the collector delay time. As the extent of base grading increases, $\tau_B$ tends to reduce, but the conduction band spike at the B-C heterojunction gets larger and its associated time delay grows. The competition between the above two mechanisms leads to a peak in $f_T$ at indium compositions around 5-7%.

Considering that mechanical strain for highly graded bases may degrade the quality of the base and emitter regions, maximum indium mole fractions in the range 3-5% may be an optimum design for improving both the DC gain and high frequency performance of graded base InGaP/InGaAs/GaAs HBTs. This result is in excellent agreement to that experimentally confirmed by Hartmann et al. [134], and once again demonstrates the capability of the present numerical simulation code to predict the behaviour of novel HBTs.

As a final remark, the reader may notice that the calculated cutoff frequencies for the above HBTs are underestimating the experimentally observed data in [93,134], especially when considering that the simulated device is free from the parasitics associated with the extrinsic base and collector regions, bonding pads, and metal interconnects. A combination of the following reasons can explain the above uncertainty in calculated values of cutoff frequencies:
Table 7.5- Summary of the calculated parameters for the graded base InGaP/InGaAs/GaAs and 
AlGaAs/AlGaAs/GaAs HBTs studied in this section ($A_{BE} = 4 \times 10^{-11} \mu m^2$).

<table>
<thead>
<tr>
<th>In (Al) Content</th>
<th>$E_n$ (kV/cm)</th>
<th>$\beta$ ($I_C = 5mA$)</th>
<th>$R_{sh,B}$ (Ohm/cm)</th>
<th>Peak $f_T$ (GHz)</th>
<th>$\tau_B$ (ps)</th>
<th>$\tau_C + \tau_{CC}$ (ps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>InGaP/InGaAs/GaAs Graded Base HBT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.00</td>
<td>0.0</td>
<td>52.4</td>
<td>324</td>
<td>44.3</td>
<td>2.10</td>
<td>1.20</td>
</tr>
<tr>
<td>0.03</td>
<td>6.2</td>
<td>85.0</td>
<td>325</td>
<td>54.8</td>
<td>1.23</td>
<td>1.27</td>
</tr>
<tr>
<td>0.05</td>
<td>10.2</td>
<td>101.6</td>
<td>327</td>
<td>61.4</td>
<td>0.97</td>
<td>1.22</td>
</tr>
<tr>
<td>0.07</td>
<td>14.2</td>
<td>103.3</td>
<td>328</td>
<td>60.7</td>
<td>0.91</td>
<td>1.32</td>
</tr>
<tr>
<td>0.10</td>
<td>20.1</td>
<td>101.0</td>
<td>330</td>
<td>54.9</td>
<td>0.86</td>
<td>1.54</td>
</tr>
<tr>
<td>0.15</td>
<td>29.6</td>
<td>92.3</td>
<td>333</td>
<td>42.6</td>
<td>0.83</td>
<td>2.20</td>
</tr>
<tr>
<td>0.20</td>
<td>38.7</td>
<td>81.4</td>
<td>337</td>
<td>24.9</td>
<td>0.82</td>
<td>4.47</td>
</tr>
<tr>
<td>AlGaAs/AlGaAs/GaAs Graded Base HBT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.00</td>
<td>0.0</td>
<td>33.0</td>
<td>324</td>
<td>37.4</td>
<td>2.12</td>
<td>1.04</td>
</tr>
<tr>
<td>0.03</td>
<td>5.4</td>
<td>35.4</td>
<td>335</td>
<td>41.8</td>
<td>1.30</td>
<td>1.01</td>
</tr>
<tr>
<td>0.05</td>
<td>8.9</td>
<td>32.2</td>
<td>343</td>
<td>41.6</td>
<td>1.02</td>
<td>1.04</td>
</tr>
<tr>
<td>0.10</td>
<td>17.9</td>
<td>19.7</td>
<td>363</td>
<td>35.6</td>
<td>0.86</td>
<td>1.02</td>
</tr>
<tr>
<td>0.15</td>
<td>26.9</td>
<td>11.0</td>
<td>386</td>
<td>29.4</td>
<td>0.81</td>
<td>1.01</td>
</tr>
<tr>
<td>0.20</td>
<td>35.7</td>
<td>5.4</td>
<td>411</td>
<td>24.5</td>
<td>0.80</td>
<td>1.05</td>
</tr>
</tbody>
</table>

- In the present simulation program, the minority electron mobility inside the base is
  assumed to be the same as majority electron mobility. But, as discussed in chapter 5 and
  also explained in [191], the minority electron mobility inside the base of HBTs is shown to
  be an increasing function of hole concentration for $p > 1 \times 10^{19} \text{cm}^{-3}$ such that for $p =
  4 \times 10^{19} \text{cm}^{-3}$ the minority electron mobility is almost 2-3 times larger than the majority
  mobility used in the present simulation code. This leads to a larger electron diffusion
  constant, and hence, a reduced base transit time.

- The velocity-field characteristics of minority electrons are observed to be largely different
  from those of majority electrons. For example, Degani et al. [327] reported the steady-state
  v-E characteristics for photoexcited minority electrons in p-In$_{0.53}$Ga$_{0.47}$As. They measured
  a steadily increasing velocity up to the value of $2.6 \times 10^7 \text{cm/s}$ at 7.5 kV/cm electric field,
  with no evidence of transferred electron effects (see subsection 5.4.4 for more details). In
  principle, this leads to an enhancement of electron drift velocity in the base region of HBTs
  which reduces $\tau_B$ (see Eq. (10.4)).

- As discussed in chapter 5, Eq. (5.67) is a steady-state velocity-field relationship; it does not
  take the velocity overshoot of electrons into account. As a result, (5.67) largely
  underestimates the average velocity of electrons inside the base and collector regions of
  HBTs when they transport from emitter to collector, and results in a calculated forward
  transit time which is much larger than the measured data. The energy transport simulations
  performed by Horio et al. [54] has revealed that the peak $f_T$ of AlGaAs/GaAs HBTs can be
  almost 2.6 and 1.5 times higher than those calculated by pure drift-diffusion approach (as
  in the present work) for the collector SCR thicknesses of 0.1 and 0.5 $\mu m$, respectively.
7.5.2. Graded Al$_x$Ga$_{1-x}$As Base

In this subsection, the performance of abrupt B-E, graded base AlGaAs/GaAs HBTs as specified in the second column of Table 7.4 will be studied as a function of Al content at the emitter end of the base, Xe. Values of Xe equal to 0.0 (non-graded base), 0.03, 0.05, 0.10, 0.15, and 0.20 are used for numerical simulation. Figure 7.12 shows the calculated band structure of the graded base HBT with Xe = 0.20 (solid lines) as compared with a non-graded HBT (dashed lines). As can be seen, the B-C structure is not modified by the introduction of base grading, and therefore, no spacer layer is required for this structure. Additionally, the problems related to the mechanical strain are avoided due to the lattice-matching between Al$_x$Ga$_{1-x}$As and GaAs. But, the B-E heterojunction is modified, and in particular, the valence band discontinuity at this junction is reduced by the introduction of base grading. This leads to an enhanced back injection of holes which, as shown in the next few paragraphs, is the main performance limiting factor for abrupt B-E, graded base AlGaAs/GaAs HBTs.

![Figure 7.12- Band structure of an abrupt AlGaAs/GaAs HBT with a graded Al$_x$Ga$_{1-x}$As (x = 0.0 \rightarrow 0.2) base (solid lines) as compared with a standard non-graded base (dashed lines).](image)

Calculated parameters of the graded base AlGaAs/GaAs HBT with Xe are given in Table 7.5. As the base electric field increases, the current gain shows an initial slight increase at Xe = 0.03, but then drops significantly with further increase in Xe. This has two reasons. First, a reduced $\Delta E_V$ at the B-E heterointerface leads to an increased back injection of holes for larger values of Xe which limits the gain of the device. Second, introduction of more Al inside the highly doped base region significantly deteriorates the minority carrier lifetime, which then results in an increased base bulk recombination and a reduced gain. Base sheet resistance is continuously increasing with higher Al inside the base due to the reduction of hole mobility. This increase of base sheet resistance is much more pronounced than the case of graded InGaAs base, and eventually limits the maximum frequency of oscillation, $f_{\text{max}}$, of the device.
Surprisingly, the cutoff frequency of the graded base HBT improves by only 12% for Xe of 0.03-0.05, followed by a reduction at higher Al contents. The reduced electron lifetime should not affect the cutoff frequency of the device significantly, since the base lifetime does not appear inside the formula for cutoff frequency (Eq. (10.49)). Reduced electron mobility and saturation velocity inside the base for higher Al contents only has a minor effect on the base transit time. Indeed, as can be observed in Table 7.5, the base transit time decreases continuously with increasing Xe. The collector delay time is not affected by the base grading, since the collector structure is unchanged. The only remaining possibility is an increased emitter delay time with higher Al contents inside the base, which is discussed in the following paragraph.

The existence of a large valence band discontinuity at the B-E heterojunction of HBTs leads to emitter injection efficiencies approaching unity, even though emitter doping is designed to be lower than the base. Indeed, Eqs. (10.47)-(10.49) and even the small-signal equivalent circuit used in the majority of high frequency studies for HBTs (see Fig. 10.4, for instance) are derived based on the assumption of a unity emitter injection efficiency. However, for BJTs with a DC emitter injection efficiency, $\gamma_0$, smaller than unity, and for HBTs with a relatively small valence band discontinuity, the formula for $h_{21}$ has an extra term inside the denominator. In this case, Eq. (10.47) has to be modified to [388: pp.313-319]:

$$h_{21} = \frac{\alpha_0}{(1-\alpha_0) + \sqrt{(1-\gamma_0)\sqrt{1+j\omega\tau_{P_E}}}}$$

(7.14)

where $\tau_{P_E}$ is a characteristic lifetime of holes back injected into the emitter. This characteristic lifetime, determined either by the minority carrier lifetime in the emitter or by the time required for the holes to diffuse through the relatively thin emitter region, may be substantially larger than the forward transit time of HBTs. Therefore, if $\gamma_0$ is noticeably smaller than unity, the extra term in the denominator of (7.14) may degrade the cutoff frequency of HBTs. Figure 7.14 shows the variation of the calculated total delay time associated with the emitter of graded base AlGaAs/GaAs HBTs with different values of Xe as a function of $J_C$. The emitter delay time initially reduces with increasing $J_C$ due to the inverse collector current dependence of the B-E dynamic resistance, $R_{be}$. At high current levels, however, the emitter delay time saturates to a level mainly determined by the time associated with the removal (or modulation) of holes back injected into the emitter. For higher Al contents at the emitter end of the base, $\Delta E_V(B-E)$ reduces, resulting in more and more holes injected into the emitter, which require longer times for being modulated. As can be observed in Fig. 7.14, this delay time can be as large as a few picoseconds for high values of Xe in graded base AlGaAs/GaAs HBTs. For comparison, the emitter delay time of an InGaP/GaAs HBT is also shown in Fig. 7.14, which reduces almost proportional to $1/J_C$ for the entire range of current density with no clear sign of saturation up to $J_C = 1 \times 10^5$ A/cm$^2$. The estimated delay time associated with the back injected holes in InGaP/GaAs HBTs is on the order of
0.1 ps, which does not affect the cutoff frequency significantly. This is due to the large $\Delta E_g$ at the InGaP/GaAs heterojunctions.

Based on the above theory one would expect to observe a more pronounced improvement of current gain and cutoff frequency for graded emitter, graded base AlGaAs/GaAs HBTs, since the back injection of holes is suppressed in these structures by a large factor of $\exp(-\Delta E_g/kT)$, where $\Delta E_g = E_g(Al_{0.3}Ga_{0.7}As) - E_g(Al_{Xe}Ga_{1-Xe}As)$.

To confirm this, a 300Å $Al_{Xe}Ga_{1-Xe}As$ ($Xe = 0.3$) graded emitter layer ($n = \times 10^{17}$ cm$^{-3}$) was inserted between the emitter and base regions of the abrupt emitter, graded base AlGaAs HBT specified in Table 7.4. The band structures of this HBT for $Xe$ values of 0.2 (solid lines) and 0.0 (dashed lines) are shown in Fig. 7.15. Variation of the calculated $\beta_0$ and $f_T$ of this structure with $Xe$ (Fig. 7.16) indeed shows 69% and 46% improvement in $\beta_0$ and $f_T$, respectively, of the graded base AlGaAs/GaAs HBT with $Xe$ values in the range 0.05-0.10, as compared to the standard uniform base device. For $Xe > 0.10$, a reduction of both $\beta_0$ and $f_T$ can be observed, which is mainly attributed to the reduction of $\Delta E_g$.

In conclusion to this section, numerical simulation of DC and high frequency figures of merits
for InGaP/InGaAs/GaAs and AlGaAs/GaAs graded base HBTs revealed that the current gain and cutoff frequency of InGaP/InGaAs/GaAs and graded emitter AlGaAs/GaAs HBTs can be improved significantly by lightly grading the composition of the base material. However, grading the base of abrupt emitter AlGaAs/GaAs HBTs does not seem to be a good idea altogether, mainly because of the limitations imposed by the back injection of holes from base to emitter in such structures. Among the above structures, the InGaP/InGaAs/GaAs graded base HBT, which requires only 3-5% indium inside the base for optimised performance, seems to be the most promising one.

7.6. Capacitance Calculations

In this section the capability of the present simulation program to calculate the junction capacitances is demonstrated. The structure used as example for this section is an InGaP/GaAs SHBT composed of a 5000 Å GaAs sub-collector layer \( n = 5 \times 10^{18} \text{ cm}^{-3} \), a 7000 Å GaAs collector layer \( n = 2 \times 10^{16} \text{ cm}^{-3} \), a 700 Å GaAs base \( n = 4 \times 10^{19} \text{ cm}^{-3} \), a 2000 Å InGaP emitter layer \( n = 5 \times 10^{17} \text{ cm}^{-3} \), a 200 Å InGaP cap layer \( n = 2 \times 10^{18} \text{ cm}^{-3} \), and a 1000 Å GaAs emitter cap layer \( n = 5 \times 10^{18} \text{ cm}^{-3} \). B-E and B-C areas of 40 and 100 μm² are considered, respectively, and no series resistances are added to the device.

Accurate calculation of differential charges requires very small mesh spacing, and therefore, a much finer mesh than that required for calculation of I-V characteristics is necessary. This is especially important for the calculation of the B-E capacitance. While maximum mesh spacing \( h_{\text{max}} \) of 500 Å was proved sufficient for the calculation of I-V characteristics, values of \( h_{\text{max}} = 200 \text{ Å} \) and 50 Å was empirically found necessary for accurate determination of \( C_{\text{BC}} \) and \( C_{\text{BE}} \), respectively.

First, the reverse-biased B-C junction capacitance is simulated as a function of \( V_{\text{CB}} \). The results are shown in figure 7.17. As the reverse bias across the B-C junction increases, depletion region gets wider, and consequently, capacitance reduces. Solution of 1-D Poisson equation inside the depletion region of a one-sided p⁺-n junction yields [316: pp.79-81]:

\[
C_j = A \sqrt{\frac{qe_s N_D}{2(V_{\text{bi}} - V)}} \quad \Rightarrow \quad \frac{1}{C_j^2} = \frac{2(V_{\text{bi}} - V)}{A^2 qe_s N_D} \quad (7.15)
\]

where \( V_{\text{bi}} \) is the built-in potential of the junction, \( N_D \) is the n-side doping concentration, and \( V \) is the voltage applied to the junction, which is positive for forward and negative for reverse bias. A plot of \( 1/C_{\text{BC}}^2 \) versus \( V_{\text{CB}} \) is also shown in Fig. 7.17, where a very nice linear behaviour is observed. Extrapolation of this line for negative \( V_{\text{CB}} \)'s (i.e., forward biases) results in \( V_{\text{bic}} = 1.22 \text{V} \). For this calculation, B-E junction is left with a constant bias of 0V.

Figure 7.18 shows the calculated \( C_{\text{BC}} \) for \( V_{\text{CB}} = 0 \) as a function of collector current density, \( J_C \). It can be seen that \( C_{\text{BC}} \) shows a sharp rise as \( J_C \) approaches the Kirk effect (or base pushout) current density \( J_{\text{Kirk}} = qN_D CV_{\text{n, sat}} = 3.2 \times 10^4 \text{ A/cm}^2 \). The theory of base pushout is
explained in more details in the next section. But, in brief, as base pushout occurs, it initially compresses the B-C depletion region, which leads to an increase in the depletion capacitance. This figure demonstrates that the capacitance calculation is still possible under rather stringent conditions as in Kirk effect, although the convergence gets much slower under such conditions and sometimes special precautions are required.

![Figure 7.17](image1.png)  
**Figure 7.17** - Calculated B-C junction capacitance versus the reverse bias applied to this junction. A plot of $1/C_{BC}^2$ versus $V_{CB}$ is also shown, which gives a nicely linear behaviour in agreement with the depletion approximation theory.

![Figure 7.18](image2.png)  
**Figure 7.18** - Calculated $C_{BC}$ versus current density at a constant $V_{CB}$ of 0V. The sharp rise in the capacitance at high current densities is due to the so-called Kirk effect.

At this stage, it must be pointed out that the experimentally observed dip in the $C_{BC}$ versus $J_C$ characteristics (see Fig. 10.26) can not be seen in the simulated plot of Fig. 7.18. The physical origin of this dip is discussed in full details in chapter 10, where it is ascribed mainly to the reduction of electron velocity inside the collector SCR with increasing $V_{CB}$ for a constant depletion width. Increasing $V_{CB}$ increases both the electric field and temperature of the B-C junction, both reducing the electron velocity inside the collector SCR of a real device. Variation of junction temperature with dissipated power is not considered in the present simulation. Also, the finite negative slope of the velocity versus field for very high electric fields in excess of $10^5$ V/cm is not taken into account in the present calculations. Therefore, the dip in the $C_{BC}$ versus $J_C$ is absent in the simulated results.

As the forward bias across the junction approaches $V_{bi}$, one may expect from Eq. (7.15) to observe a diminished depletion region, and hence, an infinite capacitance. However, it is well-known that (7.15) is based on the so-called depletion approximation [389], which is not valid for highly forward-biased junctions. As $V$ approaches $V_{bi}$, the space-charge density is no longer determined by $qN_D^+$. Instead, the number of electrons injected to the junction will be increased to a level comparable to $N_D^+$, hence compensating some of the positive charges due to the donor ions. A smaller charge density simply leads to an increased depletion region width and a reduced capacitance. Indeed, the capacitance of the B-E heterojunction of our example HBT is shown in Fig. 7.19, where a peak is observed at a forward bias of 1.43V after
which capacitance is reduced due to the large density of injected electrons. For the calculation in Fig. 7.19 the series resistances are not included. When, in particular, the emitter series resistance is included in the simulation, the peak gets smoother and even may shift to higher forward biases. This is because the real junction voltage, which determines the junction capacitance, would be somewhat smaller than the applied external voltage, $V_{BE}$, in the presence of the voltage drop across $R_E$. Also shown in the inset of Fig. 7.19 is the so-called diffusion capacitance of the B-E junction. The differential charge variation of the B-E junction is partitioned into two components; charge variation inside the emitter is attributed to the depletion (or junction) capacitance, while charge modulation inside the base is referred to as diffusion capacitance. As discussed in chapter 10, diffusion capacitance is proportional to collector current density, and hence, varies exponentially with B-E bias, as also shown in Fig. 7.19. For reverse-biased or small forward-biased junctions, current is small and the effect of diffusion capacitance can be ignored. However, for large forward biases approaching $V_{bi}$, diffusion capacitance increases significantly and would be the dominant component of the B-E capacitance.

### 7.7. Kirk Effect

In this section, the so-called base pushout (or Kirk effect) [390] is graphically explained using the results of numerical simulation. The device used as example is the same as the one studied in previous section, but with a thinner collector region of 5000Å. The simulated electron and hole density profile, effective electric field, and band structure of the device are shown in Figs. 7.20(a)-(d), respectively. Referring to these figures, the Kirk effect can be explained as follows. The velocity of electrons inside the collector SCR is almost constant at $v_{n_{sat}}$, since electric field is in excess of $10^4$ V/cm in almost the entire SCR region. Therefore, the number of electrons is approximately flat inside the collector SCR, being equal to $n_C = J_n / qv_{n_{sat}}$.

As the B-E forward bias increases, electron concentration inside the collector SCR approaches the background collector doping level $N_{DC}$. This affects the space-charge density inside this region, and certainly should modify the electric field through the Poisson equation. Assuming an unchanging dielectric constant, the 1-D Poisson equation inside the collector SCR can be written as:
\[
\frac{dE}{dz} = \frac{q(N_D - n)}{\varepsilon_s}
\]  
(7.16)

where \( E \) is the electric field. Clearly, at low current densities when \( n \ll N_D \), the gradient of electric field would be positive, starting from a maximum at the B-C junction reducing to zero towards the highly doped sub-collector region. But, as the current density approaches \( J_{\text{Kirk}} = q \cdot N_D \cdot v_{n,\text{sat}} \), this gradient reduces and becomes equal to zero for \( V_{BE} = 1.48V \), when \( n = N_D = 2 \times 10^{16} \text{ cm}^{-3} \). For even higher forward biases, \( (N_D - n) \) becomes negative, and field gradient reversal occurs. As shown in Figs. 7.20(c) and (d), this relocates the high field region (i.e., where most of the potential drops) from B-C junction to near the collector/sub-collector junction. Therefore, the potential barrier for holes at the B-C junction reduces, allowing more holes to be injected into the collector region to compensate for the extra negative charges due to electrons (Fig. 7.20(b)). The formation of a zero-field region at the base end of the collector region can also be interpreted as creation of a \textit{current-induced base} region, whose width is given by [316: p.146]:

Figure 7.20- (a) Electron and (b) hole concentration profile, (c) effective electric field, and (d) band structure of the device under study as a function of \( V_{BE} \) before and after the occurrence of Kirk effect.
Chapter 7 Numerical Simulation Results

\[ W_{\text{cib}} = W_C \left[ 1 - \frac{J_{\text{Kirk}} - qV_{n,\text{sat}}N_{\text{DC}}}{J_C - qV_{n,\text{sat}}N_{\text{DC}}} \right] \quad \text{with} \quad J_{\text{Kirk}} = \frac{qV_{n,\text{sat}}}{qW_C^2} \left( N_{\text{DC}} + \frac{2e_s V_{CB}}{qW_C^2} \right) \quad (7.17) \]

where \( W_C \) is the \( n^- \)-collector thickness. As \( J_C \) becomes larger than \( J_{\text{Kirk}} \), \( W_{\text{cib}} \) increases; and when \( J_C \) becomes much larger than \( J_{\text{Kirk}} \), \( W_{\text{cib}} \) approaches \( W_C \). This also compresses and shrinks the depletion region inside the collector, causing a sharp increase in the B-C capacitance as demonstrated in Fig. 7.18. Transport of carriers through \( W_{\text{cib}} \) is via the slow diffusion process, which certainly enlarges the intrinsic forward transit time of the device leading to the fall-off of cutoff frequency (see Fig. 7.21). Reduction of \( f_T \) and increase of \( C_{\text{BC}} \) has the most severe degradation effect on the \( f_{\text{max}} \) of the transistor, as experimentally demonstrated in chapter 10. However, as shown in Fig. 7.21, the DC current gain does not show any sign of decrease when the Kirk effect takes place. This is in clear contrast to the trends observed in Si bipolar transistors [391]. This behaviour is demonstrated experimentally in chapter 10 and is described in the following paragraphs.

In Si homojunction bipolar transistors operating above the Kirk effect threshold current, the low diffusion constant of electrons makes the diffusion velocity across \( W_{\text{cib}} \) quite small. The electron and hole densities approach (or even exceed) the level of acceptor doping in the base. The effective base width becomes \( (W_B + W_{\text{cib}}) \), resulting in a drastic drop in both \( f_T \) and \( \beta_0 \) [391]. This makes operation above the threshold current highly undesirable.

The situation is different in III-V HBTs. Since the base region of III-V HBTs is very highly doped, the hole concentration in the \( p^+ \)-base is much larger than the hole density in the current-induced base (Fig. 7.20(b)). This forms a potential fall-off in both the valence and conduction bands between the two regions (Fig. 7.20(d)). This energy barrier decouples the electron concentrations inside the two regions; the electron concentration inside the \( p^+ \)-base is not influenced by the electron concentration inside \( W_{\text{cib}} \). Therefore, the recombination rate inside the base is not changed after the Kirk effect. Also the recombination rate inside \( W_{\text{cib}} \) is relatively small, because the hole concentration is extremely smaller than that in the \( p^+ \)-base, and the electron lifetime is much longer due to a much lower doping level. Thus, the additional base current that results from recombination in the current-induced base is negligible. This means that \( \beta_0 \) will not be degraded due to the base pushout, and DC current gain characteristics is not a reliable indicator of the onset of the Kirk effect.

![Figure 7.21- Variation of DC current gain and cutoff frequency with collector current density.](image-url)
PART III - EXPERIMENTAL RESULTS

This part starts with an introduction to the processing steps used in our research laboratory for fabricating both large- and small-geometry HBTs in Chapter 8. This chapter also discusses the mask design for planar microwave HBTs. Then in Chapter 9, the DC results measured on various types of HBTs will be presented, with an emphasis on demonstration of the advantages of InGaP/GaAs DHBTs over both InGaP/ and AlGaAs/GaAs SHBTs. Additionally, variation of HBT parameters with temperature will be presented in this chapter. Small-signal equivalent circuit modelling of HBTs, methods for extracting elements of this model, and analysis of the HBT delay times are the subjects of Chapter 10. Finally, the numerical simulation results combined with the developed small-signal parameter extraction technique will be used in Chapter 11 in order to optimise the high frequency performance of InGaP/GaAs DHBTs, while maintaining its superior DC characteristics.
CHAPTER 8.
FAMILIARISATION WITH DEVICE TECHNOLOGY

In this chapter, the basic device processing techniques used to fabricate large- and small-geometry HBTs are briefly outlined. Not much emphasis has been put on details of each stage of the processing, since these have been covered in other theses submitted in our research group [90,392]. Only those parts which include topics related to the device performance improvement, or those in which the author has major contribution are discussed in more details.

The chapter starts with a summary of the fabrication steps for large-geometry (DC) HBTs. Then, in section 8.2 issues related to metal contact formation, including ohmic contacts and their characterisation, are discussed. In section 8.3, a brief overview of chemical wet etching for GaAs and related materials is presented. Section 8.4 starts with the description of the basic steps involved in the fabrication of small-geometry HBTs, followed by more details as to the two important small-geometry processing steps, implant isolation and polyimide processing, which usually are not necessary for large-geometry HBTs. Finally, this chapter concludes with a description of the mask design and the modifications made in the present mask set.

8.1. Processing Steps for Large-Geometry HBT Fabrication

The fabrication of HBTs is carried out using standard photolithography and wet chemical etching processes. For the complete fabrication of a simple large-geometry HBT five masks are required; emitter metallisation mask, emitter mesa etch mask, base metallisation mask, base mesa etch mask, and collector metallisation mask. These fabrication steps are shown in Fig. 8.1, where step 0 corresponds to the initial epitaxial layers before any processing has begun. The first step is to metallise the n-type emitter ohmic contact using Ni/AuGe/Ni/Au. Next the wafer is etched until the base is reached, after which the p-type Au/Zn/Au ohmic contact is evaporated for the base. Finally, the wafer is etched to the sub-collector and the Ni/AuGe/Ni/Au n-type ohmic contact is formed. Once this is completed, the appropriate alloying condition must be used in order to optimise the ohmic contacts for minimum resistance. All the devices on a particular wafer fabricated using the above five masks will have a common collector, since the sub-collectors of the devices are not isolated. If isolated collectors are required, an extra mask for etching the sub-collector layer down to the semi-insulating substrate in the region between neighbouring transistors has to be added.

A top view layout of an HBT fabricated using the above mask set is shown in Fig. 8.2. The current mask is repeated along the wafer in three different sizes (emitter mesa diameters of 100μm, 140μm, and 170μm).
8.2. Ohmic Contact Formation and Characterisation

8.2.1. Ohmic Contacts

A good ohmic contact is essential for any semiconductor device, especially for high frequency HBTs where it is important to have low emitter contact resistance for high $f_T$ and low base contact resistance for high $f_{max}$. The primary requirements of ohmic contacts, apart from a low specific contact resistance, are an insignificant contact metal diffusion into the semiconductor both laterally and vertically, smooth surface morphology, reliability under severe environmental and bias stress conditions, and good thermal/electrical conductivities [213].
Fabrication of ohmic contacts in practice relies on formation of tunnelling M-S junction by creating a thin heavily doped semiconductor layer at the M-S interface. The heavily doped semiconductor layer could be formed during either epitaxial growth or contact preparation. The maximum achievable doping concentration of n-type GaAs during the growth for majority of existing growth systems is typically limited to about $5 \times 10^{18}$ cm$^{-3}$, and so this is the highest doping concentration that can be used for the capping layer and the sub-collector of HBTs to minimise contact resistance. Therefore, a metallisation scheme that contains a donor dopant is required in order to increase the doping concentration under the contact and enhance the tunnelling probability. Once the multi-component metal structure is deposited onto the semiconductor, thermal treatment is applied to drive the dopant in the metal layers into the semiconductor surface to create n$^{++}$ layer. The thermal treatment is normally accomplished by rapid thermal annealing (RTA). When the multi-component metal system is selected, the quality of the ohmic contact is largely determined by the alloying conditions.

Ni/AuGe/Ni/Au system is the most frequently used multi-component metal system for n-type ohmic contact on GaAs. The first Ni layer acts as a wetting agent and prevents the AuGe metal from “balling-up” during alloying. It also aids the in-diffusion of Ge into GaAs. Ge in AuGe film (88% Au, 12% Ge by weight) is used as n$^+$ dopant in the contact system. Ge is an amphoteric dopant for III-V compounds, thus acts as n-type dopant only if it occupies the Ga cation site. Au in the AuGe alloy allows the Ga atoms to diffuse into the Au, freeing Ga sites in the crystal for occupation by Ge. The maximum achievable doping limit to which Ge can dope GaAs is believed to be $\sim 2 \times 10^{19}$ cm$^{-3}$ [71: p.222]. The entire AuGe alloy must be evaporated to dryness (since Au evaporates before Ge), otherwise the resulting AuGe ratio is likely to be incorrect and difficult to reproduce. Another layer of Ni is evaporated, but this time has a ratio of 1:3 (Ni: AuGe). This ratio is reported to provide the lowest contact resistance, independent of the alloying time [393].

The Au overlayer performs several functions. First, it acts as a selective getter for Ga atoms, hence enhancing the out-diffusion of Ga and freeing Ga sites for Ge in-diffusion. Second, alloyed AuGe/Ni metallisation has poor sheet resistance (generally near 2 $\Omega$/sq.), and so an overcoat of gold is required to enhance sheet conductivity. To probe test devices having alloyed AuGe/Ni pads without a gold overcoat is difficult, as is to obtain accurate and reproducible results. The high sheet resistance of the metallisation makes such measurements sensitive to the exact placement of the probe on the pad. Also, the resistance of the probe-to-metal contact is overly sensitive to probe pressure. An overlaying layer of Au greatly alleviates these problems and also is essential for wire bonding. Finally, the extra gold results in improved surface morphology.

In this work, the n-type ohmic contacts to GaAs were made using the Ni/AuGe/Ni/Au metallisation system with typical thicknesses of 50/300/100/2000$\AA$. Metals were deposited on the wafer using an Edward 306 high vacuum evaporation system connected to a film thickness...
monitor (FTM). After metal deposition, the ohmic contact was formed by alloying in the RTA temperature of 400°C for 20 seconds.

There is not as much study for p-type ohmic contacts as for n-type, primarily because Fermi level is usually pinned closer to the valence band than the conduction band [394]. Additionally, the p-type base region of GaAs and InP HBTs can be doped to very high concentrations (usually exceeding $2 \times 10^{19} \text{ cm}^{-3}$). Therefore, base ohmic contacts are formed even when non-alloyed metallisations are used. Au/Zn/Au and Ti/Pt/Au are the most common p-type ohmic contact systems for GaAs HBTs. The Zn in the Au/Zn/Au system is intended to increase the surface doping of GaAs, thus reducing the specific contact resistance even further. Unfortunately, due to the high melting temperature of Pt, this metal has to be sputtered onto the wafer, and a sputtering machine was not available in our research group to take advantage of this metallisation scheme. Therefore, Au/Zn/Au (300/300/1000Å) metallisation is used for the p-type ohmic contacts in this work. To avoid cross contamination between n- and p-type metals, different evaporators are used for each type. All metals are evaporated at a pressure of $\sim 10^{-6} \text{ mbar}$.

8.2.2. Transmission Line Model (TLM)

Contact resistance and other associated parameters can be measured using appropriate test structures fabricated on the wafer. These measurements are important for in-process monitoring, and for assessing the effects of design or process variations. To extract the parameters associated with ohmic contacts from experimental measurements requires the assumption of a model establishing the relationship among the various parameters. Transmission line model for planar ohmic contacts is derived based on treatment of the contact as a transmission line [71: section 11.5]. The derivation of the model equations is discussed in details in the literature, e.g., in [71: section 11.5] [395]. Here, only the final equations with some useful definitions are repeated.

In figure 8.3 three ohmic contacts of length $d$ to a semiconductor layer are shown. It is

![Figure 8.3](image_url)

**Figure 8.3** Measurement of contact-to-contact resistance and end resistance for planar ohmic contacts.
assumed that this semiconductor layer is placed on top of an insulating layer, so that any current externally applied to ohmic contacts has to pass through the semiconductor layer directly underneath it. In order to measure the resistance between contacts 1 and 2, one has to pass a known current of $I$ through these two contacts and measure the voltage drop $V$ between them. However, using the same probes for voltage sensing as for applying current will result in erroneous resistance reading, since the voltage drop across the wiring and probes will be added to the voltage drop across the two contacts. Therefore, a method known as *four-probe measurement* is usually adopted, in which two separate pairs of probes are used for the purposes of voltage measurement and current passing. In this way, there would be no current passing through the voltage probes, which leads to accurate resistance readings. The measured resistance $R_{12} = V_{12} / I$ would be due to the voltage drop across the semiconductor between the contacts as well as the voltage drops across the semiconductor underneath the contact and the contact itself. The latter component of resistance is called the contact resistance, $R_C$, which is mathematically defined as:

\[
R_C = \frac{V(0)}{I(0)} \quad \text{with} \quad I(d) = 0
\]  

(8.1)

Consequently, $R_{12}$ can be written as:

\[
R_{12} = 2R_C + \frac{R_{sh}L}{W}
\]  

(8.2)

where $R_{sh}$ is the sheet resistance of the semiconductor and $W$ is the width of the contacts. It has been shown in [71: section 11.5] [395] that the contact resistance of a planar ohmic contact can be expressed as:

\[
R_C = \frac{R_{sk}L_T}{W} \coth \left( \frac{d}{L_T} \right) \quad \text{with} \quad L_T = \sqrt{\frac{\rho_c}{R_{sk}}}
\]  

(8.3)

where $\rho_c$ is the specific contact resistance, and $R_{sk}$ is the sheet resistance of the semiconductor layer underneath the contact. It is important to mention that the sheet resistance of the semiconductor under the contact, $R_{sk}$, might be quite different from that in between the contacts, $R_{sh}$, due to, for instance, ohmic contact alloying. The *transfer length*, $L_T$, is defined as the characteristic distance from the edge of the contact where $1/e$ of the total current has been transferred from the semiconductor to the contact. Another useful quantity to consider is the *end resistance*, $R_{end}$, which is fundamentally defined as [71: section 11.5]:

\[
R_{end} = \frac{V(d)}{I(0)} \quad \text{with} \quad I(d) = 0
\]  

(8.4)

It is also shown that [71: section 11.5]:

\[
\frac{R_C}{R_{end}} = \cosh(d/L_T)
\]  

(8.5)

Hence, end resistance relates $R_C$ and $L_T$ in a straightforward way. This quantity can be measured as indicated in Fig. 8.3, using a high impedance voltmeter between pads 2 and 3 so that no current flows. Hence the voltage measured is $V(d)$, i.e., $R_{end} = V_{23} / I_{12}$. Measured
value of $R_{\text{end}}$ combined with the measurement of $R_C$ and $R_{sh}$ (to be discussed later) provide a means by which one can evaluate $\rho_C$, $L_T$, and $R_{sk}$ using Eqs. (8.3) and (8.5). Before moving to the method of finding $R_C$ and $R_{sh}$, it is worth mentioning that another widely used method for evaluating $R_{\text{end}}$ is also proposed in [396]. Consider the resistor network shown in Fig. 8.4. Note that this equivalent resistor network satisfies both the definition of $R_C$ and $R_{\text{end}}$: appropriate measurement of contact resistance would yield $R_C$, and appropriate measurement of the end resistance would yield $R_{\text{end}}$. Examination of the figure reveals a straightforward method of measuring $R_{\text{end}}$. The resistances $R_{12}$, $R_{23}$, and $R_{13}$ are measured using four-probe method, and the end resistance is:

$$R_{\text{end}} = \frac{R_{12} + R_{23} - R_{13}}{2}$$  \hspace{1cm} (8.6)

The two methods of finding $R_{\text{end}}$ (i.e., from (8.4) and (8.6)) are compared later in this section.

The basic technique used to measure contact resistance of planar ohmic contacts employs a test pattern composed of differently spaced ohmic contacts as illustrated in Fig. 8.5. Ohmic contacts are formed on a semiconductor surface and separated by variable distances, $L_j$. The contacts have a width $W$ and the pattern is isolated to restrict current flow only across the ohmic contacts. The resistance between two such contacts measured using 4-probe technique can be explained by Eq. (8.2). Therefore, a plot of measured resistances as a function of spacing, $L_j$, yields a straight line with a slope equal to $R_{sh}/W$, and an R-axis intercept equal to $2R_C$ (Fig. 8.6). From Eqs. (8.2) and (8.3) with $d \gg L_T$, the intercept with the L-axis, $-L_x$, is related to the transfer length through:

$$-L_x = 2R_C - R_{sk}$$  \hspace{1cm} (8.7)
\[ L_x = 2R_C W/R_{sh} = 2L_T (R_{sk}/R_{sh}) \] 

(8.7)

Thus, the L-axis intercept would give the value of \( L_T \) if the sheet resistance under the contact were the same as the sheet resistance between the contacts. But, under practical condition of \( R_{sk} < R_{sh} \), additional measurement of \( R_{end} \) is required in order to calculate \( L_T \). Theoretically, when \( R_C, R_{sh}, \) and \( R_{end} \) are known, \( \rho_c, L_T, \) and \( R_{sk} \) can be obtained from (8.3) and (8.5). But, practical considerations render accurate evaluation of the above parameters rather difficult. For instance, the end resistance measurement is useful only if \( d \) is not greatly larger than \( L_T \), otherwise \( R_{end} \) would be extremely small. For the base region of HBTs, \( L_T \) is on the order of a micron, which requires almost prohibitively small contacts. As a practical matter, to adopt (incorrectly) the assumption that \( R_{sk} = R_{sh} \) is sometimes useful. To compute \( \rho_c \) and \( L_T \) based on this assumption is then straightforward. These difficulties, coupled with experimental error (see below), make accurately determining \( \rho_c \) exceedingly difficult for alloyed ohmic contacts to planar structures. Claims of specific contact resistances of less than about \( 1 \times 10^{-6} \Omega \text{cm}^2 \) should be examined critically.

The measurement technique described above is quite sensitive to measurement errors. Because the R-axis intercept (equal to \( 2R_C \)) is usually small compared to the measured resistances, minor errors in resistance or contact spacing can translate into significant percentage errors in \( R_C \). In addition, when the contact length \( d \) is larger than the transfer length by a factor of >3, the measurement of \( R_{end} \) becomes practically impossible. For more accurate results, end resistances must be measured directly using (8.4) rather than (8.6). Since \( R_{13} \) and \( (R_{12} + R_{23}) \) are both much larger than \( R_{end} \), any small error in measuring those resistances will lead to extremely large relative error in \( R_{end} \). Practically, sometimes even negative values of \( R_{end} \) may be obtained using (8.6). To elaborate this, the measured sub-collector TLM data of Fig. 8.6 is repeated in Table 8.1, together with the data to evaluate \( R_{end} \) from the two methods, i.e., from Eqs. (8.4) and (8.6). For both methods, \( R_{sh} \) and \( R_C \) are obtained from the gradient and R-intercept of the graph in Fig. 8.6. When \( R_{end} \) is calculated from (8.6), the measured resistances \( R_{ij} \) are much larger than \( R_{end} \), and consequently, the resultant \( R_{end} \)’s are erroneous. For instance, \( R_{end3} \) is negative and \( R_{end7} \) is even larger than \( R_C \). Equation (8.4), which directly measures \( R_{end} \), is shown to produce more reliable data.
collector TLM in fact produces the most reliable data among the three HBT ohmic contacts. Since $R_{sh}$ (and consequently $R_{sk}$) values are significantly larger for the emitter and base ohmic contacts, their $L_T$’s (from Eq. (8.3)) are on the order of few microns, which are much smaller than $d = 50 \mu m$. Eq. (8.5), therefore, leads to an extremely small $R_{end}$ (on the order of few mΩ or even less), which can not be measured precisely. Even collector ohmic contacts of better quality ($\rho_c < 5 \times 10^{-6} \Omega \cdot cm^2$) are practically impossible to characterise using large TLM pads.

<table>
<thead>
<tr>
<th>Spacing ($\mu m$)</th>
<th>$R$ ($\Omega$)</th>
<th>$R_{end}$ ($\Omega$) from Eq. (8.6)</th>
<th>$R_{end}$ ($\Omega$) from Eq. (8.4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>$R_{12} = 2.39$</td>
<td>$R_{12} = 5.46$ $\rightarrow$ $R_{end2} = 0.470$</td>
<td>$R_{end2} = 0.22$</td>
</tr>
<tr>
<td>20</td>
<td>$R_{23} = 4.01$</td>
<td>$R_{24} = 9.23$ $\rightarrow$ $R_{end3} = -0.175$</td>
<td>$R_{end3} = 0.18$</td>
</tr>
<tr>
<td>30</td>
<td>$R_{34} = 4.87$</td>
<td>$R_{35} = 9.92$ $\rightarrow$ $R_{end4} = 0.495$</td>
<td>$R_{end4} = 0.16$</td>
</tr>
<tr>
<td>40</td>
<td>$R_{45} = 6.04$</td>
<td>$R_{46} = 11.95$ $\rightarrow$ $R_{end5} = 0.630$</td>
<td>$R_{end5} = 0.22$</td>
</tr>
<tr>
<td>50</td>
<td>$R_{56} = 7.17$</td>
<td>$R_{57} = 14.55$ $\rightarrow$ $R_{end6} = 0.665$</td>
<td>$R_{end6} = 0.27$</td>
</tr>
<tr>
<td>60</td>
<td>$R_{67} = 8.71$</td>
<td>$R_{68} = 16.57$ $\rightarrow$ $R_{end7} = 0.785$</td>
<td>$R_{end7} = 0.20$</td>
</tr>
<tr>
<td>70</td>
<td>$R_{78} = 9.43$</td>
<td>$R_{end(Average)} = 0.478$</td>
<td>$R_{end(Average)} = 0.21$</td>
</tr>
</tbody>
</table>

$R_{sh} = 11.72 \Omega/sq.$  
$R_{sk} = 0.95 \Omega/sq.$  
$R_{sk} = 2.50 \Omega/sq.$  
$R_C = 0.70 \Omega$  
$\rho_c = 2.75 \times 10^{-5} \Omega \cdot cm^2$  
$\rho_c = 1.78 \times 10^{-5} \Omega \cdot cm^2$  
$L_T = 53.8 \mu m$  
$L_T = 26.7 \mu m$

In general, the above difficulties of finding the contact parameters arises from small values of $R_C$ and $R_{end}$ as compared to the semiconductor resistance for large pads used as TLM pattern [397]. One way to avoid this problem would be to use small contact pads, but then direct probing would be difficult. A modified TLM pattern is designed in this work (see the mask design in Fig. 8.16) in which narrow metal contacts are attached to large pads for probing along the sides of the TLM structure. However, this TLM structure was proved incompatible with our HBT fabrication process, and modifications in the masks are necessary before one can take advantage of the small-TLM design. An attractive alternative TLM structure is recently proposed by Sawdai et al. [395] which makes more accurate evaluation of small-size contact parameters possible. This technique, called $L_T$-TLM by the authors, was shown to provide very accurate determination of $L_T$ (within 5%) and $\rho_c$ (within 12%) for planar contacts. Further investigation is necessary to evaluate this technique and its feasibility for our small-geometry fabrication process.

8.3. Chemical Wet Etching

In microelectronic processing, etching in liquid or gaseous form is utilised to remove material not shielded by the photoresist or any other protecting material, e.g., a metal. Wet etching makes use of liquid etchant to remove material. Generally, the etchant contains an oxidising agent (usually $H_2O_2$) for oxidising the material surface and a dissolving agent for dissolving
the resultant oxide. Acids and bases are both used as dissolving agents, with the former being more popular. Often the etching solution also contains a certain portion of water mainly for dissolving the reaction products and for diluting the solution in order to slow down the etching rate. Wet etching tends to be isotropic with a close to unity thickness to width aspect ratio. The undercutting of the masking pattern is sometimes useful, such as in the case of self-aligned B-E junctions or where reduction of B-C capacitance by collector undercutting is required \[85,94\]. However, the large amount of undercut caused during the wet etching process may impose a serious problem in processing of very small-geometry devices. In such circumstances, a dry etch process (i.e., etching in gaseous form), which can result in highly anisotropic vertical etching profiles may be adopted. In this work, all the mesa etching processes are carried out using wet etching.

One of the hardest steps in the fabrication of HBTs is etching to the base. One can easily miss the base due to its small thickness of typically 1000Å or less. This is a particularly important problem for the fabrication of AlGaAs/GaAs HBTs. Usually an etching solution that etches GaAs, also etches AlGaAs and InGaAs, although the etch rates might be slightly different. However, the \[NH_4OH: H_2O_2 (1: 200)\] solution is shown to etch GaAs, but not AlGaAs or InGaAs \[44: section 14.1\]. The selectivity, that measures the ability of the etchant to remove the top layer compared to the layer underneath, depends on the aluminium and indium concentrations. For a typical AlGaAs emitter whose Al mole fraction is 0.30 to 0.35, this 1: 200 etching solution is highly selective in removing GaAs cap layer and stops at the AlGaAs emitter \[44: section 14.1\]. But, a similar wet etching solution to selectively remove the AlGaAs emitter on top of the GaAs base does not exist. In such cases, the etching has to be carefully monitored in order to avoid overetching the base. This is usually carried out by breakdown voltage measurement technique \[90\] and/or direct thickness measurement using Tallystep. However, this method is a tedious and time consuming process. This technique is not appropriate for a fully automatic fabrication procedure and may well lead to non-uniformity across the wafer and unreliable devices. Therefore, it is highly desirable to have a selective etch process between the emitter and base materials. InGaP/GaAs and InP/InGaAs HBTs are advantageous over the more conventional AlGaAs/GaAs devices in that etching the emitter and base materials is highly selective. Referring to Table 8.2, any typical wet solution containing HCl removes InP and InGaP, either with or without an oxidising agent such as \(H_2O_2\). But, although HCl can be one ingredient of an etch to remove (In)GaAs, the solution must also contain an oxidising agent to initiate the etch of (In)GaAs \[44: section 14.3\]. Consequently, a solution containing HCl but not \(H_2O_2\) removes the In(Ga)P layer but does not attack the (In)GaAs, with a selectivity approaching infinity. Similarly a solution such as \(H_2SO_4: H_2O_2: H_2O\), which does not contain HCl, etches (In)GaAs but does not attack In(Ga)P. Because of the high degree of selectivity, the etch rate of In(Ga)P by a HCl solution needs not be carefully calibrated. A pure HCl solution is shown to have InGaP etching rates of \(>200 \text{ Å/sec}\) \[44,398\]. Therefore, etching the InGaP emitter layer of typical HBTs should
require no more than 10 seconds. However, to avoid the possible “islands” of InGaP [31], an InGaP etch duration of 30-60 seconds was used in this work.

Table 8.2 - Different types of etching solutions, etch rates, and materials etched [44,90,392,398].

<table>
<thead>
<tr>
<th>Etching Solution</th>
<th>Etching Ratio</th>
<th>Etch Rate (Å/sec)</th>
<th>Materials Etched</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH$_4$OH: H$_2$O$_2$: H$_2$O</td>
<td>8: 3: 400</td>
<td>30-40</td>
<td>GaAs, AlGaAs</td>
</tr>
<tr>
<td>NH$_4$OH: H$_2$O$_2$</td>
<td>1: 200</td>
<td>—</td>
<td>GaAs</td>
</tr>
<tr>
<td>H$_2$SO$_4$: H$_2$O$_2$: H$_2$O</td>
<td>1: 1: 20</td>
<td>~70</td>
<td>GaAs, AlGaAs, InGaAs</td>
</tr>
<tr>
<td>HF: H$_2$O$_2$: H$_2$O</td>
<td>1: 1: 10</td>
<td>~60</td>
<td>GaAs, AlGaAs</td>
</tr>
<tr>
<td>HCl</td>
<td>concentrated</td>
<td>&gt;200</td>
<td>InP, InGaP</td>
</tr>
<tr>
<td>HCl: H$_2$O</td>
<td>3: 1</td>
<td>15</td>
<td>InP, InGaP</td>
</tr>
<tr>
<td>HCl: H$_3$PO$_4$</td>
<td>1: 3</td>
<td>~30</td>
<td>InP, InGaP</td>
</tr>
</tbody>
</table>

8.4. Small-Geometry HBT Fabrication

Fabrication of small-geometry microwave HBTs is different from the large-geometry DC processes in many ways. First, for the reasons highlighted in subsection 8.4.1, a planar structure is highly desirable for MMIC applications. Therefore, ion bombardment of the inactive regions, rather than the non-planar mesa etching technique, is usually preferred for device isolation. Second, the high frequency performance of HBTs is very much sensitive to the parasitic resistances and capacitances of the device. Consequently, special techniques, such as self-aligned processing [72,399], must be employed in order to reduce such parasitics. Obviously, low resistance ohmic contact requirements become more critical for small-geometry devices. Thirdly, due to the small size of the active region, it is not possible to directly probe the ohmic contacts of the device. Therefore, connections to large contact pads (perhaps using a second layer of metallisation) are necessary in order to have access to the outside circuitry. For standard on-wafer S-parameter measurement on test devices, such contacts ought to be of a special ground-signal-ground (GSG) shape as shown in Fig. 8.8. Finally, due to the layout considerations and complexity of the process, metal crossovers are inevitable. Hence, more than one layer of metallisation with appropriate insulation between the metals is almost always required.

The above requirements lead the designer to use a fabrication process which includes a significantly larger number of masks than the simple five-mask process discussed in section 8.1. In this work, a relatively small number of masks is used to fabricate small-geometry HBTs purely for on-wafer microwave testing purposes. Obviously, fabricating monolithic circuits requires additional masks for defining passive elements, air bridges, etc. Also heat dissipation is a serious issue for high power HBT fabrications which necessitates additional backside processing steps [71: Chap.16]. The steps used in the present work for a planar self-aligned HBT fabrication process are summarised below and also visualised in figure 8.7.

1. implant isolation
2. emitter metallisation
3. emitter mesa etch and base metallisation (self-aligned)
4. base mesa etch and collector metallisation (self-aligned)
5. first level patch metallisation (to connect base and emitter fingers with pad metals)
6. polyimide spinning and window opening
7. second level interconnect and pad metallisation
The fabrication process starts with the definition of device active area. This is achieved by bombarding the region outside the active area by energetic He\(^+\) and O\(^+\) ions (Fig. 8.7(a)). This step is discussed in full details in section 8.4.1. Then, emitter metallisation window is defined by photolithography and Ni/AuGe/Ni/Au is evaporated as emitter ohmic contact (Fig. 8.7(a)). This is followed by wet etching of the cap layer and emitter region defining the emitter mesa. Using the same window opened for emitter mesa etching, Au/Zn/Au metal system is evaporated as base ohmic contact (Fig. 8.7(b)). In this way, base metal is self-aligned to the emitter mesa, which is very important for minimising the extrinsic base resistance and B-C depletion capacitance. Next, a similar process is repeated to define the base mesa, this time the collector ohmic contact is evaporated self-aligned to the base mesa (Fig. 8.7(c)), thus...
minimising the collector series resistance. Once the main structure of the device is defined using steps 1-4 above, a Ti/Au metallisation is carried out to connect individual emitter and base fingers of the device. The reason for using a separate metallisation to connect the emitter and base fingers is explained in subsection 8.4.3. Wafer planarisation and passivation is carried out using polyimide spinning on wafer (Fig. 8.7(d)), followed by window opening in polyimide for subsequent second level metallisation (Fig. 8.7(e)). This second level of metal connects the small-geometry device to the large contact pads for possible probing or wire bonding.

One important issue in the fabrication of self-aligned GaAs-based HBTs is the effect of wafer orientation on device fabrication. Most GaAs wafers are grown in the [100] direction so that alternative layers of group III atoms and group V atoms are grown on top of each other. For substrates in such an orientation, the major and minor flats are usually perpendicular to [011] and [011] directions, respectively [71]. The emitter pattern in the stripe geometry can be placed in parallel with either the major flat or the minor flat (Fig. 8.9). It is a GaAs material property that the wet etching in certain crystallographic planes occurs faster in preference to some others. This preferential (or anisotropic) etching property results in different etch profiles. When the emitter stripe is parallel to the minor flat, the resulting emitter mesa is narrower at the bottom than at the top. The converse is true when the emitter stripe is parallel to the major flat. (Note that in both cases the masking material is undercut by the wet etching.) This difference has a major implication in the device yield if the base metal is self-aligned to the emitter mesa. If the etch edge extends outward, a portion of the base contact metal may land on the emitter layer, shorting the two layers. This shorting effect can be avoided if the emitter stripe is fabricated in parallel with the minor flat to have an inward edge profile. This is depicted in the layout of the fabricated HBTs shown in Fig. 8.7(e), where '+' and '-' signs correspond to the positive and negative etch profiles as discussed above. For InGaP/GaAs HBTs, highly selective nature of the emitter etching allows an extensive overetch of the GaAs on top of InGaP layer, which pushes the emitter edge more and more inward [46]. Also the etching of InGaP in HCl solutions is shown to be vertical [44,46]. Therefore, the requirement of orienting the emitter stripe prior to fabrication is partially relaxed for self-aligned InGaP/GaAs HBTs.
8.4.1. Implant Isolation

Isolation serves a number of purposes. In active devices, it restricts the current flow to the desired path, disconnects the electrical path between neighbouring devices, reduces parasitic capacitances and resistances, and also provides a sufficiently insulating surface for construction of passive elements such as capacitors and transmission lines required in monolithic circuits [71: Chap. 10]. Device isolation can be achieved using either mesa etching down to the semi-insulating substrate, or by ion bombardment. Mesa etching, though simpler than the ion bombardment technique, has some drawbacks especially when used as part of an MMIC process. One disadvantage is the lack of planarity after mesa formation, especially that device isolation is generally one of the early (if not the earliest) stages of device fabrication. Mesa isolation in HBT wafers usually requires an etch depth of more than 1.5μm. This nonplanarity is enough to cause difficulties in subsequent resist applications, pattern exposure, and metal step coverage. Resist thickness will not be uniform near a mesa step, resulting in patterning variations. In contact photolithography, which is used in the present work, the mesa step means that the mask cannot be in good contact with material just off the mask edge. Additionally, the isolation produced by mesa etching is not as good as that achieved by ion bombardment [71: Chap. 10]. Therefore, in the present work ion bombardment is used for isolation of GaAs-based HBTs. However, due to the difficulties in implant isolation of InGaAs HBTs, mesa isolation currently is the only remaining choice for these HBTs [400].

The ion implantation process causes considerable damage to the crystal lattice. This damage creates many carrier trapping centres and renders the damaged material insulating. However, as-implanted samples usually have a large density of traps. While all the carriers are trapped at defect sites, these sites are closely spaced leading to “hopping” conduction from one site to another [400]. A post-implant annealing removes some of these traps, which reduces the probability of hopping, and leads to an increase in sheet resistivity of the implanted areas [400]. The sheet resistance increases with anneal temperature until it reaches a peak value in the temperature range 500-600°C. Any further increase in annealing temperature leads to annihilation of deep level trap centre concentration below that of the device carrier concentration, resulting in reduction of sheet resistance. In this work a post-implant anneal at
500°C for 60 seconds is used.

Since the implant isolation is the first step of our fabrication process, the (desirable) planarity means there is no visible pattern to use in subsequent alignments. Therefore, some alignment marks have to be fabricated prior to implantation by either etching or metal deposition. In this work, metal deposition is used, since alignment using metallic marks is somehow easier than using etched patterns.

Several species of ions have been used for implant isolation, including protons (H\(^+\)), helium (He\(^+\)), oxygen (O\(^+\)) and boron. The first two are light ions. Consequently, they are used for deep isolation, but the damage per ion is relatively low, so relatively high doses of implantation are required for these ions. The latter two ions are heavier and can be used for shallow implant damages. Usually a single energy/dose implant results in a Gaussian damage distribution, which is not sufficient for isolation of the entire thickness of an HBT epilayer (>1.5\(\mu\)m). Therefore, multiple energy/dose bombardments are usually employed to isolate the entire thickness of HBTs [80], although single MeV bombardments have recently been suggested for isolation of relatively thick and very highly doped materials [400].

Device isolation in this work was achieved by a multiple energy/dose O\(^+\) and H\(^+\), or O\(^+\) and He\(^+\) ion bombardment. The O\(^+\) is preferred to achieve isolation of the region up to ~0.5 \(\mu\)m deep, where the most highly doped layers are contained, since its carrier removal efficiency is much higher than for the light ions. The remaining thickness of the HBT must be isolated with higher doses of lighter species. The implant schedule for both O\(^+\)/H\(^+\) and O\(^+\)/He\(^+\) bombardments was simulated and optimised using the TRIM software at University of Surrey Ion Beam Facility. The O\(^+\) implant schedule consists of 50 keV at 2\(\times\)10\(^{12}\) cm\(^{-2}\), 100 keV at 4\(\times\)10\(^{12}\) cm\(^{-2}\), 200 keV at 6\(\times\)10\(^{12}\) cm\(^{-2}\), and 300 keV at 8\(\times\)10\(^{12}\) cm\(^{-2}\). The H\(^+\) was implanted at 100 keV to a dose of 3\(\times\)10\(^{14}\) cm\(^{-2}\), 150 keV to 4\(\times\)10\(^{14}\) cm\(^{-2}\), and 200 keV to 5\(\times\)10\(^{14}\) cm\(^{-2}\). As the He\(^+\) is heavier than H\(^+\) it needs more energy to completely isolate the conducting layers. Therefore, the following implant recipe was performed in addition to the implantations used for H\(^+\): He\(^+\) energies of 250 keV, 300 keV, 350 keV, each to a dose of 5\(\times\)10\(^{14}\) cm\(^{-2}\), and 450 keV energy at 7\(\times\)10\(^{14}\) cm\(^{-2}\). Sheet resistance of the implanted cap layer, base, and sub-collector layers were measured using special test structures (as described in subsection 8.4.4) to be all greater than 1.5\(\times\)10\(^{8}\) \(\Omega\)/sq., producing excellent device isolation.

Formation of C-H complexes and their corresponding reliability issues were briefly discussed in section 2.4. Passivation of carbon acceptors inside the base of HBTs by H incorporated during the growth of epitaxial layers is known to affect the current gain, base series resistance, and \(f_{\text{max}}\) in all the devices on the wafer, regardless of their size. In contrast, it is shown in this work that device isolation by H\(^+\) ion bombardment forms another source of H incorporation which affects the device characteristics non-uniformly, depending on its actual geometry [72]. In the case of small-geometry HBT isolation, diffusion of hydrogen from the implanted regions towards the active base area during the high temperature post-implant annealing
process can passivate the carbon dopants in the base. This effect is expected to be more significant for smaller geometry transistors. Figure 8.10 shows the variation of current gain with emitter width for both O⁺/H⁺ and O⁺/He⁺ implanted HBTs. The current gain of O⁺/He⁺ is almost independent of the actual geometry of the transistors. It actually shows some reduction for smaller geometry devices due to the emitter-size effect. On the other hand, the O⁺/H⁺ devices show a large variation in current gain, changing from 18 for an 80×160 µm² HBT to more than 100 for a 4×10 µm² device. This is attributed to the passivation of carbon dopants by hydrogen diffused from the sidewalls of the active region. The diffusion constant of H in GaAs is given as [80]:

\[ D_H = 1.5 \times 10^{-5} \exp(-0.62eV/kT) \text{ cm}^2/\text{s} \]  

(8.8)

For the annealing temperature and time used in this work, a diffusion length of ~2.9 µm will be obtained. Therefore, one expects the 4×10 µm² device to be severely affected by the H⁺ diffusion, which is exactly what Fig. 8.10 suggests. The much higher current gain in the small geometry devices compared to large ones is due to the fact that effective base doping and base bulk recombination current component are reduced in the former case by the presence of the hydrogen in almost entire area of the transistor. Even if the entire base region of the device is not affected by H⁺ diffusion, the concentration of active C acceptors will vary inside the base of the device. This leads to a variation in the B-E turn-on voltage along the base, causing current localisation and, eventually, device burn out. Therefore, bombardment by He⁺ ions is preferred to H⁺ ions in the present work.

An almost 3 µm thick photoresist (AZ1329) is used as masking material for ion bombardment to avoid any implantation damage in the active part of the device. Ion implantation, however, hardens the photoresist rendering it very difficult to remove after the bombardment process. Therefore, a thin Si₃N₄ layer, sometimes referred to as a release layer, is deposited on the wafer by plasma-enhanced chemical vapour deposition (PECVD) prior to the photoresist spinning. Once the implant isolation is finished, the thick photoresist is removed by dry etching in O₂ plasma (i.e., “ashing”) [71: section 9.2], while the Si₃N₄ layer protects the semiconductor surface. Then, the silicon nitride layer itself is removed by wet etching in HF: H₂O (1:10) solution.
Just to conclude this subsection, a summary of the entire device isolation mask level is given as follows. First, Ti/Au alignment marks are fabricated, followed by PECVD of a 300Å Si$_3$N$_4$ release layer. Then the wafer is covered by ~3 μm photoresist and bombardment windows are opened by photolithography. After the multiple energy, multiple dose O$^+/\text{He}^+$ implant isolation is carried out, the thick photoresist and the Si$_3$N$_4$ layers are removed by dry etching in O$_2$ plasma and wet etching in HF solution, respectively. Finally, the wafer is annealed for 1 min at 500°C.

8.4.2. Polyimide Processing

Dielectric films are used in GaAs processing for environmental encapsulation, surface passivation, wafer planarisation, capacitor dielectrics, and metal crossover insulators [71: section 13.3] [401-402]. Silicon dioxide, silicon nitride, and polyimide are the most favourite dielectrics used for the above purposes. While SiO$_2$ and Si$_3$N$_4$ are usually deposited on the wafer using sputtering or PECVD, polyimide can be easily applied to the wafer by spinning and baking (similar to photoresist coating process). Polyimide also has a smaller relative dielectric constant (3-4.5) compared to both SiO$_2$ (4-5) and Si$_3$N$_4$ (5.5-7.5) [71: p.295]. Therefore, it would lead to smaller parasitic capacitances when used as the metal crossover insulator. Additionally, the planarisability of polyimide is superior to that of SiO$_2$ and Si$_3$N$_4$ [392], resulting in a better step coverage. Resistivity (~3×10$^{15}$ Ω.cm) and dielectric breakdown field (10$^6$ V/cm) of polyimide are also very high [392]. Therefore, since in this work a dielectric film is mainly used to planarise the wafer and act as an insulator between metal layers, and also due to its ease of deposition, polyimide is preferred to the other two common dielectric films. In MMICs including a large number of capacitances, where reduction of capacitance area is an issue, perhaps Si$_3$N$_4$ or Ta$_2$O$_5$ [71] with larger dielectric constants may be better choices. Also for the purpose of surface passivation, sulphur containing passivation layers [403-404] or the emitter edge thinning design (see section 2.2.4.b) are quite popular. Finally, in cases where a good thermal conductivity is a major requirement of the dielectric layer, other passivation films such as SiC and amorphous Si have shown much superior performance [402].

The polyimide processing involves three steps, namely polyimide coating, curing, and etching. As mentioned earlier, deposition of polyimide was performed using the same spinning process as used for photoresist application. The spin speed and duration was adjusted to give an almost 0.7 μm thick polyimide film. Curing is simply baking the polyimide layer to drive off its solvents and moisture. If the heating process is overdone, polyimide becomes a highly stable permanent feature that is almost impossible to remove. But when cured to a lesser extent, polyimide may be patterned or removed by ashering. In this work, Hitachi PIQ 13 polyimide is spun for 30s with the spin speed of 3000 rpm. Then curing is performed by baking the sample at 300°C for 2 hours, followed by photolithography to define the via hole windows. More details as to the polyimide processing used in our research group can be
found in [90,392].

8.4.3. Mask Design

The existing mask set prior to the start of this work had some deficiencies related to the device parasitics and yield. Therefore, a new set of masks was designed to improve the shortcomings of the old set, which are explained in this subsection. One full block of this new mask is shown in Fig. 8.11. Apart from the bottom one-third of this mask and the alignment marks, the rest of the mask is covered with special microwave test HBTs with GSG-shaped contact pads. Many different geometries are tried to find the optimum device layout, as well as to study the geometry dependence of parasitic elements (see section 10.5). The bottom-right corner of the mask includes various geometries of devices without the GSG pads, which may be used for DC measurement, and also may be individually scribed and packaged for discrete applications. The bottom-left corner of the mask contains some test structures which are to be discussed in more details in the next subsection.

A close-up view of one of the small-geometry HBTs on the new mask is depicted in Fig. 8.12. This is a single emitter-finger device which has one base and collector contacts on each side. As also shown in Fig. 8.7(e), the two base fingers are connected together with a U-shaped metal. The entire part of this U-shaped metal may be defined by base ohmic contact metal, as was done in the old mask. However, the U-shaped metal was proved problematic for the fabrication of small-geometry HBTs. During the photolithography of the U-shaped base contacts with long and narrow base fingers, irregular window openings and poorly defined base fingers were formed, leading to short-circuiting of the B-E junction upon metal deposition. This problem was recognised as the main cause for the extremely low yields of the devices fabricated using the old mask set. To circumvent this difficulty, instead of the U-shaped base, only the two long base fingers were defined during the base ohmic contact metallisation. Then, the individual base fingers were connected together using the patch metal in level 5 of the small-geometry device fabrication process (see Fig. 8.14). As a result, the edge definition and the wafer yield were significantly improved. The same method may be used to connect different emitter fingers of a multi-finger HBT; a patch metal is used to connect the individual emitter fingers and also to connect them to the contact pads formed by the second layer of metal.

Unlike the base ohmic contact that has a small transfer length, $L_T$, on the order of 1-3 μm, collector ohmic contact may have a large $L_T$ in the range 10-30 μm (see the results in Table 8.1). This is due to the higher mobility of electrons compared to holes, and the thicker sub-collector of HBTs compared to the base region, both leading to a much smaller $R_{sb}$ (and $R_{sk}$) for the sub-collector than for the base region (see Fig. 8.6). Therefore, in order to efficiently collect all the electrons inside the sub-collector region, the collector ohmic contact must be much wider than the base metal stripe. Considering the fact that a wider collector contact does not affect the B-C area of the HBT (which is defined by the base mesa), it was decided to
make the collector ohmic contact much wider than the one designed in the old mask. However, this must not be overdone, since the fringing capacitances of the collector metals may become important. The new design for the collector ohmic contact is also depicted in Fig. 8.14.
When two collector contacts, instead of one, on both sides of the intrinsic area are used, the collector series resistance decreases, while the B-C capacitance becomes larger. It can be shown that the reduction of resistance for double-sided contacts is by a factor of 2-4, depending on the actual device geometry [405]. However, the B-C capacitance only increases by a small percentage due to the extra base metal-to-mesa spacing required for the extra collector ohmic contact. Although the high frequency performance of the HBTs is more sensitive to $C_{pc}$ rather than $R_{cc}$ (see section 10.5), in majority of geometries it was found that a double-sided collector ohmic contact actually improves $f_T$ and $f_{max}$ of the transistor. Therefore, most of the devices on the current mask have double-sided collector contacts.

The new mask also includes few devices designed identical to those of the old mask. Figure 8.13 shows the layout of an $8 \times 10 \, \mu m^2$ device designed in the same fashion as in the old mask. The device has only one collector contact which is much smaller than the contact used in the new mask transistors. It is clear in both designs that in order to access the collector ohmic contact(s), one has to cross over the emitter interconnect. This crossover area, which is $10 \times 10 \, \mu m^2$ in both cases, has an effective parasitic capacitance of 9.0-10.5 fF. For double-collector
devices, two crossovers are necessary. Although the area of each crossover, and hence the capacitance per crossover, is the same in both designs, but the way the emitter "bridge" is approaching the crossover area is different in the two designs. In the old mask, a straight line bridge is used, while this is replaced by a tapered shape in the new design. This should reduce the bridge resistance significantly. The average measured bridge resistances on the old and new designs were 2.7 and 1.0 $\Omega$, respectively. Additionally, the emitter bridge, which is made from second layer of metallisation, should be connected to the emitter contact via a window opening in the polyimide. This interconnection between the two layers of metal is quite resistive, especially for small polyimide windows. The new design for the emitter bridge also uses a wider polyimide window and thus results in a smaller emitter metal interconnect resistance. A special test structure is designed to measure the interconnection resistance between the two layers of metal (see Fig. 8.16). The measured values of resistance per interconnect (see next subsection) for the old mask was $\sim 2 \Omega$, compared to less than 1 $\Omega$ for the new design.

![Figure 8.15](image)

**Figure 8.15** - Comparison between the high frequency performance of 8×10 $\mu m^2$ InGaP/GaAs SHBTs fabricated using the (a) new and (b) old mask designs. Figure 8.15 compares the high frequency performance of two InGaP/GaAs single HBTs fabricated on the same wafer using the new and old mask layouts. It can be seen that the cutoff frequency is improved from 39 to 45 GHz when the new mask layout is used. This corresponds to 0.54 ps reduction in the total delay time, which is mainly due to the reduced parasitic resistances. The $f_T$ for the old mask device peaks at slightly lower value of collector current, which is due to the larger emitter and collector series resistances of this device. This also contributes to the larger delay time in the old design HBT. However, since $f_{max}$ is mostly sensitive to the B-C capacitance (which is smaller for the old mask layout), its value is not changed as such. The level of modification in the high frequency performance is still modest, because both these devices have relatively large intrinsic delay times, which are basically related to the identical vertical structures of the two devices. When more advanced vertical
structures are utilised, the improvement in the high frequency performance by using the new mask design is expected to be more pronounced. The high frequency performance of HBTs and the methods to extract their small-signal equivalent circuit elements are discussed in more details in chapter 10.

8.4.4. Test Structures

The designed mask consists of some test structures to evaluate both the process and the parasitic elements. It also consists of devices designed in configurations other than the conventional common-emitter single transistor HBTs. To discuss these test structures, the bottom-left corner of the designed mask in Fig. 8.11 is magnified and repeated in Fig. 8.16.

The interconnect test structure (marked with #2) is designed to measure the effective resistance of the polyimide via holes connecting the two levels of metallisation. It consists of a series connection of 14 identical interconnects between the two metal levels. The size of each interconnect is designed to be similar to the polyimide window on the emitter bridge. Measurement of the resistance of this test pattern gave $(34 \pm 5) \Omega$ and $(13.6 \pm 0.2) \Omega$, corresponding to a resistance per interconnect of $\sim 2 \Omega$ and less than $1 \Omega$, for the old and new masks, respectively (considering the finite resistance of the metal bars themselves).

Isolation test patterns (#3) are fabricated on entirely ion implanted areas of the emitter (cap layer), base, and sub-collector layers of the HBT. They are designed to assess the effectiveness of isolation by ion bombardment. Using the He$^+/O^+$ implant recipe given in subsection 8.4.1 for a typical HBT fabricated in this work, resistances of $(15 \pm 7) \text{M}\Omega$, $(10 \pm 4) \text{M}\Omega$, and $(35 \pm 7) \text{M}\Omega$ were measured for the emitter, base, and sub-collector isolation structures, respectively. This corresponds to a minimum sheet resistance of almost $1.5 \times 10^8 \text{\Omega/sq.}$, which provides excellent device isolation.

The crossover capacitance test structure (#9) consists of a number of (10 or 20) metal crossovers connected in parallel, each having $10 \times 10 \mu\text{m}^2$ area. The capacitance measurement on this structure gave $(9-10) \text{fF}$ per crossover. This is close to the value of $\sim 10.5 \text{fF}$ estimated from open test structures (see section 10.3.1 and Table 10.2), but is somewhat larger than the value calculated from the simple parallel plate capacitance formula ($\sim 5.7 \text{fF}$).
Common-base HBTs are suitable candidates for millimetre-wave power applications [406]. Few devices of this kind were included in the present mask (#1) for initial assessment. In these devices the large ground contact pads represent the base terminal rather than the emitter. It was observed that the $f_{\text{max}}$ in common-base HBTs is comparable to that of common-emitter devices. Few Darlington pairs (#7) are also designed in the current mask. A pair with an $(8\times10)$ $\mu$m$^2$ input transistor and a 4-finger $(8\times10)$ $\mu$m$^2$ output transistor fabricated on an InGaP/GaAs DHBT wafer has shown current gain in excess of 2000 (Fig. 8.17). But, the high frequency performance of this Darlington pair is not as good as the single-transistor HBTs. To take the most advantage of Darlington pairs, the input transistor must be significantly smaller than the output transistor.

The open and short test structures (#5) are used to evaluate and de-embed the parasitic pad capacitances and metal interconnect series resistances of the microwave HBTs. They are discussed in full details in chapter 10. A large-geometry test structure (#4) is also included in order to compare the gain (and other characteristics) of large and small transistors. It can also be used to assess the functionality of the HBT wafer after the collector metal is deposited. If
the transistor characteristics can not be measured on this device after level 4 of the small-geometry fabrication process, the fabrication shall be terminated. The large-TLM structures (#10) and devices from the old mask (#6) have been discussed earlier. Finally, as mentioned in subsection 8.2.2, the small-TLM patterns (#8) on the current mask were proved incompatible with the existing fabrication process; some modifications to the mask is necessary before one can take advantage of these patterns.

![Gummel plot and high frequency characteristics of a Darlington pair with 1-finger (8x10) μm² input and 4-finger (8x10) μm² output transistors fabricated on an InGaP/GaAs DHBT wafer (#EPI3514A).](figure.png)
CHAPTER 9. DC CHARACTERISATION OF HBTs

In this chapter, the DC characteristics of HBTs are studied with the main emphasis on InGaP/GaAs DHBTs. Issues related to the room temperature DC characterisation of HBTs are discussed in section 9.1. In particular, dependence of the offset voltage of HBTs on device structural and geometrical parameters as well as base current is investigated. Then, after explaining the reason behind a high current phenomenon in HBTs, known as “kink effect”, its occurrence is utilised in a novel technique to extract the collector and base series resistance of the device using only DC measurement results. Also the interrelation between the forward and reverse Gummel plots of HBTs is studied in great depth using both qualitative and quantitative approaches.

Section 9.2 deals with the temperatures dependence of HBT DC parameters, including leakage currents, turn-on voltage, breakdown voltage, ideality factors, and current gain. The measured collector currents and current gains at different temperatures will be compared with the results obtained from the numerical simulation program developed in the present work to demonstrate its capability to predict device behaviour in a very wide temperature range.

A number of InGaP/GaAs single- and double-HBTs will be studied in this chapter, and it will be shown that considering almost all aspects of DC characteristics, DHBTs are preferable choices over SHBTs. Since the layer structures of all the single- and double-HBTs studied in this chapter are similar with only slight variations in base and collector doping concentrations, a typical structure from each device is specified in Table 9.1 to avoid duplicating information. These typical structures were grown throughout the duration of this project as control wafers, and therefore, a number of them were available for this study. These wafers only differ in material quality (noticeably base quality), base doping (in the range 2-5×10^{19} cm^{-3}), and collector doping (1-7×10^{16} cm^{-3}). Wherever more precise values of the above doping densities are of particular interest, they would be mentioned explicitly.

Table 9.1- Typical layer structures for InGaP/GaAs single- and double-HBTs used throughout this chapter. Nominal or estimated values of the doping and thickness of various layers will be mentioned for each individual device, wherever necessary.

<table>
<thead>
<tr>
<th>Layer</th>
<th>Single HBT</th>
<th>Double HBT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cap layer</td>
<td>2800Å GaAs, n = (4-5)×10^{18} cm^{-3}</td>
<td>200Å InGaP, n = 2×10^{18} cm^{-3}</td>
</tr>
<tr>
<td>Cap layer</td>
<td>200Å InGaP, n = 2×10^{18} cm^{-3}</td>
<td></td>
</tr>
<tr>
<td>Emitter</td>
<td>1000Å InGaP, n = (3-5)×10^{17} cm^{-3}</td>
<td></td>
</tr>
<tr>
<td>Base</td>
<td>1000Å GaAs, p = (2-5)×10^{19} cm^{-3}</td>
<td></td>
</tr>
<tr>
<td>Spacer</td>
<td>—</td>
<td>200Å GaAs, n = 1×10^{16} cm^{-3}</td>
</tr>
<tr>
<td>Collector</td>
<td>5000Å GaAs, n = (1-3)×10^{16} cm^{-3}</td>
<td>(4800)Å InGaP, n = (2-7)×10^{16} cm^{-3}</td>
</tr>
<tr>
<td>Sub-collector</td>
<td>(5000-7000)Å GaAs, n = 4×10^{18} cm^{-3}</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 9  DC Characterisation of HBTs

9.1. Room Temperature DC Characteristics of HBTs

9.1.1. Output Characteristics

The common-emitter output characteristics of a bipolar transistor is a plot of $I_C$ versus $V_{CE}$ for different constant values of $I_B$. Important parameters of a bipolar transistor, such as offset voltage ($V_{CE,offset}$), saturation voltage ($V_{CE,sat}$), Early voltage ($V_A$), DC current gain, open-base breakdown voltage ($B V_{CEO}$), and open-base leakage current ($I_{CEO}$), can be extracted from the output characteristics. Figure 9.1 compares the output characteristics of large-geometry (100 μm diameter) InGaP/GaAs single- and double-HBTs as specified in Table 9.1. Since the two collector layers have almost similar doping ($\sim 2 \times 10^{16}$ cm$^{-3}$) and thicknesses (5000Å), the 50% larger breakdown voltage in the case of DHBT is purely due to the wide bandgap material in the collector of this device. The symmetry between the B-E and B-C heterojunctions of the DHBT has led to a significantly smaller offset voltage in this device (see subsection 9.1.3 for the formulation of offset voltage). However, the saturation voltage (or the knee voltage), which is the C-E voltage required to bring the transistor out of the saturation and into the forward active mode with almost constant collector current, is smaller in the case of SHBT. The large saturation voltage of DHBTs, which is due to the carrier blocking effect of the triangular potential barrier at the B-C heterojunction, is one of the main limitations in the use of these devices for low power applications. More careful designs of the B-C heterojunction, similar to those discussed in chapter 11, may fully or partly overcome this problem. The same carrier blocking effect is mainly responsible for the lower current gain of the DHBT.

![Figure 9.1- Comparison between the output characteristics of InGaP/GaAs single- and double-HBTs specified in Table 9.1. Both devices have a circular emitter mesa with 100 μm diameter.](image-url)
9.1.2. Gummel Plots

The Gummel plot is, without doubt, the most important DC characteristics of a bipolar transistor. It is a logarithmic plot of base and collector currents as a function of \( V_{BE} \), with \( V_{BC} \) set constant usually at 0V, and presents valuable information as to the B-E junction, ideality factors, leakage currents, gain variation with collector current, turn-on voltage, and emitter series resistance of the device. Figure 9.2 compares the Gummel plots of the single- and double-HBTs of Fig. 9.1 measured at room temperature and room light. Both devices have identical collector ideality factors of close to unity, demonstrating that the collector current is mainly limited by the carrier diffusion across the base rather than the thermionic-field emission across the B-E heterojunction. This means that the conduction band spike of the abrupt InGaP/GaAs B-E heterojunction is negligibly small. Also the fact that the collector current turn-on voltage for both devices is the same (1.03 V at \( I_C = 0.1 \) mA) means that the B-E junction of the two devices is essentially identical. The larger base current in the case of the DHBT leads to smaller current gain in this device, as mentioned earlier. The fact that the base ideality factor of the DHBT is closer to unity means that the extra base current in this device is mainly due to base bulk recombination or extrinsic base surface recombination, both having ideality factors close to unity. The potential spike at the B-C heterojunction blocks the efficient collection of electrons. Consequently, more electrons pile-up inside the base, and also they may be encouraged to move laterally towards the extrinsic base [42]. This increases both base bulk and extrinsic base surface recombination current components. The smaller base ideality factor of the DHBT also results in less variation of current gain with collector current (Fig. 9.3) and temperature (see section 9.2). Finally, the collector leakage current of the DHBT is almost an order of magnitude smaller than that of the SHBT. As also explained in subsection 3.2.1, the hole blocking effect of the B-C valence band discontinuity and the smaller intrinsic carrier concentration for the wide bandgap collector are responsible for the smaller collector leakage currents in the DHBT. Although it appears from Fig. 9.2 that the base leakage current is larger in the DHBT, but in fact the current reversal has happened in the base current of the SHBT, resulting in a current flow which is equal in magnitude but opposite in sign to that of the collector leakage current. (The dominant leakage current component in this case is the B-C leakage current flowing into the collector and out of the base.) In the case of DHBT, this base current reversal happens at lower B-E voltages, when the B-E leakage current becomes equal to the B-C leakage current.

9.1.3. Variation of Offset Voltage with \( I_B \)

The offset voltage, \( V_{CE, offset} \), is defined in the common-emitter configuration as the C-E voltage at which the collector current is zero when the B-E junction is forward-biased. Under this condition, it is straightforward to prove that the HBT is in saturation mode of operation, where both junctions are forward-biased. The offset voltage is of great interest for both digital and analogue applications. This voltage increases the saturation voltage of the transistor, which in turn increases the power consumption in saturation logic circuits. Additionally, a
large offset voltage limits the lower bound for collector voltage swing in amplifier applications, reducing both the output power and power-added efficiency of power amplifiers [100]. Unfortunately, $V_{CE,\text{offset}}$ is larger in HBTs (especially for AlGaAs/GaAs and InAlAs/InGaAs material system) than in Si BJTs (see Fig. 9.4). The origin of the offset voltage in HBTs and its variation with the base current is thoroughly studied by Fresina et al. [407] and Bovolon et al. [408]. In this section, a combination of the analyses used in the latter two works will be employed to explain the behaviour of offset voltage in InGaP/GaAs single- and double-HBTs.

Figure 9.2- Gummel plots of the InGaP/GaAs single- and double-HBTs of Fig. 9.1.

Figure 9.3- Variation of current gain with collector current for the InGaP/GaAs single- and double HBTs of Fig. 9.1.

Figure 9.4- Variation of offset voltage with $I_B$ for InGaP/GaAs single- and double-HBTs, as well as a large Si homojunction transistor (model BC182).

Figure 9.5- A DC Ebers-Moll model for HBTs.

Figure 9.5 shows a simple Ebers-Moll DC model for HBTs [405]. It simply includes two...
diodes representing the B-E and B-C junctions, one current source for modelling transistor action, and three resistive elements to model the series resistances of base, collector and emitter. $I_{BE}$ and $I_{BC}$ are exponentially dependent on the B-E and B-C internal biases, $V_{BE}'$ and $V_{BC}'$, respectively:

$$I_{BE} = I_{BEs} \cdot \exp \left( \frac{V_{BE}'}{\eta_{BE} V_T} \right)$$  \hspace{1cm} (9.1)$$

$$I_{BC} = I_{BCs} \cdot \exp \left( \frac{V_{BC}'}{\eta_{BC} V_T} \right)$$  \hspace{1cm} (9.2)

Here $\eta_{BE}$ ($\eta_{BC}$) is the ideality factor for the B-E (B-C) diode, and $I_{BEs}$ ($I_{BCs}$) is its corresponding saturation current. The current source is usually written in its original form as $I_{CE} = \beta_F I_{BE} - \beta_R I_{BC}$ [405], where $\beta_F$ ($\beta_R$) is the device current gain in the forward (reverse) active mode. However, since $\beta_F$ and $\beta_R$ are both bias dependent, in the present work a more general form is used to express the current source $I_{CE}$:

$$I_{CE} = I_{CC} - I_{EE} = I_{CCs} \cdot \exp \left( \frac{V_{BE}'}{\eta_{CC} V_T} \right) - I_{EEs} \cdot \exp \left( \frac{V_{BC}'}{\eta_{EE} V_T} \right)$$  \hspace{1cm} (9.3)

The internal voltages in Eqs. (9.1)-(9.3) can be expressed as:

$$V_{BE}' = V_{BE} - R_E I_E - R_B I_B$$  \hspace{1cm} (9.4)$$

$$V_{BC}' = V_{BC} + R_C I_C - R_B I_B$$  \hspace{1cm} (9.5)

The advantage of the Ebers-Moll model of Fig. 9.5 is that all the ideality factors and saturation currents can be extracted directly from measurements of the forward and reverse Gummel plots (see subsection 9.1.5). Under forward active mode of operation: $V_{BC}' < 0 < V_{BE}'$, and therefore $I_B = I_{BE}$ and $I_C = I_{CC}$. Consequently, $\eta_{BE}$, $\eta_{CC}$, $I_{BEs}$, and $I_{CCs}$ can be extracted from measurements of the base and collector currents in the forward active mode. Similarly, under reverse active mode: $V_{BE}' < 0 < V_{BC}'$, $I_B = I_{BC}$ and $I_E = I_{EE}$. Therefore, $\eta_{BC}$, $\eta_{EE}$, $I_{BCs}$, and $I_{EEs}$ can be extracted from the measured base and emitter currents in the reverse Gummel plot.

Under the condition of $I_C = 0$, one can write $I_{BC} = I_{CC} - I_{EE}$, $I_B = I_E$, and

$$V_{CE, offset} = V_{CE}' + R_E I_E = V_{BE}' - V_{BC}' + R_E I_B$$  \hspace{1cm} (9.6)

In the following, we derive an expression for $V_{CE, offset}$ which is only dependent on measurable parameters, rather than the internal voltages in (9.6). The forward and reverse Gummel plots of InGaP/GaAs single- and double-HBTs with B-E area of (10×20) $\mu m^2$ are shown in Fig. (9.6). It can be seen from this figure that $I_{BC} = I_B$ (reverse) is always significantly larger than $I_{EE} = I_E$ (reverse), unless the device is designed in such a way that it operates symmetrically in both C-UP and E-UP configurations. Therefore, one can write:
Figure 9.6- Forward and reverse Gummel plots of InGaP/GaAs (a) single- and (b) double-HBTs, together with the extracted parameters required in Eq. (9.9) for modelling the offset voltage.

\[ I_{BC} \approx I_{CC} \Rightarrow I_{BC} \cdot \exp\left(\frac{V_{B'C'}}{\eta_{BC} V_T}\right) \approx I_{CC} \cdot \exp\left(\frac{V_{B'E'}}{\eta_{CC} V_T}\right) \]  

(9.7)

Taking logarithm from both sides of (9.7), solving for \( V_{B'C'} \), and replacing into (9.6) yields:

\[ V_{CE,\text{offset}} = \eta_{CC} V_T \cdot \ln\left(\frac{I_{BC}^{\eta_{BC}}}{I_{CC}^{\eta_{CC}}}\right) - \left(1 - \frac{\eta_{CC}}{\eta_{BC}}\right) V_{B'C'} + R_E I_B \]

Then writing \( V_{B'C'} \) in terms of \( I_{BC} \) from (9.2) gives:

\[ V_{CE,\text{offset}} = V_T \cdot \ln\left(\frac{(I_{BC})^{\eta_{BC}}}{(I_{CC})^{\eta_{CC}}}\right) - (\eta_{BC} - \eta_{CC}) V_T \cdot \ln(I_{BC}) + R_E I_B \]  

(9.8)

As a final approximation, we observe from Fig. 9.6 that \( I_{BC} \gg I_{BE} \), even under the condition of \( I_C = 0 \), when \( V_{B'E'} \) is larger than \( V_{B'C'} \) by the value of \( V_{C'E',\text{offset}} \). Therefore, \( I_B = I_{BC} + I_{BE} = I_{BC} \), and one can write:

\[ V_{CE,\text{offset}} = V_T \cdot \ln\left(\frac{(I_{BC})^{\eta_{BC}}}{(I_{CC})^{\eta_{CC}}}\right) - (\eta_{BC} - \eta_{CC}) V_T \cdot \ln(I_B) + R_E I_B \]  

(9.9)

The latter assumption of \( I_B = I_{BC} \) should be used with care, especially for low-current offset voltages larger than \( \sim 0.2 \) V. Eq. (9.9), which is identical to Eq. (5) in [408], can explain the behaviour of offset voltage observed in Fig. 9.4. The first term in (9.9) is a bias independent component which defines the relative magnitude of the offset voltage in various bipolar device structures. Usually \( \eta_{BC} > \eta_{CC} \), and the second term on the RHS of (9.9) would be responsible for the slow reduction of \( V_{CE,\text{offset}} \) at low base current levels. At high current levels, the last term in (9.9) becomes dominant and the offset voltage starts to increase sharply with \( I_B \). In fact, Eq. (9.9) (or its more accurate form of (9.8)) is the basis for the well-known
“open-collector” method for the measurement of emitter series resistance [405].

In order to demonstrate the accuracy of Eq. (9.9) in explaining the variation of offset voltage with $I_B$, the parameters $I_{BCS}$, $I_{CCS}$, $\eta_{BC}$, $\eta_{CC}$, and $R_E$ are extracted for the InGaP/GaAs single- and double-HBTs in Fig. 9.6. These parameters are replaced in Eq. (9.9) to fit the variation of $V_{CE,\text{offset}}$ with $I_B$ for the above devices, as shown in Fig. 9.7. It can be observed that the agreement between Eq. (9.9) and the measured data is very good for $I_B > 5 \times 10^{-6}$ A. The discrepancy between (9.9) and the measured data for lower values of $I_B$ can be explained by a combination of the following reasons. First, the ideality factor of the B-C diode, $\eta_{BC}$, changes at lower currents. The non-exponential leakage currents may also become comparable with the exponential currents considered in the Ebers-Moll model of Fig. 9.5. Finally, the assumption $I_{BC} \gg I_{BE}$ becomes invalid at lower currents. For the SHBT in Fig. 9.6(a), $\eta_{BC} = 2$ means that $I_{BC}$ is mainly dominated by the recombination current inside the B-C SCR, which is primarily extended inside the GaAs collector. The large difference between $\eta_{BC}$ and $\eta_{CC}$ in this case results in a significant increase of $V_{CE,\text{offset}}$ at low base currents for SHBTs. On the other hand, the wide bandgap of InGaP collector and the large $\Delta E_Y$ at the B-C heterojunction of the DHBT significantly reduce both the recombination current inside the collector and the hole injection current from the base to collector, thereby reducing both $\eta_{BC}$ and $I_{BCS}$ in the case of the DHBT. Consequently, both the first and second terms on the RHS of (9.9) are smaller in the case of the DHBT. This leads to a low and bias-independent offset voltage for DHBTs. Earlier studies of the offset voltage (e.g., [98]) were neglecting the difference in ideality factors of $I_{BC}$ and $I_{CC}$, therefore the $\ln(I_B)$ term in Eq. (9.9) was absent in those studies.

The large offset voltages observed in SHBTs is mainly due to the asymmetries between the B-E and B-C junctions as explained by:

- Existence of a conduction band spike at the abrupt B-E heterojunction of SHBTs increases the turn-on voltage of this junction as compared to the B-C junction. This results in an $I_{CCS}$ much smaller than $I_{BCS}$, thereby increasing the first term of (9.9). The B-E turn-on voltage can be reduced by using a homojunction, as in Si BJTs (see Fig. 9.4) or HEBTs (see subsection 3.4.2). Also grading the B-E heterojunction, using a narrow bandgap spacer and/or a doping spike between the base and emitter [57-59] may reduce the asymmetry between the B-E and B-C turn-on voltages.

- Larger SCR recombination inside the GaAs collector as compared to the wide bandgap emitter is also responsible for further increasing $I_{BCS}$ relative to $I_{CCS}$. This also causes $\eta_{BC}$ to be close to 2, further increasing the first term on the RHS of (9.9). As mentioned earlier, this problem may be solved by using a wide bandgap collector.

- Larger B-C junction area compared to the B-E area is a final contributor to the large offset voltage of HBTs. Any technique resulting in a reduced B-C active area, such as a collector-up configuration [98,127], may lead to a significant reduction of the offset voltage.
According to Eq. (9.9), one expects the offset voltage to increase with temperature, mainly due to the increase of $V_T$ and $R_E$ at higher temperatures [408] (also see subsection 9.2.3). As a final comment, Eq. (9.9) predicts a minimum of offset voltage at $I_B = V_T(\eta_{BC} - \eta_{CC})/R_E$. Therefore, one expects to observe a broad minimum in low current levels in the case of the DHBT. Also the minimum shifts to lower current levels when the device size shrinks. This is due to the fact that smaller emitter dimensions lead to larger emitter series resistances, which is also evidenced by the comparison between the $V_{CE,offset}$ versus $I_B$ plots in Fig. 9.7 for SHBTs having B-E areas of (6×10) and (10×20) μm².

### 9.1.4. Extraction of the Series Resistances

Measurement of the emitter series resistance using the open-collector technique was briefly discussed in subsection 9.1.3. In this section, the occurrence of a high current phenomena is utilised to extract the collector and base series resistances of HBTs.

Bipolar transistors are normally biased in forward active mode, with the B-E junction forward-biased and the B-C junction reverse-biased. In order to achieve the best high frequency performance, these devices are operated at large current densities. However, large current densities in HBTs may cause the B-C junction to be forward-biased, especially in the cases where parasitic resistances of the transistor are large and the transistor is operated with small negative voltage applied to the B-C junction. Figure 9.8 shows the measured collector and base currents as a function of $V_{BE}$ at various B-C voltages for a 12×12 μm² InGaP/GaAs SHBT fabricated using a triple mesa process (i.e., without the implant isolation technique). In
Fig. 9.8(b) the base current shows a rapid rise ("kink") at high current levels, with the onset of rapid rise depending on the applied B-C voltage. By increasing the reverse bias across the B-C junction, the onset of the kink is shifted towards higher current densities or higher B-E forward biases. The kink in the base current is coincident with a decrease in the collector current, suggesting that an additional current component is flowing between the base and collector. The apparent current gain of the device at $V_{BE} = 2.0\ \text{V}$ varies from $\beta = 2$ when $V_{CB} = 0.0\ \text{V}$, to $\beta = 29$ for $V_{CB} = 1.3\ \text{V}$.

![Graphs showing variation of collector and base currents with $V_{BE}$ at different B-C biases](image)

**Figure 9.8** - Variation of (a) collector and (b) base currents with $V_{BE}$ at different B-C biases, for an InGaP/GaAs SHBT with $12\times12\ \mu\text{m}^2$ emitter dimension.

This effect has first been observed by Tiwari [409] in his study of large current densities in DHBTs. Tiwari has attributed this effect to the formation of an excess electron barrier at the B-C heterojunction leading to an increase in minority carrier charge storage inside the base and a decrease in the current gain. However, Tiwari's theory only explains the occurrence of this effect for double HBTs. Fricke _et al._ [410] have also observed the same effect for single HBTs, this time at elevated temperatures ($T = 200-300^\circ\text{C}$), but there was no detailed explanation of the effect. Liou _et al._ [411] have used a one-dimensional numerical device simulator to model this effect. They have clearly attributed the observed kink to the forward-biased B-C junction. However, they did not compare their calculated results with experimental data. Finally, Peter _et al._ [412] believe that the observed kink effect in their SiGe HBTs is purely due to the voltage drop across $R_E$ and a stronger saturation of collector current. In the present work, the observed kink in the base current of single HBTs is explained by a forward-biased B-C junction (as in [411]) using a very simple theory, and the results of the theoretical calculations are compared against measured data. Although the occurrence of this effect is detrimental to the current gain of the device, it is later shown that one can utilise it to extract (using only the DC measured results) the parasitic series resistances of the HBT.

Referring back to the DC Ebers-Moll model of Fig. 9.5 together with Eqs. (9.1)-(9.5), and

207
assuming $I_{BC} \gg I_{EE}$, one can write $I_B = I_{BE} + I_{BC}$ and $I_C = I_{CC} - I_{BC}$. In using Eqs. (9.1)-(9.5), it is important to note that they do not include any low-bias leakage current components, and therefore, they are valid only for moderate to high B-E bias conditions. Also, due to the fact that the base of the HBTs is usually more highly doped than the emitter and back injection of holes from base to emitter is blocked by the presence of the valence band discontinuity at the B-E heterojunction, high-level injection does not occur inside the base nor inside the emitter [334]. Therefore, the same ideality factors $\eta_{CC}$ and $\eta_{BE}$ obtained from the moderate bias region can be used for high bias conditions provided that the real junction potential $V_{BE}'$ is used instead of $V_{BE}$ in the exponential current formulation. However, for a highly forward-biased B-C junction in SHBTs, one may need to use a different $\eta_{BC}$, due to the high-level injection inside the collector region.

Equations (9.4) and (9.5) show that for high current conditions the internal junction biases are different from the externally applied ones. While Eq. (9.4) is simply responsible for the saturation of collector and base currents in the normal Gummel plots, the situation for Eq. (9.5) is slightly more complicated. When $R_C$ is large, the actual B-C junction may become forward-biased at high collector currents even when the externally applied B-C bias is negative. This effect is responsible for the observed kink in the base current at high bias conditions, as shown in Fig. 9.8(b). This extra component of current is in the same direction as the conventional base current and causes a sharp rise into the base current, but is flowing opposite to the conventional collector current, causing $I_C$ to reduce. As $I_B$ increases, the term $R_B I_B$ eventually becomes significant, limiting the B-C forward bias. The latter is the reason for the second saturation observed in the base current behaviour.

Although there exist some DC methods to measure the collector series resistance $R_C$ of bipolar transistors [405,413-414], none of them are accurate enough. For example, in the open-emitter method [405], where a plot of $V_{CE}$ vs. $I_B$ with $I_E = 0$ is used to extract $R_C$, the intrinsic parts of the base and collector regions will remain inactive due to the open-emitter condition. This leads to a different measured $R_C$ under this condition than in the case of normal active mode.

In the case of $R_B$ measurement, several methods are suggested in the literature, among them those in [413-416] are DC techniques. The method of Ning and Tang [415] is mainly proposed for Si BJTs and also relies on the variation of current gain in the high bias region of the Gummel plot. Therefore, it is not suitable for HBTs with an almost flat current gain at high collector current, e.g., in InGaP/GaAs HBTs with a heavily doped base. When the method of Ning and Tang was applied to several HBTs, including the one with the characteristics shown in Fig. 9.8, all of the plots showed some random behaviour with no noticeable trend. The technique in [416] uses the I-V characteristics of the B-E junction in the reverse bias breakdown regime to extract values of $R_B$ for the HBTs. Although this method is very easy to implement, it involves subtraction of some other contributing resistances which
should be calculated based on analytical assumptions for p-n junctions. It also needs prior knowledge of some physical parameters of the junction as thermal impedance, fractional change in breakdown voltage per degree Celsius and avalanche region length, which in turn may be estimated with large amounts of error. The methods in [413] and [414] are essentially iterative optimisation techniques, which are not easy to implement. Moreover, they assume that $R_B$ and $R_C$ values are identical in the forward and reverse active modes of operation.

The most reliable method to extract $R_B$ is based on s-parameters [417] (details are given in chapter 10), which involves high frequency measurements. Hence, finding a DC method to extract $R_B$ and $R_C$ using only the I-V characteristics of the transistor is of considerable interest.

Before turning to the subject of $R_B$ and $R_C$ extraction, let us first briefly explain the extraction of $R_E$ from the Gummel plot. Basically, when the kink is not observed in the base current (e.g., in the case of $V_{CB} = 1.3$ V in Fig. 9.8), $I_{BC}$ is a negligible part of the collector current which can be ignored as compared to $I_{CC}$. For a known $I_C$, $V_{BE'}$ can be found from a straight line fit to the moderate current range in the logarithmic plot of $I_C$ vs. $V_{BE}$. Then the difference between $V_{BE}$ and $V_{BE'}$ is plotted against $I_E$:

$$\Delta V_{BE} = V_{BE} - V_{BE'} = R_E I_E + R_B I_B = I_E \left( R_E + \frac{R_B}{\beta_F + 1} \right)$$

(9.10)

This plot should be a straight line, and if the forward current gain of the transistor is high enough, the slope of the line will give $R_E$. This method is applied to the collector current of the transistor with $V_{CB} = 1.3$ V and the results are shown in Fig. 9.9. The inset of Fig. 9.9 shows the graph of $\Delta V_{BE}$ vs. $I_E$ that is quite straight and its slope identifies an $R_E$ of 9.2 $\Omega$ which is in good agreement with the value of 8.9 $\Omega$ obtained from the open-collector method on the same device.

Having found the value of $R_E$, we now move on to the procedure which determines $R_C$. For $V_{CB} \geq 1.2$ V, where no base current kink is observed in the range of $V_{BE}$ shown in Fig. 9.8(b), firstly one can extract $I_{BE}$ and $\eta_{BE}$ from the fitted straight line to the moderate current levels of $I_B$ vs. $V_{BE}$. Secondly, if one subtracts this $I_B (= I_{BE})$ from any $I_B$'s in which the kink is observed ($V_{CB} < 1.2$ V), the remainder will represent $I_{BC}$:

$$\Delta I_B(V_{CB}) = I_B(V_{CB} < 1.2V) - I_B(V_{CB} = 1.3V) = I_{BC} \exp\left(\frac{-V_{CB} + R_C I_C - R_B I_B}{\eta_{BC} V_T}\right)$$

(9.11)

Now if $\Delta I_B$'s for different $V_{CB}$'s are plotted logarithmically against $I_C$, the graphs will show exponential behaviours (straight lines) in the medium $I_C$ range (0.1-1 mA for the present device). This is because $R_B I_B$ is much smaller than $R_C I_C$ in this range and may be ignored. According to Eq. (9.11), the straight lines fitted to the log($\Delta I_B$) vs. $I_C$ curves for different $V_{CB}$'s should all be parallel and equidistant with slopes of $R_C / (\eta_{BC} V_T)$ and vertical axis
intercepts of $I_0(V_{CB}) = I_{BCS} \cdot \exp(-V_{CB}/\eta_{BC}V_T)$. In Fig. 9.10 the graphs of $\Delta I_B$ vs. $I_C$ for $V_{CB} = 0.0-0.4$ V are plotted together with the straight line fits (dashed lines). One can observe that the straight line fits are quite parallel and equidistant, confirming the validity of the present theory. The slopes of these lines are: $R_C/(\eta_{BC}V_T) = 361 \pm 2.7\%$. In order to find $I_{BCS}$ and $\eta_{BC}$ (and therefore $R_C$), the values of the intercept points, $I_0$'s, are plotted against $V_{CB}$ in Fig. 9.11. The straight line fitted to the plot of $\log(I_0)$ vs. $V_{CB}$ yields:

$I_{BCS} = 4.1 \times 10^{-7}$ A and $\eta_{BC} = 3.0$

Now using $\eta_{BC} = 3.0$, the magnitude of $R_C$ can be easily found as $R_C = 27.9 \pm 2.7\% \Omega$.

The deviations of the plots of $\Delta I_B$ vs. $I_C$ from a straight line in Fig. 9.10 at high current levels are due to the term $R_B I_B$ in Eq. (9.11). In fact, these deviations will allow us to find $R_B$. For a certain value of $\Delta I_B$, one can find two values of collector currents: one from the measured results ($I_C$) and another from the straight line fit ($I'_C$). $I_C$ and $I'_C$ satisfy the following relation:

$$\Delta I_B = I_0 \cdot \exp \left( \frac{R_CI_C - R_B I_B}{\eta_{BC}V_T} \right) = I_0 \cdot \exp \left( \frac{R_CI'_C}{\eta_{BC}V_T} \right)$$

(9.12)

Therefore

$$R_B = R_C \Delta I_C / I_B \quad \text{with} \quad \Delta I_C = I_C - I'_C$$

(9.13)

Plots of $\Delta I_C$ vs. $I_B$ for different $V_{CB}$'s, corresponding to different curves in Fig. 9.10, are shown in Fig. 9.12. It can be seen that all the graphs are overlaying, showing the accuracy of the
results. The slopes of the lines are $0.53 \pm 1\%$ which give $R_B = 14.6 \pm 2.7\% \Omega$.

![Figure 9.11- The y-intercept ($I_0$) extracted from the exponential fits in Fig. 9.10 plotted against $V_{CB}$. $I_0$ vs. $V_{BC}$ shows again an exponential behaviour as expected from Eq. (9.11).](image)

![Figure 9.12- Plots of $\Delta I_C$ (as defined in Eq. (9.13)) vs. $I_B$ for different $V_{CB}$'s in the range 0.0-0.4 V. The plots are all overlaying, fitting to a line with a slope of 0.53.](image)

The above method for characterisation of parasitic resistances relies on the existence of the kink in the base current of a Gummel plot. However, when the kink does not appear in the normal Gummel plot, this method is still applicable. In the latter case, one can simply apply small forward biases to the B-C junction to observe the kink. In fact, this method has been applied to a few samples with or without the kink observed in the normal Gummel plots and the results are summarised in Table 9.2. Alternatively, one can add external collector and base resistances to the transistor and force the device to show the kink behaviour, then find the total $R_C$ and $R_B$ and subtract the external ones.

**Table 9.2- Extracted parasitic resistances for InGaP/GaAs SHBTs having different geometries.**

<table>
<thead>
<tr>
<th>Emitter dimension</th>
<th>$R_E$ ($\Omega$)</th>
<th>$R_C$ ($\Omega$)</th>
<th>$R_B$ ($\Omega$)</th>
<th>Device and layout</th>
</tr>
</thead>
<tbody>
<tr>
<td>12×12 $\mu m^2$</td>
<td>9.2</td>
<td>27.9 ± 2.7%</td>
<td>14.6 ± 2.7%</td>
<td>only one base and collector contact</td>
</tr>
<tr>
<td>10×16 $\mu m^2$</td>
<td>8.5</td>
<td>16.5 ± 4.9%</td>
<td>13.2 ± 6.1%</td>
<td>one C and double B contacts</td>
</tr>
<tr>
<td>16×20 $\mu m^2$</td>
<td>5.1</td>
<td>10.1 ± 6.4%</td>
<td>13.1 ± 15%</td>
<td>one C and double B contacts</td>
</tr>
<tr>
<td>100 $\mu m$ diameter</td>
<td>2.2</td>
<td>5.2 ± 9.7%</td>
<td>12.6 ± 13%</td>
<td>round mesa transistor</td>
</tr>
<tr>
<td>140 $\mu m$ diameter</td>
<td>1.9</td>
<td>6.1 ± 10%</td>
<td>14.3 ± 13%</td>
<td>round mesa transistor</td>
</tr>
</tbody>
</table>

The results of Table 9.2 shows that as $R_C$ gets smaller and dominated mostly by the bulk resistance rather than the ohmic contact resistance, the accuracy of the present method becomes questionable. This is because collector bulk series resistance is dependent on the B-C junction bias; for highly forward-biased B-C junctions the collector current is diverted away from the intrinsic device area. In fact the measured data always shows the largest $R_C$ for the most negative bias applied to the B-C junction due to the longest effective path of the collector current inside the $n^+$-sub-collector in this case. Therefore, the same method can be
applied to find the dependence of $R_C$ on $V_{CB}$.

The extracted series resistances are verified by the author in [418] by plugging all the extracted parameters back into the DC Ebers-Moll model of Fig. 9.5, and it was shown that the simple Ebers-Moll model of Fig. 9.5 recreates the measured data quite accurately. Here the measured data is reproduced by numerical simulation of the device. The results are shown in Fig. 9.13, where a very good agreement between the numerically calculated and measured data are observed for three different values of $V_{CB}$ (0.0, 0.8, and 1.2 V). The procedure for fitting the numerical data to the measured results was as follows. First, parameters of the HBT were fine-adjusted to create a good fit to the measured $I_C$ for $V_{CB} = 1.2$ V. These parameters include: $\Delta E_C(\text{InGaP/GaAs}) = 0.05$ eV, $R_E = 9.0 \, \Omega$, $N_{DE} = 3 \times 10^{17} \, \text{cm}^{-3}$, and $N_{AB} = 2 \times 10^{19} \, \text{cm}^{-3}$. Additionally, all the recombination rates were increased by a factor of 3.3 to recreate the measured $I_B$ ($V_{CB} = 1.2$ V). (This is mainly due to the poor quality of the base material in this device.) The above parameters were then kept fixed and $R_B$ and $R_C$ were slightly adjusted to fit the experimental data for $V_{CB} = 0.0$ V. The values of $R_B$ and $R_C$ used for simulation are 14 and 28.5 $\Omega$, respectively. Finally, the same set of parameters were used for $V_{CB} = 0.8$ V condition. The very good agreement between the measured and simulated results for different values of $V_{CB}$ once again proves the suitability of the above method to extract parasitic series resistances of the HBTs. The quality of the fits at current levels lower than 0.1 mA is affected by the large leakage currents of the device.

As a final comment, it should be pointed out that devices which show no sharp rise in their base current at room temperature, may in fact display a kink behaviour at elevated temperatures [410,418]. This is not surprising, since the reduction of electron mobility at higher temperatures gives rise to a larger collector series resistance, and hence a lower onset of the kink effect.

### 9.1.5. Forward and Reverse Gummel Plots

Design of bipolar transistors in which the roles of emitter and collector can be interchanged simply by changing the biasing conditions is quite desirable from the circuit designers’ point
of view [20,125]. Various authors have studied the Gummel plots of HBTs in forward and reverse active mode (i.e., forward and reverse Gummel plots), both theoretically [126,165] and experimentally [44,82]. Grinberg and Luryi [165] have used an analytical one-dimensional model to prove that the two output currents (collector current in the forward active mode and emitter current in the reverse active mode) must be identical, irrespective of the height of the conduction band spike in the B-E heterojunction. However, their analysis ignored the contribution of tunnelling across the B-E heterojunction and did not consider the two-dimensional effects. Liu et al. [44,82] have measured the forward and reverse Gummel plots of various types of HBTs and have shown that if \( \Delta E_c \) at the B-E heterojunction is small, the output current depends solely on the doping profile and thickness of the base. Consequently, for a given device, since the base layer is the same in either the forward or reverse active modes, the two output currents overlay each other. Also in this case, the ideality factors of both output currents should be unity, since both of them are dominated by electron diffusion across the base. However, when \( \Delta E_c \) at the B-E heterojunction of single HBTs is large, as in the case of abrupt AlGaAs/GaAs HBTs, the output current in the forward active mode will be dominated by thermionic-emission and tunnelling transport across the B-E heterointerface. Therefore, the ideality factor of the forward output current can be significantly different from unity, while that of the reverse output current for the B-C homojunction is still dominated by diffusion across the base and is very close to unity. Thus, the two output currents will not overlay in this case.

Although the theory of Liu et al. seems to explain the behaviour of forward and reverse output currents in most of the HBTs, it still considers the current transport in HBTs as a purely one-dimensional process. In fact, many of the abrupt AlGaAs/GaAs HBTs and some of the (disordered) InGaP/GaAs HBTs fabricated in our research laboratory show a reverse output current which is strikingly different from the forward current; the reverse current does not even increase exponentially with bias (see Fig. 9.14). Zhang et al. [126] have carried out a 2-D numerical simulation of forward and reverse Gummel plots for graded AlGaAs/GaAs DHBTs including the effects of recombination at the extrinsic base surface. Their output current in the forward and reverse modes are still overlaying, showing that the surface recombination on its own can not explain the odd behaviour observed in Fig. 9.14. In this section, the behaviour of reverse output current in various types of HBTs fabricated in our laboratory is studied and those similar to Fig. 9.14 are explained by the current blocking effect of the B-E heterojunction spike and the three-dimensional nature of the current transport.

Various types of SHBTs, namely InGaP/GaAs, abrupt and graded AlGaAs/GaAs, and InP/InGaAs HBTs, with large geometries (emitter mesa diameter of 100 \( \mu \)m) fabricated using the conventional techniques discussed in chapter 8 are studied here. Figures 9.14-9.16 show typical forward and reverse Gummel plots for three kinds of abrupt HBTs. In all the graphs \( I_B \) is commonly seen to increase significantly in the reverse Gummel plot. This is mainly due to the parasitic junction formed under the extrinsic base region, which causes a direct collection
of injected electrons by the base ohmic contact. Ion implantation of the extrinsic collector region, either to damage this area or to turn it into p-type, may partly or completely cure this unwanted phenomena [126]. Alternatively, selective regrowth techniques [130] or buried dielectric layers under the base electrode [125] may be utilised to eliminate this current injection.

\[ I_{c, \text{Forward}} = 1.12 \]
\[ I_{e, \text{Reverse}} = 1.00 \]
\[ I_{b, \text{Forward}} / I_{b, \text{Reverse}} < 0.01 \]

\[ E_{-3} = 1 \times 10^{-3} \]
\[ E_{-4} = 1 \times 10^{-4} \]
\[ E_{-5} = 1 \times 10^{-5} \]
\[ E_{-6} = 1 \times 10^{-6} \]
\[ E_{-7} = 1 \times 10^{-7} \]
\[ E_{-8} = 1 \times 10^{-8} \]
\[ E_{-9} = 1 \times 10^{-9} \]

0.0 0.4 0.8 1.2 1.6
\[ V_b (V) \]

0.0 0.4 0.8 1.2 1.6
\[ V_b (V) \]

**Figure 9.14** Measured forward/reverse Gummel plots for a typical abrupt AlGaAs/GaAs SHBT.

**Figure 9.15** Measured forward/reverse Gummel plots for a typical (ordered) InGaP/GaAs SHBT.

**Figure 9.16** Measured forward/reverse Gummel plots for a typical InP/InGaAs SHBT.

**Figure 9.17** Simulated forward/reverse Gummel plots for the InP/InGaAs SHBT of Fig. 9.16.

Although the Gummel plots in Figs. 9.14-9.16 are behaving almost similar as far as the increase of the base current is concerned, there is a striking difference between \( I_{E, \text{Reverse}} \) in these figures. When \( \Delta E_C \) between the wide bandgap emitter and narrow bandgap base is small (as in ordered InGaP/GaAs HBTs, see Fig. 9.15), the output current is limited by the transport of carriers across the base. Therefore, both output currents \( I_C(\text{forward}) \) and \( I_E(\text{reverse}) \) have ideality factors very close to unity and they overlay each other in the low to medium current range. Although not shown here, the forward and reverse Gummel plots of our graded
AlGaAs/GaAs SHBTs behave exactly the same way as in Fig. 9.15. The earlier saturation of \( I_E (\text{reverse}) \) as compared to \( I_C (\text{forward}) \) is due to its larger effective emitter resistance. In both the forward and reverse modes of operation an effective resistance of

\[
R_{\text{eff}} = \frac{R_E I_E + R_B I_B}{I_C} = \frac{(\beta + 1) R_E + R_B}{\beta}
\]

(9.14)

is responsible for the saturation of output current versus \( V_{BE} \). Since in the forward active mode \( \beta_F \gg 1 \), the effective resistance would be \(~R_E\). In the reverse mode, where the roles of emitter and collector are interchanged, \( \beta_R \ll 1 \), and a large resistance of \((R'_E + R'_B)/\beta_R\) is effective in saturating the output current.

When \( \Delta E_C \) at the B-E heterointerface is larger, as in the case of InP/InGaAs HBTs, forward output current is limited by tunnelling and thermionic emission of electrons across the B-E heterojunction spike. In this case, the ideality factor of the forward output current is given by Eq. 2.15. Calculations with typical parameters of our HBTs will result in ideality factors in the range (1.05-1.20), depending on the actual material and doping of the base and emitter. However, the transport of carriers across the B-C homojunction in the reverse active mode is still dominated by diffusion across the base, and hence, it has an ideality factor very close to unity. This is consistent with the measured results in Fig. 9.16. The numerically simulated Gummel plots of this InP/InGaAs HBT are also shown in Fig. 9.17, where a good agreement to the experimental data in Fig. 9.16 can be observed.

The first two types of forward and reverse Gummel plots as shown in Figs. 9.15 and 9.16 can well be explained using 1-D models in terms of current transport across the base and across the B-E heterojunction spike. However, the third type of reverse Gummel plot (Fig. 9.14) which is observed in most of our abrupt AlGaAs/GaAs HBTs and some of our (disordered) InGaP/GaAs HBTs has not been addressed and explained in the literature. In this case, the maximum reverse output current hardly approaches \(~0.1\mu A\) and never follows any exponential dependence on the base voltage. In the following, the reason behind this behaviour of the reverse output current will be explained with the aid of two-dimensional effects. As a first guess, one may consider the recombination at the extrinsic base surface as a sole responsible for this effect. However, as mentioned earlier, Zhang et al. [126] have carried out a 2-D numerical modelling of graded AlGaAs/GaAs C-UP DHBTs including the recombination at the extrinsic base surface, and their forward and reverse output currents still overlay.

We can justify the rather low reverse mode output current by the fact that there exists a large \( \Delta E_C \) between the base and emitter, which now acts as the collector in the reverse mode of operation. This conduction band barrier only existing in the intrinsic part of the device, impedes the free collection of electrons and gives them much more tendency to move towards the extrinsic base region (see Fig. 9.18). Finally, the laterally directed electrons recombine in the base bulk or surface, or even they may be collected directly by the base ohmic contact. So,
there are two key issues in the occurrence of a reverse output current as in Fig. 9.14: a very large $\Delta E_C$ at the B-E heterojunction, and a large recombination rate inside the extrinsic base, especially at the surface. The first issue is supported by the fact that abrupt AlGaAs/GaAs and disordered InGaP/GaAs heterojunctions are known to have a large conduction-to-valence band offset ratio (see chapter 5). Indeed all our GaAs based HBTs showing the type of reverse Gummel plots as in Fig. 9.14 had a forward turn-on voltage almost 0.1V larger than those showing reverse Gummel plots of Fig. 9.15 (compare Figs. 9.14 and 9.15). The second argument of large extrinsic base surface recombination was confirmed by a significant emitter size effect in the Fig. 9.14 type devices. Devices on one particular wafer of this kind showed almost 60% reduction in the current gain when the perimeter-to-area ratio was increased from 0.0375 µm$^{-1}$ to 0.45 µm$^{-1}$. Also the much smaller surface recombination velocity of InGaAs as compared to GaAs may be one of the reasons why the output current of this kind of device does not behave similar to Fig. 9.14 in the reverse mode.

![Figure 9.18- Schematic diagram of a mesa HBT structure operating in reverse active mode.](image)

This study of forward and reverse Gummel plots in various types of HBTs is of prime importance in the design of HBTs which are capable of operating in both C-UP and E-UP operation modes. Although the main requirement in the design of C-UP HBTs is to inactivate the extrinsic emitter region (see section 3.2.3), the above study shows that it is also of equal importance to minimise or smooth out any conduction band barrier at the B-C heterojunction for a symmetric operation of the transistor.

9.2. Variation of HBT DC Characteristics with Temperature

In this section, variation of various HBT characteristics with temperature is discussed. Temperature measurements are carried out in two different vacuum equipment setups. For low temperature measurements (77-300K), a cryostat which is cooled by liquid nitrogen is used, while a modified Edwards high vacuum evaporation system equipped with a heater and temperature controller is employed for high temperature measurements (300-700K) [90].

All devices used in this section are large-geometry HBTs (circular emitter with 100µm
Chapter 9 DC Characterisation of HBTs

InGaP/GaAs devices used in this study are the same as those specified in Table 9.1, with base doping level of \( \sim 4 \times 10^{19} \text{ cm}^{-3} \), and collector doping level of \( 3-4 \times 10^{16} \text{ cm}^{-3} \). The InGaP/GaAs DHBTs used for the low- and high-temperature characterisation are from two different wafers with essentially the same layer structures; the only difference between them is in device fabrication. The collectors of the HBTs measured at high temperature were not isolated by mesa etching down to the SI GaAs substrate, while the devices used for the low temperature study were all having a triple-mesa configuration with properly isolated collectors. Consequently, the latter devices were showing lower collector leakage current levels which allowed a direct comparison between the leakage current levels of single- and double-HBTs. The layer structure of the AlGaAs/GaAs HBTs utilised in this section is given in Table 9.3. Although a graded AlGaAs layer is inserted between the emitter and base of this HBT, it has been shown by the author in another study [419] that the B-E heterojunction of this HBT is not properly graded, and that a conduction band spike still exists at this heterojunction.

Table 9.3- Layer specification of a (nominally) graded AlGaAs/GaAs SHBT (wafer no. 3075-3).

<table>
<thead>
<tr>
<th>Layer</th>
<th>Material</th>
<th>Doping (cm(^{-3}))</th>
<th>Thickness ((\AA))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cap Layer</td>
<td>GaAs</td>
<td>(n = 2 \times 10^{18})</td>
<td>3500</td>
</tr>
<tr>
<td>Graded Cap</td>
<td>Al(<em>x)Ga(</em>{1-x})As ((x = 0.3 \rightarrow 0.0))</td>
<td>(n = 2 \times 10^{18})</td>
<td>400</td>
</tr>
<tr>
<td>Emitter</td>
<td>Al(<em>{0.3})Ga(</em>{0.7})As</td>
<td>(n = 3 \times 10^{17})</td>
<td>1500</td>
</tr>
<tr>
<td>Graded Emitter</td>
<td>Al(<em>x)Ga(</em>{1-x})As ((x = 0.0 \rightarrow 0.3))</td>
<td>(n = 3 \times 10^{17})</td>
<td>225</td>
</tr>
<tr>
<td>Base</td>
<td>GaAs</td>
<td>(p = 3 \times 10^{19})</td>
<td>1000</td>
</tr>
<tr>
<td>Collector</td>
<td>GaAs</td>
<td>(n = 5 \times 10^{15})</td>
<td>10000</td>
</tr>
<tr>
<td>Sub-collector</td>
<td>GaAs</td>
<td>(n &gt; 5 \times 10^{18})</td>
<td>7000</td>
</tr>
<tr>
<td>Substrate</td>
<td>S.I. GaAs</td>
<td>undoped</td>
<td>—</td>
</tr>
</tbody>
</table>

9.2.1. Gummel Plot

Figure 9.19 shows the variation of collector current of an InGaP/GaAs DHBT with \(V_{BE}\) at different temperatures ranging from 78K to 295K. A negative shift for the B-E turn-on voltage can be observed as the temperature increases. Also very low leakage currents in the range of few pA are measured for this DHBT. In fact, the real leakage current level is believed to be even less than this range, since the measured values do not vary with temperature. The observed leakage values could be due to the limitations of the measuring instrument (Semiconductor Parameter Analyser), or the system and/or wiring leakage. The very low leakage current levels of the DHBTs is due to the hole blocking effect of the large valence band discontinuity at the B-C heterojunction and the small intrinsic carrier concentration of the wide bandgap InGaP collector, as discussed previously in subsection 3.2.1. This has important consequences. First, the device can be operated at higher temperatures than the maximum temperature for SHBTs (see subsection 9.2.3). Second, it is shown in [96] that double HBTs require much shorter switching times than SHIBTs. For the purpose of comparison, the room temperature Gummel plot of an InGaP/GaAs SHBT with
Chapter 9 DC Characterisation of HBTs

exactly similar geometry and fabrication process to that of the DHBT, and measured in the same cryostat is also shown in Fig. 9.19 (dashed line). The collector leakage current of the SHBT is more than an order of magnitude larger than the (apparent) DHBT leakage current.

The measured \( I_C \) and \( I_B \) of another InGaP/GaAs DHBT are also shown in Fig. 9.20, this time at higher temperature range of (298-623)K. As mentioned earlier, the collectors of these devices are not mesa isolated, and therefore they show higher collector leakage currents as compared to those in Fig. 9.19. As can be seen, the device is still functional at 623K (350°C), and exponential current variations can be observed for both \( I_C \) and \( I_B \) at collector current levels of > 1mA. The current gains of the device calculated from Gummel plots of Fig. 9.20 are plotted versus \( I_C \) in Fig. 9.21. An almost constant current gain for about four orders of magnitude of collector current is observed in Fig. 9.21. Moreover, the current gain shows almost no variation with temperature, especially at higher current levels. This is an extremely important property of InGaP/GaAs DHBTs for the design of linear amplifiers, and for thermal stability of power amplifiers [420] (also see subsection 9.2.2).

Figure 9.19- The measured collector current of a large-geometry (circular emitter; 100 \( \mu \)m diameter) InGaP/GaAs DHBT as a function of \( V_{BE} \) at different temperatures in the range (78-295)K. The collector current of a similar InGaP/GaAs SHBT at \( T = 295K \) is also shown for comparison.

9.2.2. Turn-on Voltage

The turn-on voltages, defined at \( I_C = 1 \) mA and \( V_{CB} = 0 \), of the DHBTs in Figs. 9.19-9.20 are extracted and plotted in Fig. 9.22 as a function of temperature. The low- and high-temperature data sets are shown with different symbols, but are fitted with the same straight line. Firstly, it can be seen that the temperature measurement in the two systems is quite consistent, bearing in mind that devices from two different (but essentially identical) wafers are measured. The
slight deviation of the measured turn-on voltages at the three highest temperatures is due to the influence of leakage current near $I_C = 1$ mA at these temperatures. Therefore, these points are disregarded when fitting a straight line to the measured data. The \textit{thermal-electric feedback coefficient}, $\phi$, is defined as the rate of turn-on voltage variation with temperature, which in this case is equal to $-1.4$ mV/K. This value is somewhat larger than those reported for small-geometry transistors ($\sim -1.0$ to $-1.2$ mV/K) in [326,347,378]. This can be explained by the fact that $\phi$ increases logarithmically with collector current density [420], and the turn-on voltages for our large-geometry transistors are defined at very low current densities.

![Figure 9.20](image1.png)

**Figure 9.20**- The measured collector and base current of a large-geometry InGaP/GaAs DHBT as a function of $V_{BE}$ at various temperatures in the range (298-623)K.

![Figure 9.21](image2.png)

**Figure 9.21**- Extracted current gains from the Gummel plots of Fig. 9.20 as a function of $I_C$ at different temperatures.

![Figure 9.22](image3.png)

**Figure 9.22**- Collector current turn-on voltages (measured at $I_C = 1$ mA) versus temperature extracted from Gummel plots of Figs. 9.19-9.20.

The negative temperature coefficient of the B-E turn-on voltage has important consequences as to the thermal stability of bipolar transistors. This is especially important for the design of multi-finger power amplifiers. It is quite likely that during the high power operation of a
multi-finger bipolar transistor one individual finger gets more than its share of the total base current. This can be caused by design or fabrication asymmetries between the fingers, and even by the current crowding phenomena. Since more power is dissipated in this finger of the transistor, it gets warmer than other fingers. A negative thermal-electric feedback coefficient causes a negative shift in $V_{BE}$ and even more increase in $I_C$ of this finger. This process carries on until one finger of the transistor (the hot finger) hogs the majority of the total base current from the remaining fingers (the cold fingers). This is an intrinsic property of all bipolar transistors which can lead to thermal instabilities. Nevertheless, due to the differences in the dependence of current gain on junction temperature in various types of bipolar transistors, the above phenomena occurring during thermal instability lead to different manifestations in their I-V characteristics. In devices like silicon BJTs, where the current gain increases with increasing junction temperature [421: p.29], the current gain of the hot finger will get larger and larger, leading to even higher collector currents, until the device burns-out (i.e., the thermal runaway [420]). In contrast, the current gain of AlGaAs/GaAs HBTs decreases significantly with increasing junction temperature. Therefore, when the total base current of the device is constant (as in the common-emitter output characteristics), the fact that majority of the current flows through the hot finger causes a dramatic reduction of current gain and collapse of collector current in the output characteristics [378-379]. Although the thermal instability of bipolar transistors can be avoided by using emitter ballasting resistances [378,422], this resistance degrades the high frequency performance of the device. Therefore, it is highly desirable to design transistors, such as InGaP/GaAs DHBTs, whose current gain does not vary with either temperature or collector current [420].

9.2.3. Output Characteristics

The common-emitter output characteristics of the InGaP/GaAs DHBT at three different temperatures are shown in Figs. 9.23(a), (b), and (g). It can be seen that the device operates almost perfectly up to 573K (300°C), and it is still functional, though leaky, at 623K. InGaP/GaAs SHBTs show I-V characteristics at $T = 573K$ (Fig. 9.23(d)), which is as leaky as that of the DHBT at $T = 623K$; majority of the InGaP/GaAs SHBTs failed above 598K. The performance of the InGaP/GaAs DHBTs in this work at $T = 623K$ is even considerably better than those reported for single HBTs (at the same temperature) based on $\text{Al}_{0.45}\text{Ga}_{0.55}\text{As}$ [410] and $\text{Al}_{0.18}\text{Ga}_{0.34}\text{In}_{0.48}\text{P}$ [423] emitter materials, which both have wider bandgaps than the $\text{In}_{0.49}\text{Ga}_{0.51}\text{P}$ emitter used in the present work. As mentioned earlier, this improved high temperature performance is due to the very small collector leakage current in InGaP/GaAs DHBTs. The output characteristics of a (nominally) graded AlGaAs/GaAs SHBT, as specified in Table 9.3, is also shown in Figs. 9.23(e) and (f) for $T = 298$ and 573K, respectively. This device is measured at lower current levels compared to the InGaP/GaAs devices. Also its room temperature current gain is significantly larger than those of the InGaP HBTs due to the higher quality of the base and lower base doping concentration. Nevertheless, the current gain of this AlGaAs/GaAs HBT dramatically reduced at $T = 573K$, and also the device became
Chapter 9

DC Characterisation of HBTs

quite leaky. As far as the leakage current at high temperature is concerned, the situation for AlGaAs/GaAs SHBTs is more or less similar to that for InGaP/GaAs SHBTs, since the collector and base material of both devices are GaAs.

Figure 9.23- Common-emitter output characteristics of (a)-(b) an InGaP/GaAs DHBT, (c)-(d) an InGaP/GaAs SHBT, and (e)-(f) an AlGaAs/GaAs SHBT at T = 298K and 573K, respectively. Also shown (g) is the characteristics of the InGaP/GaAs DHBT at 623K.
The measured offset voltage, as expected, is smallest in the InGaP/GaAs DHBT and largest in the AlGaAs/GaAs SHBT. It also shows some increase at higher temperatures mainly due to the existence of $V_T$ in the first and second terms on the RHS of (9.9).

### 9.2.4. Breakdown Voltage

Figure 9.25 shows the reverse bias B-C junction characteristics of the InGaP/GaAs DHBT at various temperatures in the range 78-295K when the emitter terminal is open. Although the reverse bias leakage current, $I_{CBO}$, increases at higher T, the actual junction breakdown occurs at higher voltages when increasing temperature. The extracted open-emitter B-C breakdown voltages, $BV_{CBO}$, are plotted in Fig. 9.24 as a function of temperature. The impact ionisation coefficients of InGaP are known to decrease with increasing temperature [424], leading to a positive temperature coefficient for breakdown voltage of InGaP collector HBTs. This is similar to the trends observed for GaAs [424-426], In$_{0.48}$Al$_{0.52}$P [427], Si and Ge [316: pp. 105-107] collectors of bipolar transistors, but is opposite to the anomalous temperature dependence of breakdown voltage observed for In$_{0.53}$Ga$_{0.47}$As collectors [425-426].

### 9.2.5. Ideality Factors and Saturation Currents

Considering the definitions of the ideality factors and saturation currents as given in Eqs. (9.1) and (9.3) for the forward active mode of operation (i.e., $I_C = I_{CC}$ and $I_B = I_{BE}$), and combining them with Eqs. (7.6) and (7.7) used for analysis of temperature dependence of $\beta$, yields:

$$I_{CCS} = CC(T) \cdot \exp(-E_1 / kT)$$

$$I_{BES} = BB(T) \cdot \exp(-E_2 / kT)$$

where $E_1$ and $E_2$ are effective barriers, or the so-called activation energies, for collector and base current flows, respectively. Assuming that the temperature dependence of the
exponential pre-factors is negligible compared to the exponential dependence on $T$, and also assuming almost constant ideality factors for base and collector currents, one can extract the activation energies from a logarithmic plot of saturation currents versus $(1/T)$, i.e., the *Arrhenius* plot [428]. Furthermore, in the forthcoming analysis it is assumed that the device junction temperature is the same as the ambient temperature. This assumption is justified by the facts that $V_{\text{CB}} = 0$ is used for all of the Gummel plots, and also the transistors analysed here are all large-geometry (B-E area of $7.85 \times 10^{-5}$ cm$^2$) HBTs. This makes the dissipated power density quite small. In cases where the above assumption is not valid, the real junction temperatures and thermal resistances can be measured directly using electrical techniques [429-431].

![InGaP/GaAs DHBT](image)

**Figure 9.25**- Open-emitter reverse bias B-C junction I-V characteristics of a large-geometry InGaP/GaAs DHBT at various temperatures.

The saturation currents and ideality factors of the InGaP/GaAs DHBT were extracted from Gummel plots of Figs. 9.19 and 9.20. The large collector leakage currents at high temperatures (> 498K) make the extraction of above parameters quite difficult; erroneously large ideality factors and saturation currents may be read if the leakage currents are not taken care of. In order to extract these parameters more accurately, a constant collector leakage current (equal to the average of $I_c$ at low $V_{\text{BE}}$'s) were first subtracted from the measured data. Extracted saturation currents and ideality factors of the above DHBT are shown in Figs. 9.26 and 9.27, respectively. Straight line fits to the plots of Fig. 9.26 yield $E_1 = 1.59$ eV and $E_2 = 1.45$ eV. (Notice the different vertical scales used for base and collector saturation currents in Fig. 9.26.) Deviations of the measured saturation currents from the fitted straight lines at low temperatures are due to the increase of ideality factors (see Fig. 9.27). The above activation energies are in close agreement to the analytical values of $E_1 \approx E_{\text{gB}} + \Delta E_{\text{fn}}$ and $E_2 \approx E_{\text{gB}}$.
Chapter 9 DC Characterisation of HBTs

given in [24], assuming that the base current is dominated by base bulk recombination. (Bear in mind that both $E_{gB}$ and $\Delta E_{fh}$ are temperature dependent.) This is also discussed in section 7.4. To further confirm the above activation energies, numerical simulation of the above DHBT was performed for the temperature range 100-700K. The numerically obtained activation energies are $E_1 = 1.56$ eV and $E_2 = 1.52$ eV, which are in good agreement to the above measured values.

Variation of collector and base ideality factors in Fig. 9.27 shows that for $T > 200$K, both $\eta_C$ and $\eta_B$ remain close to unity, confirming that the base bulk recombination and/or back injection of holes remain the dominant components of base current in this temperature range. However, at lower temperatures, both ideality factors show some increase similar to those observed for InP/InGaAs [348] and AlGaAs/GaAs [432] HBTs. The slow increase of $\eta_C$ at low temperatures is due to the fact that the transport of electrons across the B-E heterojunction becomes dominated by tunnelling, since at low $T$ carriers do not have sufficient energy to overcome the conduction band barrier by thermionic-emission. However, since the conduction band spike of the InGaP/GaAs heterojunction is quite small, the observed increase of $\eta_C$ is much smaller than those for InP/InGaAs [348] and AlGaAs/GaAs [432] HBTs. The enlargement of $\eta_B$ at low temperatures, on the other hand, is quite dramatic. This can be explained by the dominance of SCR recombination current at lower temperatures. Since SCR recombination current has an ideality factor of $\sim 2$, it decreases at low $T$ with a rate much slower than that for other components of base current (such as base bulk recombination) which have ideality factors of near unity. Therefore, SCR recombination eventually becomes the dominant component of base current at low temperatures.

9.2.6. Current Gain

It was shown in Fig. 9.21 that the current gain of the InGaP/GaAs DHBT studied in this
Chapter 9 DC Characterisation of HBTs

Chapter is almost insensitive to variation of temperature in the range (298-623)K. This was expected, though, from Eqs. (7.6) and (7.7), since both $I_C$ and $I_B$ have similar ideality factors and activation energies. In order to study the variation of $\beta$ with temperature for the above DHBT in more details, numerical simulations are performed in this subsection. We start by fitting the calculated collector and base currents to the measured ones. The important point to consider is that the fitting procedure is only performed at room temperature, by adjusting only three parameters, then all the parameters (except $R_E$ which will be discussed later) will be fixed for the simulation at other temperatures.

Figure 9.28- Comparison between the measured and numerically simulated collector currents of a large-geometry (emitter diameter of 100 $\mu$m) InGaP/GaAs DHBT. The only fitting parameters used are $\Delta E_C(\text{InGaP/GaAs})$ and $R_E$, the values of which are given in the figure.

Figure 9.28 shows the measured and simulated collector current of the above InGaP/GaAs DHBT at various temperatures in the range (190-598)K. As mentioned in the last paragraph, the collector current at $T = 298$K is first fitted by adjusting $\Delta E_C(\text{InGaP/GaAs})$ to 60 meV (very similar to the value of 50 meV used for the kink effect fitting in Fig. 9.13), and $R_E$ to 3.5 $\Omega$. Then all the parameters, except $R_E$, are kept fixed for other temperatures. Since $R_E$ represents an external resistance (mainly from emitter ohmic contact) whose temperature dependence is not included in the model, slight variations of this parameter was required in order to bring perfect agreement to the high current range of the measured data. Values of $R_E$ equal to 3, 3.5, 4, 5, and 6 $\Omega$ were used for temperatures 190, 298, 398, 498, and 598K, respectively. The agreement between the measured and simulated data is extremely good, and once again proves the accuracy and validity of the semiconductor parameters defined in chapter 5. Since the developed numerical model is one-dimensional, it can not properly consider the leakage currents originating from the extrinsic B-C junction and/or surfaces.
Therefore, the simulated leakage currents are much smaller than the measured ones.

Once the collector current is correctly fitted, the next step is to reproduce the measured $I_B$ or current gain of the device. This is demonstrated in Fig. 9.29, which shows the measured and simulated current gain of the device as a function of $T$. If no further fitting parameters are considered, the calculated current gains (dashed line) are 2-3 times higher than the measured data, and also show sharper variations with $T$ both at low and high temperatures. The peaking of the current gain at medium temperatures is due to the competition between the positive temperature coefficient of the radiative lifetime, negative coefficient of the Auger lifetime, and the reduction of emitter injection efficiency at high temperatures due to the back injection of holes, as explained in subsection 7.4.2. The same peaks in the $\beta$ versus $T$ data is also observed by other authors [347,385]. However, the slow variation of the measured gain suggests that a temperature insensitive recombination lifetime, such as defect-related SRH recombination, may be dominant inside the base. Therefore, the simulations were repeated, this time by considering a thin portion of the base, just above the 200Å B-C spacer layer, as an interfacial layer with a short defect-related lifetime of 6 ps. (This is the third and final fitting parameter.) The result is shown in Fig. 9.29 with solid line. It can be seen that the current gains thus calculated are much closer to the measured data and show slight variations with temperature, just as the experimental data do. Nevertheless, still some discrepancies can be observed at very high ($T > 573K$) and very low ($T < 190K$) temperatures. The measured data shows some slight increase in gain for $T > 548K$, which is not reproduced by the calculations. The origin of this gain increase, also seen in some other works [30,423], is not clear, but it may be related to the less significance of the B-C heterojunction conduction band spike at very high temperatures.

A similar fitting procedure is repeated in the cases of InGaP/GaAs and AlGaAs/GaAs SHBTs studied earlier. The InGaP SHBT also required $\Delta E_C$(InGaP/GaAs) = 100 meV and extra base SRH recombination rate with effective lifetime of 15 ps, but no extra recombination rate was added for fitting to the AlGaAs results. The results are all shown in Fig. 9.30, where the normalised measured current gains of all three types of HBTs are compared with numerical simulations. Good agreement can be observed between the two sets of data. The main contribution to the significant temperature sensitivity of $\beta$ in AlGaAs/GaAs HBTs comes from the slower variation of SCR recombination current with $T$, as discussed in subsection 7.4.2. InGaP HBTs (with large valence band discontinuity) still show some reduction of gain at high $T$ primarily due to the negative temperature coefficient of the Auger recombination lifetime. Existence of defect-related SRH recombination sites, which are characterised by a temperature independent recombination lifetime, diminishes the current gain reduction at high temperatures. This, together with the carrier confinement effect of the B-C heterojunction, is responsible for the less temperature sensitivity of current gain in the InGaP/GaAs DHBT compared to the SHBT. Therefore, if a temperature stable current gain is needed for an specific application, one possibility is to introduce (deliberately) some deep SRH traps inside
the base layer, at the expense of having a lower gain. These traps are introduced automatically to some degrees when growing the GaAs base layer on top of the InGaP collector in DHBTs.

Figure 9.29- Comparison between the measured (symbols) and simulated (lines) current gains vs. temperature for the InGaP/GaAs DHBT specified in Table 9.1. Both simulated current gains with (solid line) and without (dashed line) the extra base SRH recombination ($\tau_{\text{int}} = 6$ ps) are shown.

Figure 9.30- Comparison between the measured (symbol) and simulated (solid lines) normalised current gains vs. temperature for the InGaP/GaAs single- and double-HBTs (Table 9.1) and the AlGaAs/GaAs SHBT (Table 9.3). The calculated results for the InGaP single- and double-HBTs include extra base SRH recombination rates with effective lifetimes of 15 and 6 ps, respectively.

The above analysis invalidates the simplistic view that “as long as the back injection of holes from base to emitter is negligible and the base current is dominated by the base bulk recombination, the current gain would remain insensitive to the variations of temperature [30].” This is also supported by the significant gain variations with temperature for InGaP/GaAs HBTs (which have negligible hole back injection) having high base doping concentrations ($\geq 5 \times 10^{19}$ cm$^{-3}$) reported in [155,326].
CHAPTER 10. HIGH FREQUENCY CHARACTERISTICS OF HBTs

Accurate extraction of the small-signal equivalent circuit parameters of HBTs is crucial for optimising device performance. The most commonly used extraction technique is numerical optimisation of the model generated s-parameters to fit the measured ones [414]. It is well known, however, that optimisation techniques may result in non-physical and/or non-unique values of the components [433]. Also the optimised parameters are largely dependent on the initial values of the optimisation process [434]. Alternative extraction methods which ensure unique determination of as many equivalent circuit elements as possible are therefore of considerable importance. Several approaches for a more accurate and more physical parameter extraction are suggested in the literature which can generally be divided into two major categories. The first method mainly makes use of special test structures and DC measurements to extract the extrinsic elements [435], while the second technique relies on analytically derived expressions for a direct calculation of the equivalent circuit element values [436].

As discussed in section 10.3, each of the above two methods, in its original form, has some drawbacks. Therefore, since 1992 various authors tried to overcome the shortcomings of these methods by combining them with measurements of the actual device under test at different biasing conditions. Among these, the so-called "cold-HBT" measurement [437], "open-collector" measurement [438], and variable current measurement [439] are the most popular.

In this thesis, a new parameter extraction technique is introduced which makes use of some of the above mentioned approaches to accurately determine the equivalent circuit elements of HBTs. Among the advantages of this new technique are a minimum use of the test structures, unique and physically meaningful values of the parameters, no reference to numerical optimisation, ease of implementation with minimum number of iterations and almost no reference to DC results, and very reasonable assumptions which make the method applicable not only to GaAs-based HBTs but to other types of III-V HBTs as well. Alongside the description of the extraction procedure, very useful information as to the accurate evaluation of forward transit time and cutoff frequency of the device without using extrapolation techniques will be discussed. Also the contribution of each equivalent circuit element to the high frequency figures of merit, $f_T$ and $f_{max}$, will be clarified.

The structure of this chapter is as follows. In section 10.1, the two most commonly used small-signal equivalent circuit configurations for HBTs, namely the T- and π-equivalent circuits are introduced and compared. Section 10.2 discusses the measurement techniques for the high frequency performance of HBTs. Then in section 10.3 a new small-signal equivalent circuit parameter extraction procedure is presented, which is followed by discussion of the cutoff frequency, maximum frequency of oscillation, and various components of delay times in section 10.4. The bias and geometry dependence of some of the extracted parameters are
explained in details in section 10.5. And finally section 10.6 is devoted to a sensitivity analysis of the high frequency figures of merit.

10.1. Small-Signal Equivalent Circuit of HBTs

An accurate physics-based small-signal model is crucial for the design of HBT-based circuits for microwave and millimetre-wave applications. In the past, several approaches were used to represent the intrinsic part of the HBT in the form of lumped-element circuits. Originally, the y-parameters of the intrinsic HBT in the common-base mode of operation were derived based on a drift-diffusion (DD) current formalism [44,388]. Although high frequency characteristics of HBTs have been computed accurately based on DD formulation using numerical techniques [375], some approximations are needed before one can use the DD equations in compact equivalent circuit models of HBTs. Such approximations are discussed in [44] and the immediate outcome is the T-shaped equivalent circuit shown in Fig. 10.1. In this figure, $C_{bic}$ is the part of the B-C depletion capacitance which is right underneath the emitter mesa (i.e., intrinsic area), $R_{be}$ is the forward-biased dynamic resistance of the B-E junction, $C_{be}$ models the charge storage inside the depletion region of the B-E junction as well as the base neutral region, and $\alpha(\omega)$ is the common-base current gain of the device. In terms of the physical parameters, $R_{be}$, $C_{be}$, and $\alpha(\omega)$ can be expressed as [44,388]:

\[
R_{be} = \frac{\eta_e kT}{qI_E} \quad (10.1)
\]

\[
C_{be} = C_{jbe} + m \cdot \tau_B / R_{be} \quad (10.2)
\]

\[
\alpha(\omega) = \frac{\alpha_0 \cdot \exp[-j\omega(\tau_C + (1-m)\tau_B)]}{1 + j\omega(m\tau_B + R_{be}C_{jbe})} \cdot \frac{\sin(\omega\tau_C)}{\omega\tau_C} \quad (10.3)
\]

\[
\tau_B = \frac{W_B^2}{2D_{nB}} + \frac{W_B}{v_{n,avg}} \quad (10.4)
\]
where $\eta_e$ is the ideality factor of the emitter current, $I_E$ is the DC emitter current, $C_{jbe}$ is the B-E depletion capacitance, $\tau_B$ and $\tau_C$ are the base transit time and B-C depletion region delay time, respectively, $\alpha_0$ is the low frequency common-base current gain, $D_{nb}$ is the diffusion constant of electrons inside the base, $v_{n,\text{avrg.}}$ is the average velocity of electrons inside the B-C depletion region of width $W_{BC}$, and other symbols have their usual meanings. $m$ is an empirical constant that matches the phase and magnitude of the common-base current gain to those of the more accurate secant hyperbolic representation [44,388]. An appropriate value of $m$ for a uniform base is found to be $5/6$ [436,440].

The common-emitter $y$-parameters of the intrinsic HBT can be calculated directly from the common-base parameters, but transforming these parameters into a useful one for lumped-element equivalent circuit needs even further approximations. The final product is the $\pi$-equivalent circuit shown in Fig. 10.2. For the purpose of small-signal modelling of HBTs both T-shaped [406,414,434,436,438,439,441-450] and $\pi$-shaped [326,435,437,451-455] equivalent circuits are examined in the literature. Laser and Pulffrey [440] have claimed that the two representations are identical provided that $r^\pi = R_{be}/(1-\alpha')$, $y_m = \alpha'/R_{be}$, and $C_{\pi} = C_{be}$, where $\alpha'$ has an expression similar to (10.3) for $\alpha(\omega)$, but without the term $[1+j\omega(m\tau_B + R_{be}C_{jbe})]$ in the denominator. Additionally, they have shown that the calculated maximum frequency of oscillation, $f_{\text{max}}$, using the above two approaches is almost identical to the one calculated by the more accurate DD formalism. However, the fact that is not considered in [440] is that $\alpha'$ is a complex function of $\omega$, and therefore, $r^\pi$ turns out to be a complex impedance rather than the more usually known resistance of magnitude $r_e/(1-\alpha_0)$. To use $r^\pi$ as a pure resistance, one has to change the input capacitance to:

$$C_{\pi} \approx C_{jbe} + (\tau_B + \tau_C)/R_{be}$$

(10.6)

which can be significantly different from $C_{be}$. This last point has also been mentioned by Liu [44: pp.663-671].

In general, the $\pi$-equivalent circuit of Fig. 10.2 suffers from three major disadvantages as compared to the T-model. First, the $\pi$-model requires some assumptions further to those used for the derivation of T-equivalent circuit from the more accurate DD approach. Second, the input capacitance in the $\pi$-model, $C_{\pi}$, includes a term related to the delay time inside the B-C depletion region, which seems peculiar in physical grounds. Third, the current source in the $\pi$-model is dependent on $v_{bc'}$, which is an internal voltage of the transistor, while in the T-model the current source is dependent on the terminal current, $i_e$. This makes the parameter extraction process for the $\pi$-model more complicated. Therefore, in this work it has been decided to use the T-shaped equivalent circuit of HBTs.

Next, the extrinsic and parasitic elements are to be added to the intrinsic equivalent circuit of
HBTs shown in Fig. 10.1. Fig. 10.3 shows a schematic diagram of a small-geometry HBT fabricated in our laboratory together with the corresponding small-signal equivalent circuit elements. The circuit elements include intrinsic as well as extrinsic and parasitic elements of the HBT. $C_{Pbc}$, $C_{Pbe}$, and $C_{Pce}$ represent the contact pad capacitances of the devices made for microwave measurement together with the fringing capacitances along device fingers for B-E, B-C, and C-E terminals, respectively. Since in the common-emitter HBTs the contact pads for the base and collector are smaller than the emitter pad, $C_{Pbc}$ is expected to be the smallest among the above three parasitic capacitances. Also $C_{Pce}$ includes the crossover capacitance between the emitter and collector metal lines (∼10fF per crossover), and therefore, it would be the largest pad capacitance. The above pad capacitances are typically < 50fF. An imperfect ion bombardment isolation of the inactive areas severely enhances these pad capacitances and degrades device performance. Hence, measured values of pad capacitances on a test structure especially made for this purpose (see section 10.2 for more details) can be used to monitor the implant isolation process.

Figure 10.3- Schematic diagram of an npn HBT together with its associated small-signal lumped-element T-equivalent circuit.

Metal interconnections for the three terminals are modelled by a series combination of inductances ($L_{ei}$, $L_{eb}$, and $L_{ec}$) and resistances $R_{ee}$ models the series resistances of the emitter ohmic contact, emitter cap layer, and the undepleted part of the emitter. $R_{bbx}$ and $R_{ccx}$ include the ohmic contact and extrinsic resistances of the base and sub-collector regions, respectively. The distributed nature of the base access resistance and B-C junction capacitance is modelled by dividing the base mesa area into intrinsic and extrinsic regions. In that respect, $R_{bbi}$ and $R_{cci}$ model the intrinsic parts of the base and sub-collector series.

$^1$ The interconnection series resistances are merged with the extrinsic device series resistances $R_{net}$, $R_{bbx}$, and $R_{ccx}$. 231
resistances, respectively. Both intrinsic and extrinsic parts of the B-C depletion region are modelled with a parallel combination of a depletion capacitance and a reverse-biased junction dynamic resistance.

Since the sub-collector region is thick (usually > 5000Å) and the electron mobility of the material forming this layer (GaAs or In$_{0.53}$Ga$_{0.47}$As) is very high, $R_{cc1}$ would be a very small resistance in npn HBTs (~1 Ω or less). Therefore, one can merge this resistance into the extrinsic part of the collector resistance without seriously affecting the accuracy of the small-signal model. The final product is a T-shaped equivalent circuit as shown in Fig. 10.4, in which $R_{cc}$ is an effective combination of $R_{cex}$ and $R_{cel}$. The equivalent circuit of Fig. 10.4 is used for the parameter extraction technique discussed in section 10.3.

**Figure 10.4** - The T-shaped small-signal equivalent circuit of HBTs.

### 10.2. High Frequency Small-Signal Measurement

The frequency response of any two-port linear network can be fully described by a set of four frequency dependent parameters. These network parameters can be represented in several forms associated with the input and output variables being chosen. Admittance parameters ($y$-parameters), for example, are associated with the currents being represented as a function of
potentials; impedance parameters (z-parameters) are associated with the voltages being written as a function of currents; hybrid parameters (h-parameters) are a combination of these; and finally scattering parameters (s-parameters) relate incident and reflected power waves.

The y-, z-, or h-parameters, though conceptually simple, may be too difficult to be measured experimentally. Measuring these parameters calls for the input and output of the device to be successively opened and short circuited. This can be hard to do, especially at microwave frequencies [456]. For example, as stated in transmission line theories, an open at a certain frequency may behave as short at another frequency [44]. Also the abrupt termination of a port often results in oscillation at high frequencies [44]. S-parameters, on the other hand, are usually measured with the device embedded between a 50Ω load and source, and there is no need to produce open or short circuits. Therefore, s-parameters are typically measured to characterise the high frequency performance of a transistors. Then, if needed, the measured s-parameters may be converted to other network parameters for analysis. A brief description of various network parameters, their transformation expressions, and formulation of various power gains and stability factor of a two-port network are presented in Appendix A. More details about the physical concept of scattering parameters and various device power gains can be found in standard textbooks such as [457].

In this work, the s-parameters are measured on-wafer using HP8510 Network Analyser and Cascade Microtech RF probes in the frequency range 50MHz to 40GHz. During high frequency measurement, there are two types of errors that need to be minimised in order to have an accurate description of the device under test: systematic errors and parasitic errors. Systematic errors are due to the measurement system and include effects such as cable losses and non-ideal probe impedance (source and load match) [458]. These errors are minimised with a network analyser calibration. Parasitic errors, on the other hand, are due to the extrinsic device test structure and include pad capacitances and interconnect inductances. De-embedding techniques must be used to isolate the intrinsic device performance from the parasitic effects that show up in the measurement. In addition, there are some environmental factors, such as temperature, humidity, light, vibration, and probe tip pressure, that need to be controlled to minimise the error.

A calibration is performed by using on-wafer calibration standards. There are several calibration methods in use today, namely LRM (Line-Reflect-Match), TRL (Thru-Reflect-Line), and SOLT (Short-Open-Load-Thru). Description of these methods and their relative advantages and disadvantages are outlined in [458]. What is important in the context of the present work is that after calibration, the reference plane will be located at the probe tips. This means that all of the systematic errors due to the network analyser, cables, bias network, and probes have been accounted for. It is important to note that a calibration is only valid for the frequency range and power chosen prior to the calibration process.

Almost always during the RF measurement of HBTs, it is required to apply a DC bias on top
of the small RF signals. This can be achieved using a bias T (Fig. 10.5), where the DC and RF signals do not have any loading effect on each other. This is because inductance (capacitance) of the bias T will behave as open circuit for RF signal (DC) and short for DC (RF signal). Also a Kelvin (force-sense) connection is made inside the bias network to minimise the effect of cable losses on the measured data [458].

Once the calibration is complete, the systematic errors down to the tip of the probe are minimised. However, there remain some extrinsic parasitics between the probes and the intrinsic device. As an example, consider the test HBT made for RF characterisation shown in Fig. 10.6. This test HBT is designed to allow RF measurements with standard GSG (ground-signal-ground) probes. This figure also shows the contact pad parasitic capacitances and the metal interconnect inductances as well as the unavoidable emitter-collector metal crossover.

The next step is to de-embed the intrinsic device from the parasitics surrounding it. A standard method to characterise the pad capacitances is to measure the s-parameters of an open test structure fabricated on the same wafer as the device under test. The open device is realised in our laboratory's HBT technology by ion implantation of both the active and inactive areas of the device. Therefore, all the etching and metallisation of the device fingers are present in the open test structure; the only difference is that there is no active area. As shown in Fig. 10.7.a, the open test device can be represented by three pad capacitances only. The y-matrix of the open device has the following form:
Chapter 10  High Frequency Characteristics of HBTs

\[
[Y_{\text{open}}] = \begin{bmatrix}
 j\omega(C_{Pbe} + C_{Pbc}) & -j\omega C_{Pbc} \\
 -j\omega C_{Pbc} & j\omega(C_{Pce} + C_{Pbc})
\end{bmatrix}
\]  

(10.7)

Figure 10.7- (a) The equivalent circuit of the open test structure, and (b) the effect of the pad capacitances on the admittance matrix of the device under test.

Now referring to Fig. 10.7.b, one can relate the de-embedded y-matrix \([y]\) to the measured y-matrix \([Y]\) (including the pad capacitances) and the values of the pad capacitances as:

\[
\begin{bmatrix}
 y_{11} & y_{12} \\
 y_{21} & y_{22}
\end{bmatrix} = \begin{bmatrix}
 Y_{11} - j\omega(C_{Pbe} + C_{Pbc}) & Y_{12} + j\omega C_{Pbc} \\
 Y_{21} + j\omega C_{Pbc} & Y_{22} - j\omega(C_{Pce} + C_{Pbc})
\end{bmatrix}
\]

or simply:

\([y] = [Y] - [Y_{\text{open}}]\)  

(10.8)

Fig. 10.8 shows the measured parasitic pad capacitances of an open test structure fabricated on EPI 1-3514A InGaP/GaAs DHBT wafer (see Table 10.1). As can be seen from this figure, the extracted pad capacitances are fairly constant with frequency, which confirms the validity of modelling the open test structure with the three capacitances as shown in Fig. 10.7.a. Also application of forward or reverse bias to the terminals of this device did not change the extracted values of capacitances significantly. This means that the capacitances shown in Fig. 10.8 are entirely due to the pad parasitics, and not due to junction capacitance. Therefore, ion bombardment has been successful in isolating neighbouring devices.

An almost similar method can be used to find the values of the series inductances from a short test structure as described in [435]. These inductances can be directly subtracted from the z-parameters of the device after de-embedding from pad capacitances as above. A third test structure has also been suggested in [435] in which the emitter layer and the intrinsic parts of the

![Figure 10.8](image-url)
base and collector layers are etched away. This last structure should help in finding the values of $R_{bbx}$ and $C_{bcx}$. However, it is quite clear that the extrinsic elements obtained from the above two test structures (especially those obtained from the last structure) are largely dependent on device geometry and number of fingers. Therefore, ideally one should have one short test structure and one structure with the intrinsic area removed for every single device geometry on the fabricated mask! Additionally, the third test structure as discussed above, requires an extra mask and fabrication step in our existing HBT technology. This is because the collector metallisation is self-aligned to the base mesa in our existing technology, and this means that all the areas etched away during the base mesa formation will be covered with collector metal during the next step. But the third test structure requires an etched base mesa without metallisation.

The parameter extraction technique used in this work does not make use of the second and third test structures. Instead, as discussed in the next section, all of the equivalent circuit elements of the device under test will be found analytically from measured s-parameters of the same device under various biasing conditions (including cold-HBT measurement).

### 10.3. Parameter Extraction Procedure

In this section, a new method for the extraction of the small-signal elements of the T-shaped equivalent circuit shown in Fig. 10.4 will be discussed. This method has been applied to devices with various geometry on various wafers from different material systems including InGaP/GaAs (both single and double HBT), AlGaAs/GaAs, and InP/InGaAs npn HBTs, and has shown extremely good consistency. This is mainly due to the very reasonable simplifying assumptions made during the analytical parameter extraction procedure, as discussed next. The suitability of this technique will be demonstrated by examples from an InGaP/GaAs DHBT wafer (#EPI 1-3514A), the layer structure of which is given in Table 10.1.

| Table 10.1- Layer structure of the InGaP/GaAs DHBT wafer no. EPI 1-3514A. |
|-----------------|---------------------|---------------|----------|------------------|
| Material        | Thickness (Å)       | Doping (cm⁻³) | Dopant   | Comments         |
| n⁺-GaAs         | 2800                | 4×10⁻ⁱ⁸       | Si       | Cap layer        |
| n⁺-In₀.₄₉Ga₀.₅₁P | 200                 | 2×10⁻¹⁸       | Si       | Cap layer        |
| n⁻-In₀.₄₉Ga₀.₅₁P | 1000                | 3×10⁻¹⁷       | Si       | Emitter          |
| p⁺-GaAs         | 1000                | 3×10⁻¹⁹       | C        | Base             |
| n⁻-GaAs         | 200                 | 1×10⁻¹⁶       | Si       | Spacer           |
| n⁻-In₀.₄₉Ga₀.₅₁P | 4800                | 5-7×10⁻¹⁶     | Si       | Collector        |
| n⁺-In₀.₄₉Ga₀.₅₁P | 1000                | 2×10⁻¹⁸       | Si       | Sub-collector    |
| n⁻-GaAs         | 7000                | 4×10⁻¹⁸       | Si       | Sub-collector    |
| S.I. GaAs       | 635µm               | undoped       | —        | Substrate        |

A final comment has to be made as to the novelty of the present approach. Although some parts of the parameter extraction technique developed by the author is similar to the ones used elsewhere in the literature (e.g., [441-442]), the overall structure of this technique is
completely new. In particular, the information extracted from “cold-HBT” measurement, the fast converging iteration used in the beginning of the method, the extraction of total delay time (τ_{E}) from measured z-parameters at lower range of frequency (and not from somehow troublesome extrapolation of h_{21} with -20dB/dec roll-off at higher frequencies), accurate extraction of forward transit time and its partitioning into base transit time (τ_{B}) and collector depletion region delay time (τ_{C}), and minimum use of DC measured data, are among new features in the present parameter extraction technique. In addition, some parts of the existing technique in [441] are modified to make the method more widely applicable. These include different plotting and/or interpretation of measured data, less restrictive assumptions, and removing the necessity for a final optimisation process. The main results of the present work are also published in [459-460].

10.3.1. Parasitic Pad Capacitance Determination

As discussed in the previous section, the first step in parameter extraction is evaluation of the contact pad capacitances and de-embedding the internal device from these parasitics. The existence of pad capacitances in the equivalent circuit of Fig. 10.4 makes the analytical formulation of network parameters in terms of circuit element values extremely difficult. After all, these contact pads are not part of the real device; they are made to facilitate RF measurement using standard probes. Three methods are suggested in the literature to characterise the pad capacitances. The first method, also recommended in the present work, is the one described earlier as measurement of an open test structure. The measured values of parasitic capacitances were almost constant with bias and frequency, and also showed a standard deviation of less than 9%. However, there is an argument that one open test structure with bombarded active area and with an unchanged device layout can not correctly represent the pad parasitics of all device geometries [437,461]. This is due to the fringing capacitances induced along the fingers of the actual device. In order to assess the effect of the fringing capacitances on the extracted pad capacitance of the device, three different open test structures are designed and fabricated on the same wafer as device under test: (o1) a test pattern with contact pads only, even without the C-E crossover; (o2) the same pattern as in (o1) but with two C-E crossovers for two collector ohmic contacts; (o3) a full pattern for an (8×10)×2 μm² HBT including all the details of device metallisation, but with the active area of the device bombarded. The results for a sample (o3) structure are shown in Fig. 10.8, and the results of the two neighbouring structures of the other kinds are summarised in Table 10.2. Comparing (o2) and (o3) one can see that the maximum difference is around 4fF, and this accounts for the entire metallisation capacitance of a two-finger device. Therefore, one may conclude that the difference in the fringing capacitances of various device geometries is in the range 1-2fF at most, and may be ignored. On the

<table>
<thead>
<tr>
<th>Open Structure</th>
<th>C_{P_b}\text{ (fF)}</th>
<th>C_{P_c}\text{ (fF)}</th>
<th>C_{P_{ce}}\text{ (fF)}</th>
</tr>
</thead>
<tbody>
<tr>
<td>(o1)</td>
<td>8.0</td>
<td>15.7</td>
<td>27.8</td>
</tr>
<tr>
<td>(o2)</td>
<td>7.9</td>
<td>16.0</td>
<td>48.9</td>
</tr>
<tr>
<td>(o3)</td>
<td>12.2</td>
<td>20.0</td>
<td>51.4</td>
</tr>
</tbody>
</table>
other hand, the difference between $C_{Pce}$ of structures (o1) and (o2) is almost 21fF, which is due to the existence of two metal crossovers, and albeit should be taken into account. Existing devices on the designed mask have either one or two collector ohmic contacts. The above analysis suggests that each C-E crossover has almost 10.5fF contribution to the total $C_{Pce}$. Therefore, devices with 2 and 1 collector contacts will have $C_{Pce}$'s of 51.4 and 40.9fF, respectively.

The other two methods of parasitic pad capacitance determination rely on the so-called "cold-HBT" measurement. Cold-HBT refers to the condition where both B-E and B-C junctions are reverse-biased, or the forward bias on them is insufficient for significant current injection. Three major changes will occur in the equivalent circuit of Fig. 10.4 under cold-HBT condition as compared to the forward active mode biasing condition. First, $R_{be}$ will be a very large resistance corresponding to a reverse-biased junction. Therefore, it does not have the normal shorting effect on the input capacitance $C_{be}$, and allows a relatively easy determination of this capacitance. Second, since the injection current is very small, the B-E diffusion capacitance is absent and $C_{be}$ reduces to $C_{jbe}$. Third, and the most important, is that the current source $\alpha(\omega)i_c$ is eliminated, since DC current is zero (or very small) and the common-base current gain would be negligibly small. Ignoring the current source makes the equivalent circuit of the cold-HBT a completely passive one. Therefore, one expects the network matrix of the cold-HBT to be symmetric (i.e., $Y_{12} = Y_{21}$, $Z_{12} = Z_{21}$, etc.). In fact, the latter condition can be used to confirm the validity of cold-HBT assumption. Under such conditions and using some first-order approximations, one can ignore the influence of series resistances and inductances at low frequencies [437,452], and arrive at the simplified equivalent circuit shown in Fig. 10.9. The y-parameters of such equivalent circuit may be written as:

$$[y] = \begin{bmatrix} G_{be} + G_{bc} + j\omega(C_{Pbe} + C_{jbe} + C_{Pbc} + C_{jbc}) & -G_{be} - j\omega(C_{Pbc} + C_{jbc}) \\ -G_{bc} - j\omega(C_{Pbc} + C_{jbc}) & G_{bc} + j\omega(C_{Pce} + C_{Pbc} + C_{jbc}) \end{bmatrix}$$

(10.9)

where $G_{be} = 1/R_{be}$, $G_{bc} = 1/R_{bc} = 1/R_{bei} + 1/R_{bcx}$, and $C_{jbc} = C_{bei} + C_{bx}$. It is now clear that $(C_{Pbc} + C_{jbc})$, $(C_{Pbe} + C_{jbe})$, and $C_{Pce}$ can be found directly from $-\text{Im}(Y_{12})$, $\text{Im}(Y_{12} + Y_{11})$, and $\text{Im}(Y_{12} + Y_{22})$, respectively. Fig. 10.10 shows the measured values of pad capacitances for an HBT with $(4\times10) \times 1 \mu m^2$ B-E area and $14\times10 \mu m^2$ B-C area under $V_{BE} = V_{BC} = 0.0$. It can be seen that the graphs are reasonably flat at the frequency range 1-10GHz. $C_{Pce}$ can be directly found from the plot of Fig. 10.10, and this value agrees reasonably well with that obtained from the open test structure. (To take into account the slight variation of $C_{Pce}$ across wafer, one can use the value obtained from Fig. 10.11)
10.10 for the particular device under test.) However, the parasitic and junction capacitances of the B-E and B-C cannot be distinguished from this graph.

To distinguish between the parasitic and junction capacitances, one can observe the voltage dependence of the measured total capacitances. First, the voltage dependence is discussed. The measured total capacitances of B-E and B-C can be written as:

\[
C_{\text{total}} = C_{\text{Pb}} + \frac{C_{\text{jbk}}}{\sqrt{1 + V_{\text{kb}}/V_{\text{bik}}}} \quad k = "e", "c"
\]

\[
C_{\text{Pbe}} + C_{\text{jbe}} = 23.9 + 10.6/\sqrt{1 + V_{\text{eb}}/1.6} \quad (\text{FF})
\]

\[
C_{\text{Pbc}} + C_{\text{jbc}} = 16.1 + 53.6/\sqrt{1 + V_{\text{cb}}/1.55} \quad (\text{FF})
\]

\[
C_{\text{Pce}} ~ 47.4 \text{ fF}
\]

where \(V_{\text{bie}}\) and \(V_{\text{bic}}\) are the built-in potentials of the B-E and B-C junctions, respectively. Here values of 1.6 and 1.55V are appointed to \(V_{\text{bie}}\) and \(V_{\text{bic}}\), respectively. B-E and B-C total capacitances are measured under various reverse biases using the method outlined above and plotted as a function of \(1/\sqrt{1 + V_{\text{kb}}/V_{\text{bik}}}\) in Fig. 10.11. The y-intercepts of this figure give values of \(C_{\text{Pbe}}\) and \(C_{\text{Pbc}}\), close but not exactly equal to the values obtained from open test structure measurement. This disagreement can partly be due to the fringing capacitances of the actual device, and partly due to the choice of built-in potentials which makes the unique separation of pad and junction capacitances difficult [461]. In addition, particularly in the case of B-E capacitance, full depletion of the emitter or collector n-regions may be a major source of error in determining the pad capacitances. For example, using \(\varepsilon_{\text{InGaP}} = 11.8\) from chapter 5, the B-E and B-C depletion region widths can be calculated from \(W = \varepsilon A / C_j\) to be \(~0.4\mu\text{m}\) and \(~0.27\mu\text{m}\), respectively. This means that the emitter is fully depleted even under zero bias, and applying more negative bias simply extends the depletion region more into the emitter cap layer. Therefore, the simple formula (10.10) will not be valid for the B-E total capacitance.

The above idea of evaluating reverse-biased junction capacitances using cold-HBT s-
parameter measurement is extended by Chang et al. [462] to low current forward-biased conditions. They simply (and correctly) argue that the conditions of cold-HBT assumption are not violated until the current injection becomes significant (or until $Y_{12}$ deviates significantly from $Y_{21}$). Using this method (called RF-CV), Chang et al. were able to extract the values of HBT junction capacitances up to $V_{BE} = 1.25V$ and $V_{BC} = 1.0V$ forward bias, which is significantly larger than the limit imposed by the normal capacitance-voltage measurement ($V_{BE} \approx V_{BC} \approx 0.7V$).

![Figure 10.12](image)

**Figure 10.12-** Area dependence of the total capacitances measured under cold-HBT condition for 1-, 2-, and 4-finger HBTs.

The last method of evaluating pad capacitances makes use of the area dependence of the total capacitances measured from cold-HBT (under no bias) [461]. In the present work, a large number of 1-, 2-, and 4-finger HBTs on the same wafer are measured under no bias condition and the area dependence of the total capacitances are studied. However, one must take the etch undercut of the emitter and base mesas into account when plotting total capacitance versus area. Etch undercuts of 0.5μm and 0.75μm on each side of the emitter and base mesas, respectively, seemed reasonable considering the etch rate and duration, and also created a minimal data scatter for the plots of total capacitances versus area. This plot is shown in Fig. 10.12, which gives $C_{Pee} \approx (51.5 \pm 5)\text{fF}$. The plot of B-E total capacitance versus area is reasonably linear, but there is some scatter in the data of the total B-C capacitance versus area. This scatter is most likely due to the non-uniformity in the collector doping across the wafer. C-V measurements across this wafer estimated the collector doping to be in the range $(5-7) \times 10^{16} \text{cm}^{-3}$, which can cause up to 20% scatter in $C_{jbe}$. This large scatter may cause a significantly erroneous value for $C_{Pbe}$ when using the area dependence of the total capacitance. The large doping variation across the wafer is due to the difficulties in controlling low doping densities in InGaP, according to the grower of the epitaxial layer. The
depletion region widths can be estimated from slopes of the graphs as:

\[ C_{\text{Pbc}} + C_{j\text{be}0} = 15.5 + 0.57A_{\text{BC}} \text{ (fF)} \Rightarrow W_{\text{BC}} = 0.18 \mu\text{m} \]

\[ C_{\text{Pbe}} + C_{j\text{be}0} = 17.7 + 0.85A_{\text{BE}} \text{ (fF)} \Rightarrow W_{\text{BE}} = 0.12 \mu\text{m} \]

which seem much more reasonable than those obtained from bias dependence of capacitances. However, still the entire emitter region is depleted under zero bias. The pad capacitances are also not far from those estimated by open test structure measurement.

In summary, \( C_{\text{Pce}} \) can be estimated directly from cold-HBT data of the actual device under test. However, as discussed above, the bias or area dependence of the total B-E and B-C capacitances measured under cold-HBT condition suffer from problems that still make the open test structure measurement a favourable method for evaluating \( C_{\text{Pbe}} \) and \( C_{\text{Pbc}} \). Nevertheless, a plot similar to the one in Fig. 10.12 should be very useful in scaling device parameters of HBTs, as discussed in subsection 10.5.2.

10.3.2. Z-Parameters and Second-Order Approximations

Once the contact pad capacitances are accurately measured, one can de-embed the internal device from parasitic capacitances surrounding it. After some algebraic manipulation, the z-parameters of the equivalent circuit diagram of Fig. 10.4 without pad capacitances can be written as:

\[
Z_{11} = \frac{R_{\text{bbi}}[(1-\alpha)Z_{\text{bci}} + Z_{\text{bcx}}]}{R_{\text{bbi}} + Z_{\text{bci}} + Z_{\text{bcx}}} + Z_B + Z_E
\]

(10.11a)

\[
Z_{12} = \frac{(1-\alpha)R_{\text{bbi}}Z_{\text{bci}}}{R_{\text{bbi}} + Z_{\text{bci}} + Z_{\text{bcx}}} + Z_E
\]

(10.11b)

\[
Z_{21} = \frac{[(1-\alpha)R_{\text{bbi}} - \alpha Z_{\text{bcx}}]Z_{\text{bci}}}{R_{\text{bbi}} + Z_{\text{bci}} + Z_{\text{bcx}}} + Z_E
\]

(10.11c)

\[
Z_{22} = \frac{(1-\alpha)Z_{\text{bci}}(R_{\text{bbi}} + Z_{\text{bcx}})}{R_{\text{bbi}} + Z_{\text{bci}} + Z_{\text{bcx}}} + Z_C + Z_E
\]

(10.11d)

Therefore,

\[
Z_{11} - Z_{12} = \frac{R_{\text{bbi}}Z_{\text{bcx}}}{R_{\text{bbi}} + Z_{\text{bci}} + Z_{\text{bcx}}} + Z_B
\]

(10.12a)

\[
Z_{22} - Z_{21} = \frac{Z_{\text{bci}}Z_{\text{bcx}}}{R_{\text{bbi}} + Z_{\text{bci}} + Z_{\text{bcx}}} + Z_C
\]

(10.12b)

\[
Z_{12} - Z_{21} = \frac{\alpha Z_{\text{bci}}Z_{\text{bcx}}}{R_{\text{bbi}} + Z_{\text{bci}} + Z_{\text{bcx}}}
\]

(10.12c)

The impedance blocks in Eqs. (10.11) and (10.12) can be derived from Fig. 10.4 as:
Chapter 10
High Frequency Characteristics of HBTs

\[ Z_{\text{bcj}} = \frac{R_{\text{bcj}}}{1 + j\omega C_{\text{bcj}}R_{\text{bcj}}} \]  \quad (10.13)

\[ Z_{\text{bcx}} = \frac{R_{\text{bcx}}}{1 + j\omega C_{\text{bcx}}R_{\text{bcx}}} \]  \quad (10.14)

\[ Z_E = R_{ee} + j\omega L_e + \frac{R_{be}}{1 + j\omega C_{be}R_{be}} \]  \quad (10.15)

\[ Z_B = R_{bbx} + j\omega L_b \]  \quad (10.16)

\[ Z_C = R_{cc} + j\omega L_c \]  \quad (10.17)

Three new parameters are also defined as:

\[ R_{\text{be}} = \left( \frac{1}{R_{\text{bcj}}} + \frac{1}{R_{\text{bcx}}} \right)^{-1} \quad \text{(i.e., parallel combination of } R_{\text{bcj}} \text{ and } R_{\text{bcx}}) \]  \quad (10.18)

\[ C_{\text{bc}} = C_{\text{bcj}} + C_{\text{bcx}} \quad \text{(i.e., parallel combination of } C_{\text{bcj}} \text{ and } C_{\text{bcx}}) \]  \quad (10.19)

\[ r = \frac{C_{\text{bcj}}}{C_{\text{bc}}} \]  \quad (10.20)

Equations (10.11) and (10.12), in their original form, can not be used for the purpose of parameter extraction; simplifying assumptions are needed to make them more useful for this task. If one assumes:

\[ R_{\text{bbi}} << R_{\text{bc}} \]  \quad (10.21)

\[ [\omega C_{\text{bc}}r(1-r)R_{\text{bbi}}]^2 << 1 \]  \quad (10.22)

\[ (\omega C_{\text{bc}}R_{\text{bc}})^2 >> 1 \]  \quad (10.23)

then after some algebraic manipulation, the common denominator of the ratios on the RHS of (10.11) and (10.12) may be written as:

\[ R_{\text{bbi}} + Z_{\text{bcj}} + Z_{\text{bcx}} = \frac{\omega^2 C_{\text{bc}}^2 R_{\text{bcj}} R_{\text{bcx}}}{(1 + j\omega C_{\text{bcj}}R_{\text{bcj}})(1 + j\omega C_{\text{bcx}}R_{\text{bcx}})(1/R_{\text{bc}} - \omega^2 C_{\text{bcj}} C_{\text{bcx}} R_{\text{bbi}} - j\omega C_{\text{bc}})} \]  \quad (10.24)

Consequently, one can use (10.24) to arrive at the following simplified equations:

\[ Z_{11} - Z_{12} = R_{\text{bbx}} + rR_{\text{bbi}} + j\omega \left[ L_b - r^2 (1-r)R_{\text{bbi}}^2 C_{\text{bc}} \right] + \frac{jR_{\text{bbi}}}{\omega C_{\text{bc}}^2} \left( \frac{r}{R_{\text{bcx}}} - \frac{1-r}{R_{\text{bcj}}} \right) \]  \quad (10.25)

\[ Z_{22} - Z_{21} = R_{\text{cc}} - r(1-r)R_{\text{bbi}} + \frac{1}{\omega^2 C_{\text{bc}}^2 R_{\text{bc}}} + j\omega L_c - \frac{j}{\omega C_{\text{bc}}} \]  \quad (10.26)

It is worth pointing out that assumptions (10.22) and (10.23) are "second-order" approximations, as opposed to first-order approximations suggested in [442] for the
intermediate frequency range (e.g., $\omega C_{bc} R_{bbi}, \omega C_{bcx} R_{bbi} << 1$). This makes the range for applicability of (10.21)-(10.26) wider. If, for instance, "<<" means "at least 10 times smaller", then with some rather conservative values of $R_{bc} = 10k\Omega$, $R_{bbi} = 20\Omega$, $r = 0.5$, and $C_{bc} = 100 - 2000fF$, the above approximations would be valid for $0.5GHz \leq f \leq 50GHz$.

Therefore, even in InP/InGaAs HBTs where $R_{yyi}$ is 2-3 times larger than GaAs-based HBTs, there would be a wide enough frequency range over which (10.21)-(10.26) are valid and small-signal parameters can be extracted using the technique discussed in this thesis. Another result of the above discussion is that extremely low frequency range (characterised by $\omega C_{yg} R_{yg} \ll 1$) and extremely high frequency range (characterised by $\omega C_{yg} R_{yy} \gg 1$) as defined in [442] require measurement frequencies as low as 50MHz or as high as 500GHz, which cannot be achieved using presently available network analysers.

Other useful relations that can be directly (without any assumptions) derived from (10.11) and (10.12) are:

$$\alpha = \frac{Z_{12} - Z_{21}}{Z_{22} - Z_{21} - Z_{C}} \quad (10.27)$$

$$Z_E = Z_{12} - \frac{(1 - \alpha) R_{bbi} Z_{bcx} Z_{bc} + Z_{bc} + Z_{bcx}}{R_{bbi} + Z_{bc} + Z_{bcx}} = Z_{12} - (1 - \alpha)(Z_{11} - Z_{12} - Z_B) \frac{Z_{bc}}{Z_{bcx}}$$

$$= \frac{Z_{12} - (Z_{22} - Z_{12} - Z_C)(Z_{11} - Z_{12} - Z_B) Z_{bc}}{Z_{22} - Z_{21} - Z_C} \frac{Z_{bc}}{Z_{bcx}} \quad (10.28)$$

Therefore, if $Z_C$, $Z_B$, and $(Z_{bc}/Z_{bcx})$ are known, $Z_E$ can be accurately determined. However, as will be shown later in this chapter, elements of $Z_E$ can be determined at low frequencies without accurate knowledge of the above impedance blocks.

Under forward active mode of operation, and especially at high current regime, the assumption:

$$\left(\omega C_{bc} R_{bc}\right)^2 << 1 \quad (10.29)$$

would be valid in a wide frequency range and (10.15) can be approximated as:

$$Z_E = R_{ee} + R_{be} + j\omega(L_e - R_{be}^2 C_{be}) \quad (10.30)$$

The above formulae and approximations will be used in the next few subsections to extract the small-signal parameters of an InGaP/GaAs DHBT with B-E and B-C areas of $(3\times10)x4\mu m^2$ and $32\times10\mu m^2$, respectively (accounting for the estimated mesa undercuts), and with the layer structure as shown in Table 10.1. As mentioned previously, extraction of parameters starts from de-embedding the internal device from parasitic pad capacitances surrounding it. Then, a series of s-parameter measurements is carried out under constant $V_{CB}$ and variable collector current (including $I_C = 0$, i.e., the cold-HBT). The cold-HBT measurement is to assist in finding $R_{bbx}$, which is assumed bias independent, as well as $C_{Pce}$.
Measured data under variable $I_C$ will be used to separate current dependent elements from those insensitive to current. These are the subjects of next few subsections.

10.3.3. Cold-HBT Measurement

The idea of cold-HBT measurement is not a new one [414,437,441-442,446,453,461], but in this work some additional information are extracted from the device under cold condition which help us find some of the extrinsic series resistances of the device under hot condition.

As mentioned earlier, under cold condition $\alpha = 0$, $Z_{12} = Z_{21}$, and $C_{bc} \approx C_{jbe} \cdot (10.21)-(10.26)$ would still be valid. However, $R_{be}$ is very large and (10.29)-(10.30) are no longer valid. Instead, assuming $(\omega C_{jbe} R_{be})^2 >> 1$ and (10.21)-(10.23), one can write:

$$Z_{12}({\text{cold}}) = Z_{21}({\text{cold}}) = R_{ee} + (1-r)R_{bbi} + \frac{1}{\omega^2 C_{jbe}^2 R_{be}} + j\omega \left[\frac{1}{1-r} R_y (1-r)^2 R_{bbi} C_{bc}\right]$$

It will be shown later (Eq. (10.32b)) that $(1-r)/R_{bci} \approx r/R_{bcx}$, and therefore, the last terms on the RHS of (10.25) and (10.31) are extremely small and can be ignored. Additionally, one can assume $r = \frac{B/E \text{ area}}{B/C \text{ area}}$ under low current injection condition; $r = C_{bci}/C_{bc}$ may vary at higher levels of current as discussed in the forthcoming sections.

![Figure 10.13- Real parts of the z-parameters as a function of frequency obtained under cold-HBT condition.](image)

![Figure 10.14- A plot of $\omega \cdot \text{Im}(Z_{ij})$ versus $\omega^2$ under cold-HBT condition.](image)

As seen from (10.25), (10.26), and (10.31), at high frequencies under cold condition the real parts of $(Z_{11} - Z_{12})$, $(Z_{22} - Z_{21})$, and $Z_{12}$ saturate at $[R_{bbx} + rR_{bbi}]$, $[R_{cc} - r(1-r)R_{bbi}]$, and $[R_{ee} + (1-r)R_{bbi}]$, respectively (see Fig. 10.13). In case any of the real parts is not completely saturating at high frequencies, one can plot it versus $1/\omega^2$ then fit a straight line.
through the data points and extrapolate to the y-intercept. Since \( r \) at low currents is equal to the ratio between the areas of the B-E and B-C junctions, one obtains a system of three equations with four unknowns, namely \( R_{bbi} \), \( R_{bbx} \), \( R_{ee} \), and \( R_{cc} \). The approach employed in this thesis to find these four elements is to assume a reasonable value for \( R_{ee} \), which, for instance, can be obtained from DC open-collector measurement [405]. It is important to mention that the above value of \( R_{ee} \) only serves as an initial guess and it will be corrected in one of the early stages of parameter extraction for hot-HBT, after which only one or maximum two iterations will result in converging values of all the series resistive elements. Once \( R_{ee} \) is known, the other three resistances can be evaluated, but only \( R_{bbx} \) will be assumed constant and fed into the parameter extraction procedure for hot-HBT.

Although the information obtained from imaginary parts of the z-parameters measured under cold condition is not required for the parameter extraction of hot-HBT, it is constructive to show the variation of \( \text{Im}(Z_{ij}) \) to verify the assumptions used to derive (10.25), (10.26), and (10.31). Fig. 10.14 shows a plot of \( \omega \cdot \text{Im}(Z_{11} - Z_{12}) \), \( \omega \cdot \text{Im}(Z_{22} - Z_{21}) \), and \( \omega \cdot \text{Im}(Z_{12}) \) versus \( \omega^2 \). Linear variation of these plots confirms, once again, the validity of previously mentioned approximations. The y-intercept of the plots are \( -1/C_{bc} \), and \(-1/C_{jbe} \), respectively. A zero y-intercept for the plot of \( \omega \cdot \text{Im}(Z_{22} - Z_{21}) \) versus \( \omega^2 \) supports the earlier statement that the last terms in (10.25) and (10.31) are indeed very small. Deviation of \( \omega \cdot \text{Im}(Z_{12}) \) from linearity at low frequencies is due to the violation of the assumption \((\omega C_{jbe} R_{bc})^2 \gg 1 \) at very low frequencies. But as far as extraction of parameters is concerned, there is a wide enough linear region where all the above mentioned parameters can be extracted. The B-E and B-C junction resistances can also be obtained from the cold-HBT data. But since they are not needed for the extraction of HBT parameters under forward active mode of operation, these are not discussed any further here.

### 10.3.4. Hot-HBT Measurement

Next the s-parameters of the device under forward active mode with variable \( I_C \) and constant \( V_{CB} = 0.8 \text{V} \) in this case) will be measured. The reason for having a constant \( V_{CB} \) rather than a constant \( V_{CE} \) will be clarified later. In this analysis we will have the following additional assumptions which all seem physically justifiable: (i) \( R_{bbx} \) is assumed bias independent, (ii) \( R_{ee} \) is constant at low current levels, but it may change at higher currents, and (iii) \( r = \frac{B/E \text{ area}}{B/C \text{ area}} \) only under low current injection condition; it may change at higher levels of current. It is quite clear that the above assumptions are much less restrictive than those used in other published works, e.g., [441]. In [441] it is assumed that all the extrinsic series elements \( (L_e, L_b, L_c, R_{ee}, R_{cc}, \text{ and } R_{bbx}) \) and \( C_{bcx} \) are absolutely bias independent, while this is not required in the present work.

Bearing in mind that (10.25) and (10.26) are still valid under hot-HBT condition, one can obtain the values of \([R_{bbx} + rR_{bbi}], [R_{cc} - r(1-r)R_{bbi}], [L_b - r^2(1-r)R_{bbi}^2C_{bc}], C_{bc}, \text{ and} \)
Lc by plotting real parts of \((Z_{11} - Z_{12})\) and \((Z_{22} - Z_{21})\) versus \(f\), and \(\omega \cdot \text{Im}(Z_{11} - Z_{12})\) and \(\omega \cdot \text{Im}(Z_{22} - Z_{21})\) versus \(\omega^2\). The plots are similar to those in Figs. 10.13 and 10.14, and therefore, are not repeated here. Since \(R_{bbx}\) is known from cold results, \(R_{bbl}\) can be determined and consequently \(R_{ce}\) and \(L_{b}\) are evaluated. \(C_{bc}\) and \(L_{c}\) are obtained independently from y-intercept and gradient of the plot of \(\omega \cdot \text{Im}(Z_{22} - Z_{21})\) versus \(\omega^2\), respectively. This way of plotting the data is preferable to that suggested in other works, e.g., [441]. In [441], after \(C_{bc}\) is obtained, \(L_{c}\) is evaluated from \(\omega^{-1} \cdot [\text{Im}(Z_{22} - Z_{21}) + 1/(\omega C_{bc})]\) and it is stated that this parameter is very much sensitive to the value of \(C_{bc}\), while clearly this problem does not exist in the present work.

If \(\text{Re}(Z_{22} - Z_{21})\) is plotted against \(1/\omega^2\), the gradient of the fitted line will be equal to \(1/(R_{bc}C_{bc}^2)\) (Fig. 10.15). Since \(C_{bc}\) is known, \(R_{bc}\) can be determined. As determined from Fig. 10.15, \(R_{bc}\) increases with a larger reverse bias across the B-C junction. If \(R_{bc}\) is extremely large, then the term \(1/(\omega^2 C_{bc}^2 R_{bc})\) would be negligibly small and almost comparable to the terms already neglected in the derivation of (10.26). Therefore, the plot of \(\text{Re}(Z_{22} - Z_{21})\) versus \(1/\omega^2\) may result in a non-linear variation or even a negative slope. In such circumstances, one can use a large value for \(R_{bc}\), bearing in mind that this element does not affect the small-signal parameters significantly. As to the determination of \(R_{bci}\) and \(R_{bcx}\), one can consider the inverse dependence of resistance on area:

\[
\frac{Z_{bci}}{Z_{bcx}} = \frac{R_{bci}}{R_{bcx}} = \frac{(B/C \text{ area}) - (B/E \text{ area})}{B/E \text{ area}} = \frac{1-r}{r}
\]

\[
R_{bc} = rR_{bci} = (1-r)R_{bcx}
\]

At this stage all of the terms on the RHS of (10.27) and (10.28) are known, and therefore, \(\alpha\) and \(Z_e\) can be evaluated. Variation of \(\alpha\) with frequency will be discussed in the next section. As to the determination of \(Z_e\), magnitude of \((1-\alpha)\) will become large at high frequencies. Consequently, any small error in the determination of \(Z_B\), \(Z_C\), or \((Z_{bci}/Z_{bcx})\) may result in a significant deviation of \(Z_e\) at higher frequencies. But the effect of the above impedance blocks on \(Z_e\) is just minimal at lower range of frequency where \((1-\alpha) = 0\). Therefore, \((R_{ce} + R_{bc})\) can be evaluated from the real part of RHS of (10.28) at low frequencies. Fig. 10.16 shows the variation of \(\text{Re}(Z_E)\) with frequency for different values of DC collector current. All of the
plots in Fig. 10.16 saturate at low frequencies at \((R_{ee} + R_{be})\), and at higher frequencies asymptotically approach \(R_{ee}\), as expected from the formulation of \(Z_E\) in equation (10.15) [414]. Determined values of \((R_{ee} + R_{be})\) can be plotted against \((1/I_c)\) to differentiate between \(R_{ee}\) and \(R_{be}\) (Fig. 10.17). The \(y\)-intercept of this plot gives a corrected \(R_{ee}\), which has to be used in order to obtain a corrected value of \(R_{bbx}\) through the cold-HBT measured resistances. \(R_{be} = \eta kT/qI_c\) can also be found from the gradient of the plot in Fig. 10.17, which gives an ideality factor, \(\eta\), of 1.03. The sudden increase of \((R_{ee} + R_{be})\) at the highest current is due to device self-heating which is known to increase both \(R_{ee}\) and \(R_{be}\) [326,414].

![Figure 10.16- Plots of Real\((Z_E)\) versus frequency under variable \(I_C\). All the curves saturate at low frequencies to the value of \(R_{ee} + R_{be}\).](image)

![Figure 10.17- Variation of \(R_{ee} + R_{be}\) as determined from Fig. 10.16, with \((1/I_C)\). Extrapolated \(y\)-intercept of the fitting line gives \(R_{ee}\).](image)

Although any change in \(R_{ee}\) will be directly reflected to a change in \(R_{bbx}, R_{bbi,}\) and \(R_{cc}\) with more or less similar magnitude, the resultant variation of the latter parameters only has a minor effect on \(R_{ee}\) and \(R_{be}\), which are determined in the low frequency region where \((1 - \alpha) \approx 0\). Therefore, the above procedure will be a very fast converging iteration with only one or two steps required. It is worth mentioning that in [441] \(R_{bbi}, R_{bbx},\) and \(R_{cc}\) are determined by numerical optimisation and/or using assumptions related to extremely high frequencies, while fully analytical methods in measurable range of frequency are employed in the present work.

Once the iteration procedure is converged, one can plot the imaginary part of \(Z_E\) versus \(\omega\) to find \([L_e - R_{be}^2 C_{be}]\) and thus \(L_e\). Imaginary part of \(Z_E\), obtained from the RHS of (10.28), is very much sensitive to the value of \(r\). Therefore, an accurate value of \(r\) considering the undercutting of B-E and B-C junctions is crucial in determining \(L_e\), otherwise one would observe a non-physical saturating behaviour for \(\text{Im}(Z_E)\). \([L_e - R_{be}^2 C_{be}]\) is an increasing function of \(I_C\) (see Fig. 10.18), but once corrected for the variation of \(R_{be}^2 C_{be}\), shows an almost constant value of \(L_e\) versus collector current. Values of all the extracted parameters for the device under study with variable \(I_C\) and constant \(V_{CB} = 0.8V\) are summarised in Table
10.3. Extraction of delay times $\tau_B$, $\tau_C$, $\tau_{EC}$, and capacitances $C_{jbe}$, $C_{be}$ listed in Table 10.3 will be discussed in section 10.4.

![Graph showing Im(Z_e) versus $\omega$ under various collector current levels. The gradient of the fitting line is equal to $L_e - R_{be}^2 C_{be}$.](image)

**Figure 10.18** - Im($Z_e$) versus $\omega$ under various collector current levels. The gradient of the fitting line is equal to $L_e - R_{be}^2 C_{be}$.

### 10.3.5. Comparison between the Measured and Calculated Data

Section 10.3 and the forthcoming section 10.4 present a completely analytical HBT parameter extraction technique, which has been successfully applied to a (3×10)$^2 \times 4$ μm$^2$ InGaP/GaAs DHBT. All the extracted parameters under variable collector current and constant $V_{CB}$ of 0.8V are summarised in Table 10.3. Fig. 10.19 compares the measured s-parameters with those calculated (by SPICE) using extracted elements at collector currents of 2.3 and 25mA. Excellent agreement between the measured and SPICE-calculated data can be observed in the entire range of frequency and for both high and low current injection levels. It is known, however, that polar plots of s-parameters do not perfectly reflect the quality of agreement between measured and calculated data. Also the high frequency portion ($f > 10$GHz) of the s-parameters in a polar plot is compressed in a small area of the plot. Therefore, the real and imaginary parts of the measured and calculated $z$-parameters are shown for $I_C = 9$mA in Figs. 10.20(a) and (b), respectively. The calculated $z$-parameters also show a very good fit to the measured ones. The quality of the fit is better than all other reports to date.

Excellent agreement between the measured and calculated parameters eliminates the necessity for a final optimisation step. Indeed, to prove the latter an optimisation process is carried out using HP-ADS optimisation facility with the extracted parameters as initial guess. The values of the elements after optimisation for $I_C = 9$mA are also shown in Table 10.3. An average mixed relative and absolute error is defined as in [441]:

\[
\text{Relative Error} = \frac{|E_{measured} - E_{calculated}|}{E_{measured}} \\
\text{Absolute Error} = |E_{measured} - E_{calculated}|
\]
Table 10.3 - Directly extracted and optimised parameters under variable collector current and constant $V_{CB}$ of 0.80V for an InGaP/GaAs DHBT with B-E area of (3×10)$\times$4 $\mu$m$^2$ as specified in Table 10.1.

The relative errors of calculation, as defined in [441], are also given.

The excellent accuracy of the presented method can be attributed to physically accurate...
expressions used for elements of the equivalent circuit, and very reasonable assumptions made during the extraction procedure. Consequently, this approach can be used as a standard technique to extract the equivalent circuit elements of various types of HBTs, including InP- and GaAs-based single and double HBTs.

Figure 10.19- Comparison between the measured (symbols) and simulated (solid lines) s-parameters for the device under study at $V_{CB} = 0.80V$. Blue lines/symbols (open triangles) and red lines/symbols (open squares) correspond to $I_C = 2.3$ and 25mA, respectively.

10.4. High Frequency Figures of Merit ($f_T$ and $f_{\text{max}}$)

Cutoff frequency, $f_T$, and maximum frequency of oscillation, $f_{\text{max}}$, are defined as the frequencies where current gain $h_{21}$ and maximum available power gain (MAG), respectively, reduce to unity. Expression relating $h_{21}$ and MAG to the $z$-parameters of a two-port network are given in Appendix A. Figure 10.21 shows a plot of $h_{21}$, MAG, maximum stable gain (MSG), and Mason’s Unilateral gain (U) [463], all defined in Appendix A, as a function of frequency for the device considered in previous section under $V_{CB} = 0.80V$ and $I_C = 25mA$ conditions.
Chapter 10

High Frequency Characteristics of HBTs

50

100

Ic = 9mA
V_{CB} = 0.8V

Re(Z21)

Re(Z22)

Re(Z11)

Re(Z12)

Frequency [GHz]

Figure 10.20- A comparison between the calculated (solid lines) and measured (symbols): (a) real and (b) imaginary parts of the z-parameters as a function of frequency under $I_c = 9 \text{mA}$ and $V_{CB} = 0.80 \text{V}$ condition. A different $I_c$ level, as compared to Fig. 10.19, is used to demonstrate the applicability of the present approach for a wide range of bias.

Figure 10.21- Variation of current gain, $h_{21}$, power gains $U$, MSG, MAG, and the Rollett stability factor, $K$, with frequency for the device studied in the previous section under $I_c = 25 \text{mA}$ and $V_{CB} = 0.80 \text{V}$ condition. The medium frequency fit to $h_{21}$ is also shown with a dashed line.

As seen in Fig. 10.21 and equations (A.8)-(A.10) in the Appendix A, when the Rollett stability factor ($K$) is less than 1, the HBT network is potentially unstable. In such circumstances, maximum stable gain would be the useful gain, and MAG is identical to MSG. It can be
shown that MSG always follows a roll-off very close to -10dB/dec (i.e., proportional to 1/ω) [406]. At higher frequencies, HBT becomes unconditionally stable (K > 1) and MAG deviates from MSG. After passing through a transition region, the slope of MAG approaches -20dB/dec (i.e., proportional to 1/ω^2) [406], and finally MAG reaches unity (0dB) gain at fmax. Mason’s unilateral gain, on the other hand, shows a much more uniform roll-off with a gradient close to -20dB/dec. But, the interesting fact is that U reaches 0dB gain at exactly the same frequency as MAG does. Indeed it has been shown in the literature that [375]:

\[ U(f_{\text{max}}) = \text{MAG}(f_{\text{max}}) = 1 \]  \hspace{1cm} (10.34)

always holds true. For the device shown in Fig. 10.21, the value of fmax is within the measurement range and can be directly observed. However, for the state-of-the-art devices with fmax > 800GHz [2], maximum frequency of oscillation can not be reached directly. In such cases, either U or MAG can be extrapolated with a slope of -20dB/dec to find fmax. But, since U is much more well behaved than MAG, it is preferable to use U for evaluating fmax.

The trends explained above for the power gains and stability factor are observed for all of the HBTs studied in this work and all others reported in the literature. The cutoff frequency of the device shown in Fig. 10.21 is ~40GHz, as defined from frequency where h21 becomes unity. In the medium range of frequency, 20Log(h21) follows an almost -20dB/dec roll-off (see the discussion in subsection 10.4.2). Extrapolation of this single-pole behaviour (the dashed fitting line in Fig. 10.21) will result in fT = 26GHz, which is significantly different from the one obtained from where h21 is equal to unity. Deviation of h21 from the single-pole, -20dB/dec roll-off behaviour at higher frequencies can be due to one or a combination of the following:

- importance of higher order poles and zeros in the transfer function
- the transit time effect [449]
- frequency dispersion related to extrinsic base surface recombination [464]

Since the state-of-the-art HBTs have cutoff frequencies in excess of the measurable range of frequency for the present network analysers, fT is conventionally obtained using extrapolation of the medium frequency range h21 with -20dB/dec roll-off. Additionally, the total delay time, \( \tau_{\text{EC}} \), of an HBT is related to fT through \( \tau_{\text{EC}} = \frac{1}{2\pi f_{\text{T}}} \). But, as shown in subsection 10.4.2, this expression is based on a single-pole approximation for h21, which only holds for the low to medium range of frequency. Therefore, in this thesis the frequency where extrapolated h21 with -20dB/dec meets the 0dB gain, i.e., ~26GHz in Fig. 10.21, is referred to as fT. But the latter still imposes another problem. The gradient of the fit to the medium frequency h21 is almost always different from -20dB/dec. For the devices measured in this work, the gradient varies in the range (18.5-19.5)dB/dec, depending on the device, fitting range of frequency, and biasing condition. For the graph in Fig. 10.21 this is equal to -19.2dB/dec. Therefore, this fitting line will not exactly follow the single-pole approximation of h21, and hence, \( \tau_{\text{EC}} \).
obtained from this fitting is not accurate. Even if one decides to pass a line with constant slope of -20dB/dec through one point of the graph, the question remains as to which point the line should pass through. Therefore, the value of \( \tau_{EC} \) thus found is not unique.

In subsection 10.4.2, a novel technique to evaluate a unique value for \( \tau_{EC} \) is suggested which only relies on measured values of z-parameters at low frequencies, and therefore, avoids the above mentioned ambiguities. This comes following a discussion on forward transit time, \( \tau_F = \tau_B + \tau_C \), and its components in subsection 10.4.1.

### 10.4.1. Forward Transit Time

The forward transit time, \( \tau_F \), is the sum of the base transit time, \( \tau_B \), and the B-C depletion region delay time, \( \tau_C \). This delay time only depends on the vertical structure of the HBT, and is not affected by the lateral parasitics of the device. Therefore, it is important to evaluate \( \tau_F \), since it can be treated as the minimum bound for the delay time achievable by a known epitaxial structure.

In the previous section, direct extraction of the parameter \( \alpha(\omega) \) was explained using equation (10.27). Here, it will be shown that the frequency and bias dependence of this parameter includes sufficient information to extract \( \alpha_0 \), \( C_{jbe} \), \( \tau_B \), and \( \tau_C \). Using (10.3) one can write:

$$ \frac{1}{|\alpha|^2} = \frac{(\omega \tau_C)^2}{\alpha_0^2 \sin^2(\omega \tau_C)} \left[ 1 + \omega^2(m\tau_B + R_{be}C_{jbe})^2 \right] $$

Then, expanding the Taylor series of \( \sin(\omega \tau_C) \) and ignoring the \( \omega^4 \) terms (and higher powers of \( \omega \)), the following equation can be derived:

$$ \frac{1}{|\alpha|^2} \approx \frac{1 + \omega^2[(m\tau_B + R_{be}C_{jbe})^2 + \tau_C^2 / 3]}{\alpha_0^2} \quad (10.35) $$

Therefore, \( \alpha_0 \) and the term inside the square bracket in (10.35) can be extracted from the y-intercept and gradient of \( 1/|\alpha|^2 \) versus \( \omega^2 \), respectively. This plot is shown in Fig. 10.22 for the device studied in section 10.3 at various current levels. A linear behaviour can be observed in this plot for the low to medium frequency range.

If one further assumes:

$$ \omega^2(m\tau_B + R_{be}C_{jbe})^2 \ll 1 \quad (10.36a) $$

$$ \omega^2[(1 - m)\tau_B + \tau_C]^2 \ll 1 \quad (10.36b) $$

then Eq. (10.3) can be rewritten as:

$$ \alpha = \frac{\alpha_0 \cdot \sin(\omega \tau_C)}{\omega \tau_C} \left[ 1 - j\omega(\tau_B + \tau_C + R_{be}C_{jbe}) \right] \quad (10.37a) $$
\[
\frac{1}{\alpha} = \left( \frac{\omega\tau_C}{\alpha_0 \cdot \sin(\omega\tau_C)} \right) \left[ 1 + j\omega(\tau_B + \tau_C + R_{\text{be}}C_{\text{jbe}}) \right]
\]  
(10.37b)

**Figure 10.22** - Variation of \(1/|\alpha|^2\) with \(\omega^2\), and \(\text{Im}(\alpha)/\text{Re}(\alpha)\) with \(\omega\) under various \(I_C\) levels.

Therefore, \((\tau_B + \tau_C + R_{\text{be}}C_{\text{jbe}})\) can be obtained from a plot of \(\frac{\text{Im}(1/\alpha)}{\text{Re}(1/\alpha)}\) versus \(\omega\) (Fig. 10.22). The values of \(\tau_B + \tau_C + R_{\text{be}}C_{\text{jbe}}\) at various current levels are plotted against \(1/I_C\) in Fig. 10.23. Since \(R_{\text{be}}\) is already known as a function \(1/I_C\), \(C_{\text{jbe}}\) can be determined from the gradient of the above plot. This method assumes that \(C_{\text{jbe}}\) does not vary significantly with slight changes of \(V_{\text{BE}}\) in the current range considered. Linearity of this plot confirms that the latter assumption is a reasonable one. The obtained value of \(C_{\text{jbe}}\) is \(\sim 218\,\text{fF}\). The forward transit time can also be accurately evaluated from the y-intercept of the above plot. This method of characterising the forward transit time is only relying on an accurate value of \(R_{\text{cc}}\), which is to be used in equation (10.27) to find \(\alpha(\omega)\); collector series inductance, as shown in the following, has a negligible effect on \(\alpha\). Therefore, the present method is expected to be much more reliable than the conventional method of plotting total delay time, \(\tau_{\text{EC}}\), versus \(1/I_C\), which additionally requires a prior knowledge of \(R_{\text{ce}}\) and \(C_{\text{bc}}\) (see Eq. (10.49) and Fig. 10.23). Also one should bear in mind the difficulties in accurate and unique evaluation of \(\tau_{\text{EC}}\), as discussed previously, which make the conventional method even more unreliable. Furthermore, it is shown by Lee [453] that the new method of determining \(\tau_F\) is much less sensitive to errors in de-embedding pad capacitances. However, the method suggested by Lee [452-453] is slightly different from the one used in this thesis. Lee has suggested to use:
Figure 10.23- Variation of $(\tau_F + R_{be}C_{jbe})$ (obtained from Fig. 10.22), $(\tau_F + R_{be}C_{jbe} + R_{cc}C_{be})$ (see the text), and $\tau_{EC}$ (determined from Fig. 10.24) with $1/I_c$. Extrapolated y-intercepts of the fitting lines give $\tau_F$, $(\tau_F + R_{cc}C_{be})$, and $[\tau_F + (R_{ee} + R_{cc})C_{be}]$, respectively.

\[ \text{Im}\left(\frac{Z_{22} - Z_{12} - Z_C}{Z_{12} - Z_{21}}\right) = \text{Im}\left(\frac{1-a}{a}\right) = \text{Im}\left(\frac{1}{a}\right) = \tau_B + \tau_C + R_{be}C_{jbe} \]

while the last approximate equality in the above expressions clearly ignores the term $\alpha_0 \cdot \sin(\omega \tau_C)/\omega \tau_C$ in (10.37). Therefore, the present method is expected to give more accurate results.

In the above discussion, it was made clear that the new method of extracting $\tau_F$ only relies on an accurate value of $Z_C$ to be subtracted from $Z_{22}$. In what follows, it will be shown that only an accurate value of $R_{cc}$ is required for precise determination of $\tau_F$. First, one can define $\alpha_{Z_C}$ as:

\[ \alpha_{Z_C} = \frac{Z_{12} - Z_{21}}{Z_{22} - Z_{21}} \]  

(10.38)

then from (10.12c) and (10.27) it follows:

\[ \alpha_{Z_C} = \frac{\alpha}{1 + \frac{\alpha \cdot Z_C}{Z_{12} - Z_{21}}} = \frac{\alpha}{1 + \frac{Z_C (R_{bbi} + Z_{bci} + Z_{bcx})}{Z_{bci} \cdot Z_{bcx}}} \]

and using assumptions (10.21)-(10.23), $\alpha_{Z_C}$ simplifies to:

\[ \alpha_{Z_C} = \frac{\alpha}{1 + (R_{cc} + j\omega L_c) \cdot (1/R_{bc} - \omega^2 C_{bc}^2 r(1-r)R_{bbi} + j\omega C_{bc})} \]
If one further assumes that $R_{cc}$ has the same order of magnitude as $R_{bbi}$, which means that (10.21)-(10.23) would still be valid when $R_{bbi}$ is replaced by $R_{cc}$, and additionally assumes:

$$\omega^2 L_c C_{bc} << 1$$  \hspace{1cm} (10.39)

then it is straightforward to show that:

$$\alpha_{Z_c} = \frac{\alpha}{1 + j\omega (C_{bc} R_{cc} + L_c / R_{bc})}$$ \hspace{1cm} (10.40)

For the values of parameters listed in Table 10.3, the term $(C_{bc} R_{cc})$ is at least 1300 times larger than $(L_c / R_{bc})$, and therefore, it is safe to ignore the latter term in (10.40) and write:

$$\alpha_{Z_c} = \frac{\alpha}{1 + j\omega C_{bc} R_{cc}}$$ \hspace{1cm} (10.41)

Therefore, it is clear that an accurate value of $L_c$ is not required for determination of $\tau_F$ using the new method discussed above. Also (10.37) and (10.41) result in:

$$\alpha_{Z_c} = \frac{\alpha_0 \cdot \sin(\omega \tau_{C})}{\omega \tau_{C}} \left[1 - j\omega (\tau_F + \tau_C + R_{be} C_{jbe} + R_{cc} C_{bc})\right]$$ \hspace{1cm} (10.42)

which means that if $\text{Im}(\alpha_{Z_c}) / \text{Re}(\alpha_{Z_c})$ is plotted against frequency, a linear graph with the gradient $- (\tau_F + R_{be} C_{jbe} + R_{cc} C_{bc})$ will be obtained. These values determined under various collector currents are also plotted in Fig. 10.23 as a function of $1/L_c$. The gradient of this plot is similar to that of $(\tau_F + R_{be} C_{jbe})$ versus $1/L_c$, but its y-intercept is larger by the value of $R_{cc} C_{bc}$.

After finding $C_{jbe}$ from the plot in Fig. 10.23, the plots in Fig. 10.22 for each bias condition give us two equations for the two unknowns $\tau_B$ and $\tau_C$. Therefore, $\tau_B$ and $\tau_C$, and consequently $C_{be}$ (from (10.2)) can be evaluated. At this stage it must be pointed out that in [447], the term $\sin(\omega \tau_{C})/(\omega \tau_{C})$ is neglected in the formulation of $\alpha(\omega)$, and it is suggested to use a plot of $1/|\alpha|^2$ versus $\omega^2$ to directly find $R_{be} C_{be}$. This is equivalent to ignoring the term $\tau_C^2 / 3$ in (10.35), and if used, would result in an unphysical variation of extracted $C_{be}$'s with $V_{CB}$. Also Li and Prasad [441-442] used the invalid assumption of $C_{be} \propto I_C$ (which ignores $C_{jbe}$) and a plot of $[L_c - R_{be}^2 C_{be}]$, obtained from Fig. 10.18, versus $1/L_c$ to find $C_{be}$. Therefore, they obtained a non-linear plot (Fig. 13 in [441]) and an inaccurate value of $C_{be}$ which necessitated numerical optimisation. But even optimisation is relatively insensitive to $C_{be}$, and hence, $C_{be}$ could not be determined accurately in their work. Obviously, the above problems do not exist in the present method. Additionally, the formula for $\tau_B$ in some of the earlier works (e.g., [441]) is incorrectly stated as $\tau_B = R_{be} C_{be} (= 1/\omega_\alpha)$. Based on this, the authors observed values of $\tau_B$ which were unexpectedly varying with both $I_B$ and $V_{CE}$ (see Tables II and III and Fig. 15 in [441]). In contrast, $\tau_B$'s obtained in the present work are almost constant with bias ($1.61 \pm 0.08$ ps), since a physically correct formula for $C_{be}$ (equation
(10.2)) is used here.

### 10.4.2. Cutoff Frequency and Total Delay Time

For the HBT equivalent circuit of Fig. 10.4, after de-embedding the pad capacitances, $h_{21}$ can be expressed as (see Appendix A for the relationship between $h$- and $z$-parameters):

$$h_{21} = \frac{-Z_{21}}{Z_{22}} = \frac{\alpha Z_{bc1} Z_{bcx} - (1 - \alpha) R_{bbi} Z_{bc1} - Z_E (R_{bbi} + Z_{bc1} + Z_{bcx})}{(1 - \alpha) Z_{bc1} (R_{bbi} + Z_{bcx}) + (Z_E + Z_C) (R_{bbi} + Z_{bc1} + Z_{bcx})} \tag{10.43}$$

In the following, some “first-order” approximations are used to simplify the above equations and establish the relationship between the total delay time and the cutoff frequency. Obviously, these first-order approximations do not limit the range of applicability of the parameter extraction technique recommended in this chapter, since the extraction of parameters is already completed with only the second-order approximations discussed in the previous sections. To start we assume:

$$\omega C_{bc} (1 - r) R_{bbi} << 1 << \omega C_{bc} R_{bc} \tag{10.44}$$

Then, it follows that $R_{bbi} + Z_{bcx} = Z_{bcx} = 1/j \omega C_{bcx}$ and $Z_{bc1} = 1/j \omega C_{bc1}$. Therefore, (10.43) can be simplified to:

$$h_{21} \approx \frac{\alpha - j \omega C_{bc} [Z_E + (1 - r) R_{bbi}]}{(1 - \alpha) + j \omega C_{bc} (Z_E + Z_C)}$$

Next, $Z_E$ and $Z_C$ are replaced by their appropriate relations from (10.30) and (10.17), and all the terms including $\omega^2$ and higher powers of $\omega$ are neglected. These include the terms such as $\omega^2 C_{bc} L_e$ and $\omega^2 C_{bc} L_c$, and additionally means that $\sin(\omega \tau_C) = \omega \tau_C$. Hence, bearing in mind that $\alpha_0 = 1$, it follows from (10.37a) that:

$$\alpha = \alpha_0 - j \omega (\tau_B + \tau_C + R_{bc} C_{jbe}) \tag{10.45}$$

and $h_{21}$ further simplifies to:

$$h_{21} = \frac{\alpha_0 - j \omega [(\tau_B + \tau_C + R_{bc} C_{jbe})]}{(1 - \alpha_0) + j \omega [(\tau_B + \tau_C + R_{bc} C_{jbe}) + (R_{ee} + R_{bc} + (1 - r) R_{bbi}) C_{bc}]} \tag{10.46}$$

At frequencies much lower than the cutoff frequency, it can be shown that the terms inside the square brackets are much lower than $\alpha_0$ (but not much lower than $1 - \alpha_0$). Therefore, as a final approximation the term inside the square bracket in the numerator can be ignored, and the well known approximation for $h_{21}$ follows:

$$h_{21} \approx \frac{\alpha_0}{(1 - \alpha_0) + j \omega \tau_{EC}} = \frac{\beta_0}{1 + j(1 + \beta_0) f / f_T} \tag{10.47}$$

where:
Chapter 10 __________________________________________________________

High Frequency Characteristics of HBTs

\[ \beta_0 = \alpha_0 / (1 - \alpha_0) \]  

(10.48)

\[ 1/2 \pi f_T = \tau_{EC} = \tau_B + \tau_C + R_{bc}C_{jbc} + (R_{ec} + R_{be} + R_{cc})C_{bc} \]  

(10.49)

Therefore, in the frequency range where the above assumptions are valid, \( |h_{21}| \) is almost proportional to \((1/f)\), and hence, \(20 \log|h_{21}|\) follows a -20dB/dec roll-off. However, at frequencies approaching \( f_T \), the term inside the square bracket in the numerator of (10.46) starts to become significant, and \( |h_{21}| \) flattens (i.e., a zero in the transfer function which cancels the effect of the dominant pole). At even higher frequencies, the \( \omega^2 \) terms and others which were assumed insignificant at lower frequencies start to dominate, and one may even observe the resonance peaking as shown by simulation in [449].

The following conclusions can be drawn out of the above discussion. First, as seen from (10.43), \( R_{bbi} \) and \( L_b \) do not have any effect on \( h_{21} \) (and hence \( f_T \)). Even \( R_{bbi}, L_e, \) and \( L_c \) have second-order effects on \( h_{21} \), and therefore, influence \( f_T \) only marginally. Second, the DC current gain, \( \beta_0 \), and parameters determining the current gain such as recombination lifetimes, surface recombination, emitter-size effect, heterointerface quality, etc. do not have any effect on the cutoff frequency. Therefore, a transistor with a poor DC current gain may have a large \( f_T \). Also enhancing the DC current gain does not degrade the cutoff frequency, as some may think.

Higher order zeros and poles and other effects discussed earlier may cause the gradient of the logarithmic fit to the variation of \( |h_{21}| \) with frequency to deviate significantly from the theoretical -20dB/dec roll-off. Therefore, extrapolation of the variation of \( |h_{21}| \) to 0dB gain may result in erroneous and non-unique values of \( f_T \) and \( \tau_{EC} \). In this thesis, a new method of characterising the total delay time based on low frequency measured data is introduced. In the following, it will be shown that:

\[ \text{Re}(Z_{22}) = \tau_{EC} / C_{bc} \quad \text{at low frequencies} \]  

(10.50)

Since \( C_{bc} \) is one of those elements which can be evaluated quite precisely at low frequency, the above equation serves as a low frequency rule to find an accurate value of total delay time. (10.50) uses some of the first-order assumptions already used to derive (10.46). Therefore, this expression is at least as accurate as the definition of \( f_T \) and \( \tau_{EC} \) in (10.47) and (10.49). To prove (10.50), we start from (10.11d) and use the assumptions \( \omega C_{bc} R_{bbi} >> 1 \) (which means that \( Z_{bc} = 1/j \omega C_{bc} \) and \( Z_{bbe} = 1/j \omega C_{bc} \)) and (10.29) to write:

\[ \text{Re}(Z_{22}) = \text{Re} \left[ \frac{(1-\omega)}{j \omega C_{bc}} \left( \frac{[1+j \omega(1-r)C_{bc}R_{bbi}]}{[1+j \omega(1-r)]C_{bc}R_{bbi}} \right) + R_{cc} + R_{ee} + R_{be} \right] \]

\[ = \text{Re} \left[ \frac{(1-\omega)}{j \omega C_{bc}} \left( \frac{[1+j \omega(1-r)C_{bc}R_{bbi}]}{1+[\omega(1-r)C_{bc}R_{bbi}]^2} \right) \right] + R_{cc} + R_{ee} + R_{be} \]

then assuming \( \omega(1-r)^2 C_{bc} R_{bbi} \ll 1 \), which is less restrictive than (10.44), \( \text{Re}(Z_{22}) \)
Chapter 10

High Frequency Characteristics of HBTs

simplifies to:

\[ \text{Re}(Z_{22}) = \text{Re} \left( \frac{(1 - \alpha)}{j\omega C_{bc}} \right) + R_{cc} + R_{ee} + R_{be} = \frac{-\alpha_i}{\omega C_{bc}} + R_{cc} + R_{ee} + R_{be} \]

where \( \alpha = \alpha_r + j\alpha_i \) is written in terms of its real and imaginary parts. Now if one uses the low frequency relation for \( \alpha(\omega) \) (i.e., (10.45)), then (10.50) follows. The assumptions used to derive (10.50) are valid for almost an order of magnitude of frequency (typically \( 1.5\text{GHz} \leq f \leq 15\text{GHz} \)), and therefore, there exists a wide enough range for \( \tau_{EC} \) to be extracted. Fig. 10.24 shows the variation of \( \text{Re}(Z_{22}) \) with frequency for various \( I_C \)'s. The plot at all different current levels saturate at low frequencies. The constant range of the plots gets narrower for lower current levels, primarily due to the narrower range of validity for assumption (10.29). The values thus calculated are compared in Table 10.3 with those obtained from extrapolation of \( |h_{21}| \) at higher frequencies. The slight difference between the two sets of \( \tau_{EC} \)'s is mainly due to the fact that \( |h_{21}| \) in all of the cases rolls off with gradient less than 20dB/dec (18.5-19.5dB/dec in these cases). Therefore, using \( \text{Re}(Z_{22}) \) is expected to give more accurate delay times. Finally, it should be pointed out that both the above sets of calculated \( \tau_{EC} \)'s are for the device de-embedded from pad capacitances, which do not belong to the actual device anyway.

10.4.3. Unilateral Power Gain

To date, derivation of a simplified formula for \( U \) and \( f_{\text{max}} \) in HBTs is discussed by many authors [375,440,449-450,454-455]. The well-established relationship:

\[ \text{Re}(Z_{22}) = \text{Re} \left( \frac{(1 - \alpha)}{j\omega C_{bc}} \right) + R_{cc} + R_{ee} + R_{be} \]

\[ \text{Re}(Z_{22}) = \text{Re} \left( \frac{-\alpha_i}{\omega C_{bc}} \right) + R_{cc} + R_{ee} + R_{be} \]

**Figure 10.24**- Variation of \( \text{Re}(Z_{22}) \) with frequency under various \( I_C \) levels. The sudden increase in \( \text{Re}(Z_{22}) \) at \( I_C = 40\text{mA} \) is due to the device self-heating (see section 10.5).
is critically discussed in all above works, and majority of them concluded that (10.51) either overestimates or underestimates \( f_{\text{max}} \) of HBTs. In the following a simplified formula for \( U \) and \( f_{\text{max}} \) will be presented based on the T-shaped equivalent circuit of Fig. 10.4 (without the parallel pad capacitances).

The expression for the Mason’s unilateral gain based on z-parameters is given in Appendix A and is repeated here:

\[
U = \frac{|Z_{21} - Z_{12}|^2}{4\{\text{Re}(Z_{11})\text{Re}(Z_{22}) - \text{Re}(Z_{12})\text{Re}(Z_{21})\}} \tag{10.52}
\]

Using (10.29) and the first-order assumptions (10.44) it can be shown that:

\[
\text{Re}(Z_{11}) \approx R_{ee} + R_{bc} + R_{bbi} + R_{bbx} - (1 - r)\alpha_r R_{bbi} \tag{10.53a}
\]

\[
\text{Re}(Z_{22}) \approx R_{ee} + R_{bc} + R_{cc} - \frac{\alpha_i}{\omega C_{bc}} \approx \frac{\tau_{EC}}{C_{bc}} = \frac{1}{\omega_T C_{bc}} \tag{10.53b}
\]

\[
\text{Re}(Z_{12}) \approx R_{ee} + R_{bc} + (1 - r)R_{bbi} (1 - \alpha_r) \tag{10.53c}
\]

\[
\text{Re}(Z_{21}) \approx R_{ee} + R_{bc} + (1 - r)(1 - \alpha_r)R_{bbi} - \frac{\alpha_i}{\omega C_{bc}} \tag{10.53d}
\]

\[
|Z_{21} - Z_{12}|^2 \approx \frac{|\alpha|^2}{\omega^2 C_{bc}^2} \tag{10.53e}
\]

where \( \omega_T = 2\pi f_T = 1/\tau_{EC} \). Now replacing (10.53) into (10.52) and using the low frequency approximation for \( \alpha \) (i.e., (10.45)) with assuming \( \alpha_r = \alpha_0 \approx 1 \), yields the following low frequency expression for \( U \):

\[
U = \frac{|\alpha|^2}{4\omega^2 C_{bc}^2 [R_{Bl} \tau_{EC} / C_{bc} + (R_{ee} + R_{be}) R_{cc}]} \tag{10.54}
\]

where \( R_{Bl} = R_{bbx} + rR_{bbi} \). From the above expression for \( U \), it is clear that the series inductances do not have any effect on unilateral gain and \( f_{\text{max}} \), since the input and output matching circuits can cancel the effects of any series reactances. Further assuming \( \omega^2 (\tau_B + \tau_C + R_{be} C_{jbe})^2 \ll 1 \) at lower frequencies, results in \( |\alpha|^2 \approx 1 \) and:

\[
U = \frac{\omega_T}{4\omega^2 R_{Bl} C_{bc} \left[ 1 + \frac{(R_{ee} + R_{be}) R_{cc} C_{bc} \omega_T}{R_{Bl}} \right]} \tag{10.54}
\]

This means that \( U \) has a -20dB/dec low frequency roll-off (1/\( \omega^2 \) dependence) which can be
extrapolated to higher frequencies to obtain $f_{\text{max}}$ as:

$$f_{\text{max}} = \sqrt{\frac{f_T}{8\pi (R_b C_{bc})_{\text{eff}}}}$$  \hspace{1cm} (10.55a)$$

with the effective time constant $(R_b C_{bc})_{\text{eff}}$ defined as:

$$(R_b C_{bc})_{\text{eff}} = R_b C_{bc} + (R_{ce} + R_{be}) R_{cc} C_{bc}^2 \omega_T$$  \hspace{1cm} (10.55b)$$

(10.55) is similar to the more simplistic expression for $f_{\text{max}}$ given by (10.51), but there exists an additional time constant term on the RHS of (10.55b) that makes it more accurate. For instance, using the data shown in Table 10.3, the second term on the RHS of (10.55b) forms as large as \(\sim 30\%\) of the total $(R_b C_{bc})_{\text{eff}}$ time constant. This may introduce \(\sim 20\%\) error in estimated value of $f_{\text{max}}$ when (10.51) is used. The calculated values of $f_{\text{max}}$ from both (10.51) and (10.55) together with the extrapolated $f_{\text{max}}$ from the measured $U$ are shown in Table 10.3. It is clear that (10.55) gives much more accurate values of $f_{\text{max}}$, and that (10.51) overestimates $f_{\text{max}}$, as also suggested in [455]. Finally, it should be pointed out that (10.55) is equivalent to equation (44) in [454] when B-C network is represented with only two resistances ($R_{bb1}$ and $R_{bbx}$) and two capacitances ($C_{bc1}$ and $C_{bcx}$).

10.5. Bias and Geometry Dependence

In this section, the extraction method described in previous sections is employed to study the bias and geometry dependence of some of the physical parameters.

10.5.1. Bias Dependence of the Parameters

Figure 10.25 shows the variation of the extracted total B-C junction capacitances in Table 10.3 versus collector current. Also shown are the measured capacitances under constant $V_{\text{CB}}$ of 0.5 and 1.2V, and constant $V_{\text{CE}}$ of 2.1V. It can be observed that $C_{bc}$ reduces with increasing $I_C$ under constant $V_{\text{CB}}$ condition, similar to the trends observed in [131] and [465]. The reduction of $C_{bc}$ with current is attributed to the following effects:

(i) Current-induced broadening of the B-C depletion layer: As the collector current increases, more and more electrons are injected to the B-C

![Figure 10.25](image-url)
depletion region. This reduces the total space-charge density \((N_D - n)\) inside this region. Since the total space-charge inside the B-C depletion region is almost fixed with the reverse bias across this junction, the depletion region should get broader in order to compensate the reduction of the space-charge density. A wider depletion region translates into a smaller depletion capacitance.

(ii) Variation of space-charge with \(V_{CB}\) due to electron velocity modulation: It is shown in [465] that the intrinsic part of the B-C capacitance can be written as:

\[
C_{bcI} = \frac{\varepsilon A_E}{W_{BC}} - \tau_C \left| \frac{dI_C}{dV_{CB}} \right|_{W_{BC}=cte} - I_C \left| \frac{d\tau_C}{dV_{CB}} \right|_{W_{BC}=cte}
\]

where \(\tau_C\) is the electron delay time in the collector given by:

\[
\tau_C = \int_0^x \left(1 - \frac{x}{W_{BC}}\right) \frac{dx}{v(x)}
\]

where \(v(x)\) is the position dependent velocity of electrons. When \(v(x) = v_{n,avg} = cte.,\) (10.57) reduces to (10.5). The first term on the RHS of (10.56) is the dielectric capacitance associated with the changes in charge located at the edges of the depletion layer. The second and third terms describe the modulation of electron space-charge by \(V_{CB}\). The second term which is proportional to the output conductance is negligibly small for HBTs biased in the forward active mode [465]. Modulation of electron space-charge due to modulation of the average velocity in the collector is described by the third term. Increasing \(V_{CB}\) increases both the electric field and temperature of the B-C junction, thus reducing the electron velocity inside this region [131] (i.e., \((d\tau_C / dV_{cb})\)\(|_{W_{BC}=cte}\) is positive). Therefore, the overall intrinsic capacitance reduces at higher collector currents. This capacitance reduction phenomena has recently been employed as one of the improvement factors in realising HBTs with highest ever reported \(f_{max}\) of 820GHz [2].

The above current dependent phenomena happens inside the intrinsic part of the device where injection of electrons occurs. Therefore, one expects \(C_{bcx}\) to remain constant, reflecting all the change in \(C_{bc}\) to a similar change in \(C_{bci}\). Consequently, it is chosen in this thesis to keep \(V_{CB}\) constant during the recommended parameter extraction procedure. \(C_{bcx}\) will be calculated using low current (or cold) measured \(C_{bc}\) and \(r = (B-E\text{ area})/(B-C\text{ area})\) under the same \(V_{CB}\) value as in the high current data; any change in \(C_{bc}\) with current will be directly reflected into a similar change in \(C_{bci}\) and \(r\). However, one should notice that the variation of \(r\) at high current does not change anything in the determination of \(R_{bbx}\) from cold-HBT data.

Under a constant \(V_{CE}\) of 2.1V, \(C_{bc}\) shows an initial increase with \(I_C\). This is due to the fact that higher \(I_C\) requires higher \(V_{BE}\), and hence, a lower \(V_{CB}\) under constant \(V_{CE}\) condition. Therefore, both \(C_{bci}\) and \(C_{bcx}\) will increase initially, and it would be difficult to differentiate
between them. Variation of $C_{bc}$ with current under constant $V_{CE}$ was also observed in [441], but the authors did not explain this behaviour.

Other interesting features of the data in Table 10.3 include a reduction of $R_{bhi}$ (and $R_{cc}$ to a smaller extent) at higher currents due to the emitter current crowding. The values of inductances $L_b$, $L_c$, and $L_e$ seem to be more or less constant with bias; at highest bias point they all show some increase due to the device self-heating. $\alpha_0$ shows a continuous increase with collector current and saturates at higher currents before starting to fall-off at the highest bias point. This reflects to a similar trend for the variation of common-emitter DC current gain, $\beta$, with current. The sudden change of many of the parameters at $I_c = 40\,\text{mA}$ is most probably due to the device self-heating rather than Kirk effect. Kirk effect (or base pushout) is expected to happen at collector currents around 100mA for the dimension and collector doping level of the device under study. (See also the following discussion on delay times.) Since device temperature rise is expected to increase both $R_{ee}$ and $R_{be}$ [326], the sudden increase of $(R_{ee} + R_{be})$ at $I_c = 40\,\text{mA}$ is evenly divided between the two elements.

The base transit time remained almost constant $(1.61 \pm 0.08)\,\text{ps}$ as the collector current was varied in the range $(2-40)\,\text{mA}$. This further supports the idea that Kirk effect is not happening at $I_c = 40\,\text{mA}$. The base transit time is given by (10.4) where $v_{n,\text{avg.}}$ is an average velocity of electrons at the base end of the B-C depletion region. $v_{n,\text{avg.}}$ is higher than the static saturation velocity of electrons (described in chapter 5) due to the velocity overshoot effect. For this parameter a value of $2.05 \times 10^7 \,\text{cm/s}$ can be adopted as in [326]. For the base doping density in the present device, minority electron mobility of $\sim 2000 \,\text{cm}^2/\text{V.s}$ is expected [246: p.602] which results in $D_{nB} = 50 \,\text{cm}^2/\text{s}$. Therefore, using $W_B = 900\,\text{Å}$ one obtains $\tau_B = 1.25\,\text{ps}$ from (10.4). The larger measured value of $\tau_B$ is due to the carrier trapping behind the triangular potential barrier at the B-C heterojunction [60], as discussed in chapter 11. Also since minority electron mobility inside the base varies with temperature as $1/T$ [384], $D_{nB}$ will be almost temperature independent. Consequently, $\tau_B$ does not change significantly as self-heating occurs, though it may change slightly through the reduction of $v_{n,\text{avg.}}$.

The collector depletion layer delay time is expressed as in (10.5), where $W_{BC}$ can be estimated using $C_{bc} = 175\,\text{fF}$ as $\sim 0.19\,\text{μm}$. Therefore, average value of $\tau_C$ at low current levels $(0.79 \pm 0.07\,\text{ps})$ results in $v_{n,\text{avg.}} = 1.2 \times 10^7 \,\text{cm/s}$ inside InGaP collector, which is supposed to be somewhat smaller than the average velocity of electrons inside GaAs [60]. When self-heating occurs, this velocity is expected to be significantly reduced [326] (see also chapter 5), hence causing a sharp increase of $\tau_C$ under high current condition. The increase of $R_{be}$, $R_{cc}$, and $R_{ee}$ with temperature are other contributors to the enlargement of $\tau_E$ at the highest current level in Fig. 10.23. The $y$-intercept of the linear fit to the low current variation of $\tau_{EC}$ with $(1/I_C)$ is $5.65\,\text{ps} = \tau_F + (R_{ee} + R_{bc})C_{bc}$. Using the average values of $R_{ee}$, $R_{cc}$, and $C_{bc}$ in Table 10.3, the term $(R_{ee} + R_{bc})C_{bc}$ can be calculated as $3.35\,\text{ps}$, which results in $\tau_F \approx 2.30\,\text{ps}$. This is close, but not exactly equal, to the $y$-intercept of $(\tau_F + R_{be}C_{jbe})$ versus $(1/I_C)$,
which is 2.41ps. As discussed in section 10.4, the latter method is expected to give more accurate values of $\tau_F$.

The relatively early occurrence of the self-heating in the device studied so far in this chapter masks other high current phenomena, and does not allow us to observe the effect of base pushout in this device. In fact, the relatively high collector doping (5-7×10^{16} \text{cm}^{-3}) dictates a large current density for the onset of Kirk effect, $J_{Kirk} = qN_{DC}v_{n,\text{avg}}$. To study the Kirk effect on the small-signal parameters of HBTs, another DHBT with similar layer structure is considered here; the only major difference between this DHBT and that described in Table 10.1 is that collector doping of the new device is lower (1-2×10^{16} \text{cm}^{-3}). This results in a lower onset current for Kirk effect in the new device. Additionally, the substrate of the new wafer (400μm) is thinner than that in Table 10.1 (635μm), which results in a better thermal conductivity and a less pronounced self-heating in the new device.

Figure 10.26 shows the variation of measured $f_T$, $f_{\text{max}}$, $C_{bc}$, and $\beta_0$ versus collector current for an (8×10)×2 μm² device on the wafer described above under $V_{CB} = 0.8V$ condition. At collector currents below 20mA $C_{bc}$ reduces with $I_C$ as described previously. When the base pushout occurs at collector currents around 20mA, the B-C depletion width suddenly starts to shrink and $C_{bc}$ sharply increases. $f_{\text{max}}$ also follows the opposite trend. The fall-off of $f_T$, however, happens at a slightly higher current level (~30mA) and is less pronounced. More interestingly, the small-signal current gain $\beta_0$ does not show any decrease when the Kirk effect takes place. This is in clear contrast to the trends observed in Si bipolar transistors [391]. The above behaviour is also observed by Camnitz and Moll [131] and is described in full details in section 7.7. The flattening of $\beta_0$ followed by a slight reduction at the highest current level is most possibly due to the device self-heating. It is important to note that a sudden decrease of the DC current gain observed when measuring the Gummel plots (under $V_{CB} = 0$) of some of III-V HBTs is due to the quasi-saturation of the device (or Kink effect) as described in section 9.1.4. For example, the DC Gummel plot ($V_{CB} = 0$) of the device under study showed a sharp increase of $I_B$ at $I_C = 70$mA, but this feature was removed when $V_{CB}$ was increased to 0.8V.

Base pushout occurs in high
current density operation when the negative space-charge density is high enough for the electric field to vanish in the collector adjacent to the metallurgical base. In response, holes diffuse out from the base to form a quasi-neutral region, called the current-induced base with thickness $W_{cib}$. Operation above the threshold current for the Kirk effect reduces the width of the collector space-charge region by $W_{cib}$, and hence, $C_{bc}$ increases sharply.

Considering the current-induced base region, the base and collector delay times of equations (10.4) and (10.5) will be modified to:

$$
\tau_{B_{,Kirk}} = \frac{(W_B + W_{cib})^2}{2D_{n,avrg.}} + \frac{W_B + W_{cib}}{\nu_{n,avrg.}}
$$

(10.58)

$$
\tau_{C_{,Kirk}} = \frac{W_{BC} - W_{cib}}{2\nu_{n,avrg.}}
$$

(10.59)

where $D_{n,avrg.}$ is an appropriate average diffusion constant of the base and current-induced base regions, which would be larger than $D_{nB}$. In III-V HBTs, $D_{nB}$ is significantly larger than that in Si BJTs. Additionally, a highly doped base region allows a much thinner base layer for a reasonable base resistance. Consequently, $\tau_B$ in III-V HBTs is small and comparable to $\tau_C$, while in Si BJTs $\tau_B$ is much larger than $\tau_C$. As the base pushout occurs, the effective base width increases, while the effective B-C depletion region width shrinks. Hence, the delay added by the current-induced base is at first partially offset by the reduction in delay of the consequentially smaller collector space-charge region. As a result, the fall-off of $f_T$ occurs at a slightly higher current level and is less pronounced than that of $f_{max}$.

Figure 10.27 shows the extracted components of the emitter-to-collector delay time (Eq. (10.49)) for the device under study using the methods described in sections 10.3 and 10.4. As can be seen, the increase of $\tau_B$ and $(R_{ee} + R_{cc}) \cdot C_{bc}$ due to the base pushout is to some level compensated by the reduction of $\tau_C$.

### 10.5.2. Geometrical Scaling Rules

In this section layout and geometry dependence of some of the extracted parameters for devices fabricated on the wafer specified in Table 10.1 will be studied. Figure 10.28 shows the top view layout and cross-sectional view of a typical planar small-geometry HBT fabricated
using the mask set designed in this work. Details of the mask design and fabrication are given in Chapter 8. Here, empirical equations describing the variation of equivalent circuit elements with geometrical parameters will be given. Such equations, also regarded as scaling rules, are quite useful in designing the optimum layout for a known vertical layer structure and fabrication technology. The geometrical parameters used for the scaling, also defined in Fig. 10.28, include emitter length (EL), emitter width (EW), emitter metal width (EMW), base metal width (BMW), base-emitter separation (BES), collector-base separation (CBS), and base width (BW), all in units of micron (μm). Additionally, number of fingers for emitter (F_E), base (F_B = F_E + 1), and collector (F_C) are other important geometrical factors for determining the values of parasitic elements.

![Figure 10.28](image-url)

**Figure 10.28**: (a) Top view layout, and (b) vertical cross-section of a 2-finger planar HBT.
Devices used in this study had CBS = 2 or 3μm, BMW = 2, 3, or 4μm, EL = 10 or 20μm, BES = 1 or 2μm (few fully self-aligned test devices had BES = 0), FE = 1, 2, or 4, and FC = 1 or 2.

Before turning to the actual relationships between the elements and geometrical parameters, a short discussion on the discrepancy of extracted parameters is necessary. As can be seen in the forthcoming paragraphs, the discrepancy of the extracted parameters is relatively large. Even devices with exactly similar layout on different areas of the wafer show differences as large as ±50%. This inconsistency is due to many growth and fabrication related discrepancies, as well as some level of inaccuracy in the parameter extraction procedure. Some of the sources of parameter variation can be listed as follows:

- Thickness and doping non-uniformity across the wafer. For instance, the collector doping level was measured by C-V method in the range 5-7×10^{16} cm^{-3}.
- Variation of ohmic contact specific resistance (partly caused by the above reason) across the wafer. TLM measurement results on the same wafer shows ρc = (8±3)×10^{-6} Ω.cm^2 for the emitter contacts, and the base ohmic contacts have even larger standard deviations.
- Non-uniform ion implantation. Devices on some corners of the wafer were not properly isolated.
- Misalignment of various mask levels with respect to each other.
- Insufficient accuracy of some of the extracted parameters. Le, Cjbe, r, and Rbc are the least accurate extracted parameters using the methods described in this chapter.

Nevertheless, when a large number of devices are studied, some clear geometrical relationships can be observed for the averages of the parameters. Some of these relations are presented in the following.

\( R_{ee} \): Emitter resistance is composed of three components: pad-to-finger access resistance (including the via hole resistance), ohmic contact resistance, and semiconductor (mainly cap layer) resistance. The first term is almost constant, while the second and third components are inversely proportional to the total emitter metal area \((= F_E \cdot EL \cdot EMW)\) and total emitter mesa area \((= F_E \cdot EL \cdot EW)\), respectively. A simple calculation shows that the semiconductor resistance is much less than the ohmic contact resistance, and therefore, \( R_{ee} \) can be approximately written as:

\[
R_{ee} = K_1 + K_2 / (F_E \cdot EL \cdot EMW)
\]

where \( K_1 \) and \( K_2 \) are constants. Fig. 10.29 shows a plot of the extracted emitter series resistances versus inverse of the total emitter metal area. The average of the values for a particular emitter metal area are shown with solid circles, and the vertical lines identify the range of measured results. Fitting to the average points gives \( K_2 = 639Ω \cdot μm^2 = 6.39\times10^{-6}Ω \cdot cm^2 \) and \( K_1 = 3.95Ω \). \( K_2 \) is the specific contact resistance of the emitter ohmic contact, and is within the range of \((8 \pm 3)\times10^{-6}Ω \cdot cm^2 \) measured by TLM. The emitter access
resistance, $K_1$, is mainly composed of the emitter bridge resistance and interconnect (polyimide via hole) resistance, as described in chapter 8.

![Image 10.29- Emitter series resistance versus inverse of the total emitter metal area for the wafer no. EPI 1-3514A.](image)

![Image 10.30- Medium frequency effective base resistance as a function of $1/(F_E \cdot EL)$.](image)

![Image 10.31- Collector series resistance versus $1/(F_C \cdot EL)$.](image)

![Image 10.32- The base series inductance versus the ratio $EL/(BMW \cdot F_B)$.](image)

**Base series resistance:** The extrinsic base resistance, $R_{bbx}$, is composed of the base access resistance, base contact resistance, and the resistance of the region between the emitter mesa edge and base contact edge (i.e., the emitter undercut region). Therefore, it depends on many parameters such as $F_B$, BMW, EL, transfer length ($L_T$) and specific contact resistance ($\rho_c$) of the base ohmic contact, and the undercut width. The intrinsic base resistance also depends on $F_E$, EL and EW [44]. Therefore, formulation of base series resistance as a function of all the above parameters is a very difficult task and requires an extremely large set of data. Instead, it has been decided here to develop a simpler relation between the base series resistance and the most important parameters, namely $F_E$ and EL. Additionally, the formulation of medium frequency effective base resistance, $R_{Bi} = R_{bbx} + r \cdot R_{bbi}$, is considered instead of the intrinsic and extrinsic resistances individually. There are two reasons for the latter choice. First, $R_{Bi}$ can be extracted directly and precisely without prior knowledge of $r$ or $R_{bbx}$ from cold-HBT data, and therefore, it has a minimised extraction error. Second, $U$ and $f_{max}$ are formulated based on $R_{Bi}$ (see Eqs. (10.54)-(10.55)); knowledge of the extrinsic and intrinsic components is not required for the determination of $U$ and $f_{max}$. 

268
Figure 10.30 shows the variation of $R_{bi}$ with $1/(F_e \cdot EL)$. The fitting to the average of these data gives an access resistance (y-intercept) of 1.82Ω. Unlike emitter, base has neither a via hole nor a narrow bridge, and therefore has a lower access resistance. Scattering of the data can partly be explained by the fact that variation of BMW and EW are not considered in the graph of Fig. 10.30. Devices with BMW = 2μm and wider emitter fingers usually had larger base resistances. Other reasons listed in the beginning of this section are still among the possibilities for discrepancy.

$R_{cc}$: The collector series resistance is almost inversely proportional to $F_C$ and EL. Figure 10.31 shows the variation of $R_{cc}$ with $1/(F_C \cdot EL)$. Discrepancies in this figure may arise from the dependence of $R_{cc}$ on $F_E$ and BW. Observed values of $R_{cc}$ decrease for a larger number of emitter fingers and slowly increase with a wider base mesa, but a clear functional form for these dependencies could not be found.

$L_g$: Similar to the series resistances, the base series inductance is expected to include an access term in series with the parallel combination of single-finger inductances. The latter term is proportional directly to the finger length (~EL), and inversely proportional to the number of base fingers ($F_B$) and their width (BMW). Therefore, the base inductance can be written as:

$$L_b = 19.6 + 6.9 \frac{EL}{BMW \cdot F_B}$$  \hspace{1cm} (10.61)

The constants in the above relation are found from the fitting in Fig. 10.32. The fact that the base fingers are longer than the emitter length (EL) can be regarded as an additional cause for discrepancy in the graph of Fig. 10.32.

$L_e$ and $L_c$: The geometrical dependence of $L_e$ is similar to that of $L_b$, when BMW and $F_B$ are replaced by EMW and $F_E$. As to the collector series inductance, the entire collector metallisation (including the collector pad and ohmic contact) has a completely unchanged pattern in all different geometries on the designed mask (apart from the number of collector contacts which is either 1 or 2). Therefore, the finger component of $L_e$ is expected to be only inversely proportional to the number of collector contacts, $F_C$. The observed fitting functions are as follows:

$$L_e = \left(38 + \frac{4EL}{F_E \cdot EMW}\right) \pm 12 \text{ [pH]}$$  \hspace{1cm} (10.62)

$$L_c = \left(22 + \frac{34}{F_C}\right) \pm 8 \text{ [pH]}$$  \hspace{1cm} (10.63)

It can be observed that the deviations in $L_e$ and $L_c$ are relatively large, especially for $L_c$. In the case of $L_c$, an increasing trend with increasing BW was observed, which can not be clearly justified. It is possible that the double collector fingers of the transistor form part of a loop, diameter of which is controlled by BW.
Chapter 10    

High Frequency Characteristics of HBTs

$C_{bc}$ and $C_{jbe}$: The B-C and B-E junction capacitances are modelled using the second term in Eq. (10.10), in which only $C_{jbc0}$ and $C_{jbe0}$ are geometry dependent. The geometry dependence of $C_{jbc0}$ and $C_{jbe0}$ was shown in Fig. 10.12, and the following relations were derived:

\[
C_{jbc0} = 0.57 A_{BC} \text{[fF]} \Rightarrow W_{BC} \approx 0.18 \mu m \tag{10.64}
\]

\[
C_{jbe0} = 0.85 A_{BE} \text{[fF]} \Rightarrow W_{BE} \approx 0.12 \mu m
\]

where $A_{BC}$ and $A_{BE}$ are the B-C and B-E junction areas (in $\mu m^2$), respectively, after correction for etch undercuts. It is important to mention that the junction capacitance expressed by Eq. (10.10) is developed based on the depletion approximation which fails to hold under large forward bias condition [316]. When the forward bias across the junction approaches the built-in potential, the carrier density inside the space-charge region gets large and the depletion approximation is no longer valid (also see section 7.6 and Fig. 7.19). Therefore, one has to use (10.10) with care for the B-E junction of an HBT in forward active mode. Full depletion of the emitter and collector regions is another issue to be considered, as discussed in subsection 10.3.1.

**Pad Capacitances:** The parasitic pad capacitances are discussed in subsection 10.3.1, and it was stated that a constant parasitic capacitance is a reasonable assumption. The only exception was $C_{pce}$ which was largely dependent on the number of collector contacts because of the collector-emitter metal crossover. The final results of subsection 10.3.1 are summarised below:

\[
C_{pce} = (30.5 + 10.5F_C) \pm 5 \text{ fF}
\]

\[
C_{pbe} = 20.0 \pm 2 \text{ fF} \tag{10.65}
\]

\[
C_{pbc} = 12.2 \pm 2 \text{ fF}
\]

$\tau_B$ and $\tau_C$: The low bias values of $\tau_B$ and $\tau_C$ are expected to be geometry independent. However, relatively large deviations were found in the extracted values of these delay times, especially in the case of $\tau_C$: $\tau_B = (1.83 \pm 0.25) \text{ps}$, $\tau_C = (0.71 \pm 0.25) \text{ps}$. The variation of $\tau_B$ and $\tau_C$ in different devices was completely random, and did not show any particular dependence on layout. The large discrepancy in $\tau_C$ can be partly described by the collector doping variation across the wafer.

10.6. Sensitivity Analysis

In this section, the sensitivity of $h_{21}$ and $U$ to the parameters of the equivalent circuit in Fig. 10.4 is studied. The values of the parameters for the original circuit are given in Table 10.4, and for each graph only one variable is stepped. The graphs in Figs. 10.33(a)-(p) summarise the results. In order to show a correct graphical view of the sensitivity to various parameters, those parameters which normally have comparable magnitude are assumed to be exactly the same in the original circuit. For instance, $R_{ce}$, $R_{cc}$, $R_{bb1}$, and $R_{bbx}$ all assume the value of $10 \Omega$.  

270
Table 10.4 - Values of the original small-signal equivalent circuit parameters for sensitivity analysis.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_{ee}$</td>
<td>10Ω</td>
</tr>
<tr>
<td>$R_{cc}$</td>
<td>10Ω</td>
</tr>
<tr>
<td>$R_{bbx}$</td>
<td>10Ω</td>
</tr>
<tr>
<td>$R_{bbi}$</td>
<td>2Ω</td>
</tr>
<tr>
<td>$R_{be}$</td>
<td>25kΩ</td>
</tr>
<tr>
<td>$C_{be}$</td>
<td>160fF</td>
</tr>
<tr>
<td>$C_{ibc}$</td>
<td>160fF</td>
</tr>
<tr>
<td>$r$</td>
<td>0.50</td>
</tr>
<tr>
<td>$C_{pbe}$</td>
<td>20fF</td>
</tr>
<tr>
<td>$C_{pce}$</td>
<td>60fF</td>
</tr>
<tr>
<td>$C_{pbc}$</td>
<td>12fF</td>
</tr>
<tr>
<td>$L_{b}$</td>
<td>30pH</td>
</tr>
<tr>
<td>$L_{c}$</td>
<td>30pH</td>
</tr>
<tr>
<td>$L_{e}$</td>
<td>30pH</td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>0.99</td>
</tr>
<tr>
<td>$\tau_B$</td>
<td>1.20ps</td>
</tr>
<tr>
<td>$\tau_C$</td>
<td>0.60ps</td>
</tr>
</tbody>
</table>

From Figs. 10.33(a)-(b), it can be seen that both $h_{21}$ and $U$ are affected by $R_{ee}$ and $R_{cc}$. The effects of $R_{ee}$ and $R_{cc}$ on the medium frequencies are comparable. However, differences can be observed at very high or very low (DC) frequencies. $h_{21}$'s of the variable $R_{ee}$ graph show some converging characteristics at high frequencies which cannot be observed in the variable $R_{cc}$ graph. This is due to the existence of $R_{ee}$ (and absence of $R_{cc}$) in the numerator of (10.46), which introduces a zero in the transfer characteristics and starts to become important at frequencies approaching and above $f_{i}$. Starting from the definitions of $h_{21}$ and $U$, it is straightforward to show that at very low frequencies (approaching DC):

$$|h_{21}|_{DC} = \frac{\alpha_0 R_{bc}}{R_{ee} + R_{be} + R_{cc} + (1-\alpha_0)R_{bc}}$$  \hspace{1cm} (10.66)

$$U_{DC} = \frac{\alpha_0^2 R_{bc}}{4(R_{ee} + R_{be})}$$  \hspace{1cm} (10.67)

The effects of $R_{ee}$ and $R_{cc}$ on the DC current gain is negligible in this case, since $(1-\alpha_0)R_{bc} >> (R_{ee} + R_{be} + R_{cc})$, and $|h_{21}|_{DC} = \alpha_0/(1-\alpha_0) = \beta_0$ according to (10.66). However, $U$ is almost inversely proportional to $R_{ee}$ at very low frequencies.

$R_{bbx}$ and $R_{bbi}$ almost have no effect on $h_{21}$ and $U_{DC}$, but they both affect the unilateral gain at medium to high frequencies, hence influencing $f_{\text{max}}$. $R_{bbi}$ has a smaller effect on $f_{\text{max}}$ when compared to $R_{bbx}$. This is because $R_{bbi}$ is multiplied by $r$ and then added to $R_{bbx}$ in the expressions for $U$ and $f_{\text{max}}$ (see Eqs. (10.54)-(10.55) and the definition of $R_{B1}$). Therefore, minimising components of $R_{bbx}$, such as base ohmic contact resistance, is more important than $R_{bbi}$ for improving the high frequency performance. $R_{be}$ influences $h_{21}$ and $U$ only at low biasing current levels. Otherwise, $R_{be}$ is much smaller than $R_{ee}$ and can be ignored as compared to this resistance and others in the corresponding expression. $R_{be}$ has almost no effect on $h_{21}$ and $U$ at frequencies higher than ~3GHz. However, it largely affects $U$ at low frequencies (see Eq. (10.67)), and has some influence on $|h_{21}|_{DC}$. When $R_{be}$ is small, $|h_{21}|_{DC}$ varies with changes in $R_{ee}$, $R_{be}$, $R_{cc}$, and $R_{bc}$, as shown in Eq. (10.66). But, when $R_{bc}$ is so large that $(1-\alpha_0)R_{bc} >> (R_{ee} + R_{be} + R_{cc})$, then the current gain is only defined by $\beta_0$ and is independent of other parameters.
Figure 10.33- Results of stepping of various parameters of the HBT small-signal equivalent circuit shown in Fig. 10.4. Variations of $h_{21}$ and $U$ are shown when: (a) $R_{ee}$ (b) $R_{cc}$ (c) $R_{bbx}$ (d) $R_{bbi}$ (e) $R_{be}$ (f) $R_{bc}$ (g) $C_{be}$ (h) $C_{jbe}$ (i) $C_{pe}$ (k) $L_{b}$ (l) $L_{c}$ (m) $\alpha_{0}$ (n) $r$ (o) $\tau_{b}$ (p) $\tau_{c}$ are varied within the specified ranges. In each case the dashed lines correspond to the lowest value of the stepping parameter.
Figure 10.33- Continued
Chapter 10 High Frequency Characteristics of HBTs

Figure 10.33. Continued

C_{bc} is without doubt the most influential parameter in determining the high frequency performance of HBTs, while C_{jbe} has almost no effect on the current/power gain characteristics. This has important implications on high frequency device design and fabrication. Many techniques have been explored in the literature to minimise the B-C capacitance, among them undercutting the extrinsic base-collector mesa [85,94], buried collector structures [15,66], extrinsic collector ion bombardment [466], and more recently, the transferred substrate method [2,467]. An alternative method is to use the collector-up structure instead of the conventional emitter-up [20,468]. This latter technique reduces C_{bc} with the expense of increasing C_{jbe}. But, C_{jbe} is shorted by a small resistance R_{be}, and as shown in Fig. 10.33(h), has almost no effect on the high frequency figures of merit. Designing the HBT in collector-up configuration, however, has its own difficulties as discussed in subsections 3.2.3 and 9.1.5.

The pad and interconnect parasitics are shown in Figs. 10.33(i)-(l) not to have a major
influence on $\beta_{21}$ and $U$. $C_{Pce}$ has absolutely no effect on $\beta_{21}$, $U$, and $MSG^\dagger$. This is why the unavoidable metal crossover is designed to exist for the collector-emitter metals, and not the collector-base or base-emitter metals. $C_{Pbe}$ again has absolutely no effect on $U$ (and $MSG$), but slightly changes $\beta_{21}$. $C_{Pbc}$ is the most influential of the three pad capacitances, which affects both $\beta_{21}$ and $MSG$. The interconnect inductances also have no effect on the power gain, since the input and output matching circuitry, designed to deliver the maximum power, cancel the effect of any series reactance in the input or output by adding the complex conjugate of that component in series. $\beta_{21}$ is influenced by the variation of inductances only at frequencies above $f_T$. A larger $L_b$ results in a flatter current gain at high frequencies. This may have some applications to make the current gain bandwidth wider. It can be shown that when $\omega^2 L_b (C_{Pbe} + C_{Pbc})$ approaches unity, the gain starts to become flat; it then finds a resonance peak at $\omega = 1/\sqrt{L_b (C_{Ppe} + C_{Pbc})}$ before sharply falling down.

$\alpha_0$ is seen to have a large influence on $|\beta_{21}|_{DC}$, while its effect on $U$ is just minimal (compare (10.66) with (10.67)). The fact that $\alpha_0$ does not cause any change in $f_T$ or $f_{max}$ of the transistor means that a larger DC current gain does not necessarily result in a lower (or higher) cutoff frequency, as also pointed out by Liu et al. [469]. The effect of $r$ on $U$ is only through the parameter $R_{B1} (= R_{bbx} + r.R_{bbi})$, and is not much significant. Both $\tau_B$ and $\tau_C$ affect $f_T$ largely, and $f_{max}$ to a smaller extent. In fact, when the parasitics of the transistor (especially $C_{bc}$) is reduced, minimising the forward transit time becomes even more crucial in improving the high frequency performance of the transistor.

From the sensitivity analysis of this section it can be concluded that $C_{bc}$, $\tau_B$, and $\tau_C$ are the most important parameters for high frequency HBTs. Base, collector, and emitter series resistances occupy the next level of importance. The pad and interconnect parasitics are shown not to have a significant effect on $f_T$ and $f_{max}$, unless other intrinsic and parasitic delay components are reduced to the values much lower than those used in this example.

---

$^\dagger$ By definition, $\beta_{21}$ is the current gain when output is shorted. Therefore, $C_{Pce}$ will be shorted when $\beta_{21}$ is determined. Also $MSG = |Y_{21}/Y_{12}|$, and according to (10.7)-(10.8), among the pad capacitances, only $C_{Pbc}$ affects $MSG$.  

275
CHAPTER 11. OPTIMISATION OF THE B-C HETEROSTRUCTURE OF InGaP/GaAs DHBTs

The experimental results in chapter 9 demonstrated that InGaP/GaAs DHBTs are quite advantageous over both InGaP and AlGaAs SHBTs in terms of their DC characteristics. These advantages included larger breakdown voltage, lower offset voltage (which is also relatively insensitive to the bias variation), less variation of current gain with temperature, and lower B-C leakage current leading to the capability to operate at higher temperatures. However, as also mentioned in the previous chapters, the existence of a finite conduction band spike at the B-C heterojunction of DHBTs impedes the free collection of electrons, increases the base bulk and extrinsic base surface recombination, and degrades the current gain. The insertion of an n-GaAs spacer layer between the p\(^+\)-GaAs base and n-InGaP collector was shown to improve the current gain of InGaP DHBTs. However, it will be demonstrated in this chapter that the carrier trapping behind the triangular potential barrier forming inside the GaAs spacer may significantly degrade the high frequency performance of these devices compared to SHBTs. This effect is first explored in this chapter with the aid of the small-signal parameter extraction method developed in chapter 10. After exposing the origin of the longer delay times in InGaP DHBTs with spacer design, numerical simulation is utilised to optimise the B-C heterostructure of these devices for an improved high frequency performance. Finally, one of the optimised structures designed in this chapter is verified experimentally. A direct comparison between InGaP/GaAs double- and single-HBTs shows that with careful design of the B-C heterostructure it is possible to realise DHBTs with high frequency performance as good as SHBTs, while the former devices still offer superior DC characteristics.

Design of power HBTs will not be complete without thermal considerations and power measurement. Self-heating is known to be the dominant limiting factor in the design of power HBTs [470-471], especially in the case of GaAs substrates with poor thermal conductivity. Efficient heat removal is one of the most challenging objectives in the design of GaAs power devices. Various techniques such as wafer thinning and via hole formation [71: chap.16], thermal shunts [470], and the flip-chip technique [471], are suggested in the literature for more efficient thermal management. These techniques require complicated top- or back-side fabrication processes, which unfortunately could not easily be carried out in our research laboratory. Consequently, only electronic requirements of a microwave power amplifier are emphasised in this work, and the cutoff frequency-breakdown voltage product \((f_T \times BV_{CEO})\) is defined as the microwave power figure of merit. This figure of merit is also used in some other works, e.g., [136]. Finally, large-signal power parameters such as maximum output power and power added efficiency (PAE) are not measured here mainly due to the unavailability of measurement set-ups.
11.1. Small-Signal Parameters of InGaP Single- and Double-HBTs

In this section, the high frequency performance and small-signal equivalent circuit elements of InGaP/GaAs single- and double HBTs are compared using the methods described in chapter 10. The layer structures of the two HBTs are similar to those in Table 9.1 with the base doping of $4 \times 10^{19}$ cm$^{-3}$. Collector doping of the single- and double-HBTs is estimated as $1.7 \times 10^{16}$ (GaAs) and $2 \times 10^{16}$ cm$^{-3}$ (InGaP), respectively. Both devices are fabricated using the processing steps detailed in chapter 8, but with the old mask set. DC characteristics of these HBTs were measured using an HP4145B Semiconductor Parameter Analyser (SPA). The large-geometry test SHBTs and DHBTs on these wafers have DC current gains of 28 and 20, B-C breakdown voltages ($B V_{CBO}$) of 21 and 32 V, and C-E offset voltages of 120 and 25 mV, respectively (i.e., the DC results shown in sections 9.1.1 and 9.1.2). On-wafer s-parameter measurements were then carried out using HP8510 Network Analyser and Cascade Microtech RF probes. The measured $f_T$ and $f_{\max}$ for an $8 \times 10$ $\mu$m$^2$ emitter SHBT were 37 and 25 GHz, while those for the DHBT were significantly lower at 23 and 19 GHz, respectively. In order to determine the reason behind the low $f_T$ and $f_{\max}$ of the DHBT, small-signal equivalent circuit elements of Fig. 10.4 were extracted from the measured s-parameter data for both transistors. The results are summarised in Table 11.1.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>SHBT</th>
<th>DHBT</th>
<th>Parameter</th>
<th>SHBT</th>
<th>DHBT</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I_C$ [mA]</td>
<td>19</td>
<td>15</td>
<td>$V_{CE}$ [V]</td>
<td>2.5</td>
<td>2.5</td>
</tr>
<tr>
<td>$C_{pbe}$ [fF]</td>
<td>30</td>
<td>33</td>
<td>$L_b$ [pH]</td>
<td>54</td>
<td>48</td>
</tr>
<tr>
<td>$C_{pbc}$ [fF]</td>
<td>13</td>
<td>12</td>
<td>$L_{bc}$ [pH]</td>
<td>58</td>
<td>60</td>
</tr>
<tr>
<td>$C_{pce}$ [fF]</td>
<td>45</td>
<td>39</td>
<td>$L_e$ [pH]</td>
<td>50</td>
<td>52</td>
</tr>
<tr>
<td>$R_{be}$ [$\Omega$]</td>
<td>1.4</td>
<td>1.7</td>
<td>$R_{ee}$ [$\Omega$]</td>
<td>13</td>
<td>15</td>
</tr>
<tr>
<td>$R_{bxx}$ [$\Omega$]</td>
<td>30</td>
<td>34</td>
<td>$R_{bb}</td>
<td>$ [$\Omega$]</td>
<td>12</td>
</tr>
<tr>
<td>$R_{ccb}$ [$\Omega$]</td>
<td>10</td>
<td>15</td>
<td>$R_{bc}$ [$k\Omega$]</td>
<td>~20</td>
<td>~30</td>
</tr>
<tr>
<td>$C_{ibc}$ [fF]</td>
<td>102</td>
<td>95</td>
<td>$C_{bc}$ [fF]</td>
<td>48</td>
<td>55</td>
</tr>
<tr>
<td>$r$</td>
<td>0.533</td>
<td>0.533</td>
<td>$\alpha_0$</td>
<td>0.955</td>
<td>0.944</td>
</tr>
<tr>
<td>$\tau_B$ [ps]</td>
<td>1.5</td>
<td>2.6</td>
<td>$\tau_C$ [ps]</td>
<td>1.6</td>
<td>2.1</td>
</tr>
<tr>
<td>$f_T$ [GHz]</td>
<td>37</td>
<td>23</td>
<td>$f_{\max}$ [GHz]</td>
<td>25</td>
<td>19</td>
</tr>
</tbody>
</table>

A brief look at the data in Table 11.1 reveals that the main contributors to the smaller $f_T$ and $f_{\max}$ in the DHBT are the larger $\tau_B$, $\tau_C$, $R_{cc}$, and $C_{bc}$, as well as the smaller collector current for the peak performance. The larger $C_{bc}$ in the case of the DHBT is most possibly related to the higher collector doping level in this device, which leads to the narrower collector depletion width. But, then, this should result in a lower collector SCR delay time, $\tau_C$, in the case of the DHBT. However, since the average velocity of electrons inside the InGaP collector is smaller than that inside the GaAs collector of SHBTs, the resultant $\tau_C$ is still larger in the DHBT. Simple calculations of the depletion region width, $W_{BC}$, for the biasing
condition and doping densities used here yields the values of 0.46 and 0.41 μm, for the SHBT and DHBT, respectively. These, in turn, together with the extracted \( \tau_C \)'s would give \( v_{n,\text{avg.}}(\text{GaAs}) \approx 1.4 \times 10^7 \text{ cm/s} \), and \( v_{n,\text{avg.}}(\text{InGaP}) \approx 1.0 \times 10^7 \text{ cm/s} \), which are reasonable considering the possibility of velocity overshoot in a thin portion (~1000Å) of the collector [54-55]. The calculated \( C_{bc} \)'s based on the above \( W_{BC} \)'s are, however, somewhat smaller than the extracted values in Table 11.1.

The larger \( \tau_B \) in the DHBT deserves more attention, since the base layer in both devices is \( \text{p}^+ \)-GaAs. In fact, the larger base transit time in the case of the DHBT is related to the carrier trapping behind the triangular potential barrier formed at the GaAs-spacer/InGaP-collector heterojunction (see the band structure in Fig. 11.1). This causes an additional delay time for electron transfer from emitter to collector, thus degrading the high frequency performance of the device. This potential barrier is also known to reduce the maximum current density prior to the \( f_T \) fall-off [100]. Therefore, the peak \( f_T \) of the DHBT is obtained in smaller current levels as compared to the SHBT. This leads to a larger \( R_{be} \) and \( \tau_E \) in the DHBT. These disadvantages of the DHBT as to the high frequency performance are confirmed in the next section with the aid of numerical simulation results.

The larger \( R_{cc} \) in the case of DHBT can be attributed to the resistance of the undepleted InGaP collector, or to the InGaP-Collector/GaAs-sub-collector heterojunction. The slight discrepancy between the other extracted parameters of the SHBT and DHBT are related to the non-uniformity in the layer structure or fabrication of the two wafers.

### 11.2. Numerical Simulation of the Delay Times

The main conclusion of the last section was that the conduction band spike formed at the B-C heterojunction of DHBTs significantly degrades their high frequency performance. Therefore, appropriate techniques should be employed in order to either completely remove this potential barrier, or at least reduce its height. These techniques, which are briefly explained in subsection 3.2.1, include continuous compositional grading between the base and collector materials, inserting a thin spacer layer and/or a doping spike between the base and collector, using a short-period linearly graded chirped superlattice (CSL), using \( \text{Al}_{0.11}\text{Ga}_{0.89}\text{As} \) at the collector end of the base region of InGaP/GaAs DHBTs, and finally step-grading the B-C heterojunction. The relative merits and disadvantages of the above methods were also discussed in subsection 3.2.1, where it was mentioned that the continuous grading of heterojunctions such as InGaP/GaAs and InP/InGaAs is practically difficult to achieve due to the necessity of a simultaneous control over both group V and group III sublattice elements. Instead, step-grading of the B-C heterojunction using InGaAsP quaternary material may be employed. InGaAsP lattice-matched to InP has widely been used to step-grade the B-C heterojunction of InP/InGaAs HBTs [107-109,135,472], and impressive DC and RF characteristics have been observed. Numerical simulation of such structures for minimised carrier blocking is also reported in [473], where it is shown that the optimised structure
Chapter 11 Optimisation of the B-C Heterostructure of InGaP/GaAs DHBTs

consists of two 200Å n-InGaAsP intermediate layers such that the band offsets at the heterointerfaces are approximately equal. The use of InGaAsP lattice-matched to GaAs for step-grading the B-C heterojunction of InGaP/GaAs DHBTs has been suggested for the first time by the present author [60]. In the following, it will be shown that introducing two thin layers of InGaAsP lattice-matched to GaAs with intermediate bandgap values between those of GaAs and InGaP can significantly improve the high frequency performance of InGaP/GaAs DHBTs.

To further elucidate the advantage of the new structure over conventional designs of InGaP/GaAs DHBTs, a comparison of the numerically simulated (intrinsic) cutoff frequencies for various SHBT and DHBT structures is carried out. The layer structures of the devices are summarised in Table 11.2. All devices have similar emitter, base, and sub-collector structures, as well as similar collector thicknesses of 5000Å. The difference is only in the material and doping of the collector. Device DHBT-A is a simple DHBT without any special treatment for the B-C heterojunction spike. In device DHBT-B a GaAs spacer layer is inserted between GaAs base and InGaP collector layers to lower the potential spike, similar to the experimental DHBT studied in the previous section. Device DHBT-C employs a doping spike in the collector which both lowers the conduction band spike and reduces the width of this barrier, thus improving the tunnelling transmission across it. Finally, DHBT-D is the proposed step-graded structure in the present work, which uses two intermediate bandgap InGaAsP layers sandwiched between the GaAs spacer and the InGaP doping spike.

Table 11.2- Layer structure of the single and double HBTs used for the numerical simulation.

<table>
<thead>
<tr>
<th>Layer</th>
<th>SHBT</th>
<th>DHBT-A</th>
<th>DHBT-B</th>
<th>DHBT-C</th>
<th>DHBT-D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cap Layer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>300Å n⁺-In0.5Ga0.5As</td>
<td>1×10¹⁹ cm⁻³</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>300Å n⁺-In1.5GaAs (x = 1 → 0.5)</td>
<td>1×10¹⁹ cm⁻³</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000Å n⁺-GaAs 5×10¹⁸ cm⁻³</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>300Å n⁺-InGaP 2×10¹⁸ cm⁻³</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emitter</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000Å n-InGaP 4×10¹⁷ cm⁻³</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>800Å p⁺-GaAs 4×10¹⁹ cm⁻³</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n⁻-GaAs spacer 3×10¹⁶ cm⁻³</td>
<td>200Å</td>
<td>300Å</td>
<td>200Å</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n⁻-In0.12Ga0.88As0.75P0.25,3×10¹⁶ cm⁻³</td>
<td></td>
<td></td>
<td></td>
<td>200Å</td>
<td></td>
</tr>
<tr>
<td>n⁻-In0.29Ga0.71As0.40P0.60,3×10¹⁶ cm⁻³</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>200Å</td>
</tr>
<tr>
<td>n⁺-InGaP doping spike 2×10¹⁸ cm⁻³</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n⁻-InGaP collector 3×10¹⁶ cm⁻³</td>
<td>5000Å GaAs</td>
<td>5000Å</td>
<td>4800Å</td>
<td>4650Å</td>
<td>4350Å</td>
</tr>
<tr>
<td>Sub-collector</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7000Å n⁺-GaAs 5×10¹⁸ cm⁻³</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semi-insulating GaAs Substrate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 11.1 shows the simulated conduction band diagram of the B-C heterojunction with various designs. This figure clarifies the significance of the step-graded structure in minimising the energy barriers in the conduction band, and thus the base transit time. As can be seen in this figure, the conduction band spike in the spacer design is lowered to give
electrons arriving from the base more chance to jump over it. But, the triangular barrier thus formed is deep and wide enough so that still a large number of electrons may get trapped behind this barrier. This introduces some extra delay time, and degrades the high frequency performance of DHBT-B as compared to the SHBT.

![Figure 11.1](image)

**Figure 11.1** - Numerically simulated conduction band diagram of the B-C heterojunction of various InGaP/GaAs DHBTs as specified in Table 11.2 at thermal equilibrium.

Carrier blocking effect in the case of DHBT-A is so severe that the simulated current gain for this device is even less than unity. Figure 11.2 shows the simulated common-emitter output characteristics of this DHBT. It can be seen that the current gain of this device is very much dependent on $V_{CE}$, and approaches unity only at very high C-E voltages. On the other hand, all other four structures have current gains much larger than unity. The calculated current gains at $J_C = 2 \times 10^4$ A/cm$^2$ and $V_{CB} = 0.5$ V were 48.5, 1.2, 40.5, 48.8, and 48.6 for the SHBT, DHBT-A, DHBT-B, DHBT-C, and DHBT-D, respectively.

The output characteristics of the DHBT-B with spacer design is significantly improved over the DHBT with abrupt B-C design. A typical simulated output characteristics for such a device is shown in Fig. 7.2 and shall not be repeated here. However, the current gain of this DHBT at $J_C = 2 \times 10^4$ A/cm$^2$ is almost 15% smaller than that for the SHBT, and more importantly it shows some gradual

![Figure 11.2](image)

**Figure 11.2** - Common-emitter output characteristics of the abrupt InGaP/GaAs DHBT studied in this section (i.e., DHBT-A).
reduction with further increasing the current density. To understand the reason behind this behaviour, the band structure of this device is simulated at three different $J_C$'s of 0.0, $\approx 2.5 \times 10^4$, and $\approx 6 \times 10^4$ A/cm$^2$. The resultant conduction band profile near the B-C heterojunction is shown in Fig. 11.3, where the curves are vertically shifted with respect to each other for more clarity. As the current density increases, more and more electrons are injected into the collector SCR, where they may neutralise some of the background donor concentration (i.e., early stages of the base pushout). This reduces the electric field inside the collector SCR, thus decreasing the voltage drop across the spacer region. As a result, the conduction band spike will be raised in energy, and may surpass the conduction band edge inside the base. Eventually, the DHBT-B with 200Å spacer behaves similar to the DHBT-A with an abrupt B-C heterojunction. This effect has also been pointed out by some other authors [100,409]. Its most detrimental consequence is an early reduction of the cutoff frequency to be discussed next.

Figure 11.4 shows the calculated cutoff frequencies versus collector current density for the structures specified in Table 11.2. At very low current densities, the total delay times, $\tau_{EC}$, in all five structures are similar, since they are dominated by the large emitter charging time, $\tau_E$, due to the large $R_{be}$ (see Eq. (10.49)). At higher current densities, however, the emitter charging time gets progressively smaller, and other components of delay time start to dominate. In the case of the abrupt DHBT-A, carrier blocking effect of the B-C heterojunction spike increases the base transit time, $\tau_B$, dramatically (more than 30 ps for the present calculations). Consequently, the peak $f_T$ of DHBT-A is limited to less than 5 GHz. The insertion of a 200Å spacer between the base and collector in DHBT-B improves the base transit time for low-to-medium current densities. However, as shown in Fig. 11.3, when $J_C$ approaches $10^4$ A/cm$^2$, the conduction band spike will be raised in energy, stopping the free collection of electrons. Therefore, $f_T$ of DHBT-B peaks at $J_C \sim 10^4$ A/cm$^2$, far below the onset of the Kirk effect in an SHBT with comparably doped collector. Adding a thin doping spike layer in the collector (as in DHBT-C) introduces an extra electric field component which shifts the potential spike to lower energies, thus eliminating the early fall-off of the cutoff frequency. Not surprisingly, the closest performance to that of the SHBT belongs to the proposed step-graded structure (DHBT-D). In this structure, the relatively large potential
spike at an abrupt InGaP/GaAs heterojunction is divided into three smaller barriers, which can easily be overcome by the energy of electrons. The presence of the doping spike after the InGaAsP layers is also expected to shift all these tiny barriers to lower energies, effectively eliminating any possibility of charge trapping behind them. Moreover, the larger electric field introduced by the doping spike reduces the depletion region width inside the collector and its associated delay time, $\tau_C$, partly making up for the smaller velocity of electrons inside InGaP as compared to GaAs. But, at the same time, the large doping inside the collector increases $C_{bc}$. This is also compensated to some extent by the smaller dielectric constant of InGaP (11.8) compared to that of GaAs (12.9) (see subsection 5.2.2). As a result, the high frequency performance of the InGaAsP step-graded DHBT is almost identical to that of the SHBT, while still the former may benefit from superior DC characteristics.

![Figure 11.4](image)

**Figure 11.4**: Numerically simulated cutoff frequencies versus collector current density for the SHBT and various DHBT structures specified in Table 11.2. $V_{CE} = 2.0\, \text{V}$ is kept constant.

Next, the number of InGaAsP layers and their composition was optimised for highest speed. It was observed that using three InGaAsP layers, instead of the two in Table 11.2, did not improve the cutoff frequency significantly. The peak cutoff frequency was increased by only 2 GHz, and it was decided not to add the third InGaAsP layer, since the breakdown voltage of the device may be largely compromised. As to the composition of the $\text{In}_{1-x}\text{Ga}_x\text{As}_y\text{P}_{1-y}$ layers, it was observed that variation of the arsenic mole fraction, $y$, of the two layers in the ranges $(0.7 \pm 0.1)$ and $(0.35 \pm 0.1)$ did not change the $f_t-J_C$ characteristics notably, and therefore, the design is not much sensitive to the slight random variations of composition during the growth.
11.3. Experimental Verification of an Improved Design InGaP DHBT

Although the growth of InGaAsP layers lattice-matched to InP is routinely carried out for optoelectronic applications, not many growth systems are optimised for the growth of this quaternary material lattice-matched to GaAs. Consequently, the growers of the epitaxial layers used throughout this work were not immediately able to produce good quality GaAs HBTs comprised of InGaAsP layers. Therefore, the experimental results in this section concentrate on the InGaP/GaAs DHBT with doping spike design, which was the second best among those in Table 11.2. Still, it will be demonstrated that the doping spike design of Table 11.2, with a modified collector doping, is capable of producing high frequency performance almost as good as that of the SHBT, while the breakdown voltage and offset voltage of this DHBT is still superior to those of the SHBT.

Optimum design of the devices of type DHBT-C is very much sensitive to the thicknesses of both the spacer and the doping spike layers, as well as to the doping level of the latter. Although the spacer thicknesses of 30-50Å have been tried in the past [105-106,474-476], the calculations in the present work show that the conduction band spike at the GaAs-spacer/InGaP-doping spike interface may rise above the conduction band edge of the base layer at high current densities. The relatively large saturation voltages observed in the output characteristics of some of the above works [474,476], or the finite positive output resistance seen in all of them [105-106,474-476] confirm the calculations in the present work. Additionally, the close proximity of a highly doped collector region to the highly doped base layer increases the possibility of Zener breakdown. Indeed, an attempt to reproduce the design of Song et al. [105] (with 50Å GaAs spacer and 150Å InGaP doping spike) by the author failed because of the Zener breakdown. The output characteristics of this DHBT was previously shown in chapter 3 (see Fig. 3.3).

The effect of various collector designs on the DC characteristics of InGaP/GaAs single- and double-HBTs is systematically studied in [104]. It is found that when the spacer thickness is increased beyond 300Å, the breakdown voltage and offset voltage of the device may be largely compromised. On the other hand, a spacer thickness of 50Å resulted in 35% reduction of current gain and a relatively large saturation voltage. Therefore, it was concluded that a 300Å thick spacer together with a 50Å doping spike layer produces the optimum DC results. But the authors in [104] did not show any high frequency performance. To the best of author’s knowledge, the high frequency performance of InGaP/GaAs DHBTs is only reported in [4] and [105], where both have used the doping spike design, but no direct comparison with an SHBT with a similar specification has been made. In this section, the DC and RF performance of InGaP/GaAs SHBT and doping spike DHBT are demonstrated and directly compared.

The two structures studied in this section are specified in Table 11.3. These structures are almost similar to the SHBT and DHBT-C in Table 11.2. But, there are some differences between these HBTs and the corresponding ones in Table 11.2, which are either forced by the
limitations of the growth system, or are deliberately introduced to improve the performance. The InGaAs cap layers, which are included in order to improve the emitter ohmic contact quality, could not be doped to more than $1 \times 10^{18}$ cm$^{-3}$ in the existing MOCVD growth system. Despite this, the specific contact resistance of the emitter is significantly improved due to the narrower bandgap of In$_{0.5}$Ga$_{0.5}$As compared to the GaAs cap layer. The n$^-$-InGaP collector of the DHBT is more highly doped than the GaAs collector of the SHBT. This, together with the doping spike present in the collector of the DHBT, is expected to shrink the depletion layer width, and compensate for the lower velocity of electrons inside InGaP as compared to GaAs. Finally, the base layer is much lower doped than the level necessary for microwave HBTs (i.e., $4 \times 10^{19}$ cm$^{-3}$). This is again due to an unoptimised growth system in which the DC current gain was significantly degraded for base doping levels of $4 \times 10^{19}$ cm$^{-3}$ or higher. State-of-the-art epitaxial growth facilities are capable of growing base layers doped in excess of $10^{20}$ cm$^{-3}$ [66] with the DC current gains comparable to those obtained in the present work for $N_{AB} = 2.7 \times 10^{19}$ cm$^{-3}$. Nevertheless, a comparative study to show the advantages of DHBTs over SHBTs is more of a concern in the present work than producing state-of-the-art results.

Table 11.3- Layer structure of the single- and double-HBTs experimentally studied in this section.

<table>
<thead>
<tr>
<th>Layer</th>
<th>SHBT (MR1106)</th>
<th>DHBT (MR1108)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cap Layer</td>
<td>300Å n$^-$-In$<em>{0.5}$Ga$</em>{0.5}$As $1 \times 10^{18}$ cm$^{-3}$</td>
<td>300Å n$^-$-In$<em>{0.5}$Ga$</em>{0.5}$As (x = 1 $\rightarrow$ 0.5) $1 \times 10^{18}$ cm$^{-3}$</td>
</tr>
<tr>
<td></td>
<td>1000Å n$^+$-GaAs $5 \times 10^{18}$ cm$^{-3}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>300Å n$^+$-InGaP $2 \times 10^{18}$ cm$^{-3}$</td>
<td></td>
</tr>
<tr>
<td>Emitter</td>
<td>1000Å n- InGaP $5 \times 10^{17}$ cm$^{-3}$</td>
<td></td>
</tr>
<tr>
<td>Base</td>
<td>800Å p$^+$-GaAs $2.7 \times 10^{19}$ cm$^{-3}$</td>
<td>300Å</td>
</tr>
<tr>
<td>n$^-$-GaAs spacer</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>n$^+$-InGaP doping spike</td>
<td>$2 \times 10^{18}$ cm$^{-3}$</td>
<td></td>
</tr>
<tr>
<td>n$^-$-collector</td>
<td>5000Å GaAs $3.5 \times 10^{16}$ cm$^{-3}$ (†)</td>
<td>4650Å InGaP $6 \times 10^{16}$ cm$^{-3}$ (†)</td>
</tr>
<tr>
<td>Sub-collector</td>
<td>7000Å n$^+$-GaAs $5 \times 10^{18}$ cm$^{-3}$</td>
<td></td>
</tr>
</tbody>
</table>

Semi-insulating GaAs Substrate

(*) Estimated from base sheet resistance of (380 ± 20) Ω/sq. using Fig. 6.6.
(†) Estimated from the capacitance-voltage measurements.

11.3.1. DC Results

The measured DC results of large-geometry (100 μm diameter, circular emitter) HBTs from wafers MR1106 and MR1108 are compared in this section. In order to demonstrate the effectiveness of the InGaAs cap layer in reducing the specific resistance of the emitter ohmic contact, DC results of the HBTs prior to annealing are presented. Figure 11.5 shows the Gummel plots of the two HBTs. It can be observed that both devices have similar turn-on voltages and ideality factors. The DC current gain of the DHBT is almost 17% smaller than that of the SHBT (Fig. 11.6). The peak current gain in the DHBT and SHBT reaches 29 and 35, respectively. It is also important to observe in Fig. 11.6 that the current gain of the doping spike DHBT does not change with the variation of $V_{CB}$ (or $V_{CE}$). This confirms that the carrier blocking effect of the B-C conduction band spike is negligible.
Chapter 11

Optimisation of the B-C Heterostructure of InGaP/GaAs DHBTs

Figure 11.5- Gummel plots of the single- and double HBTs specified in Table 11.3. Devices are large-geometry (100 µm emitter diameter) and unalloyed.

Figure 11.6- Variation of DC current gain with $I_C$ for the single- and double-HBTs specified in Table 11.3 for two different conditions: $V_{CB} = 0$ and $V_{CE} = 4.0$ V. Devices are large-geometry (100 µm emitter diameter) and unalloyed.

Figure 11.7- Common-emitter output characteristics of the single- and double-HBTs studied in this section for (a) low, and (b) high base current levels. Devices are large-geometry (100 µm emitter diameter) and unalloyed.

Figure 11.7 shows the common-emitter output characteristics of the two devices at low and high base current levels. First, as shown in Fig. 11.7(a), the breakdown voltage, $B/V_{CEO}$, of the DHBT (12.8 V) is almost 45% higher than that of the SHBT (8.8 V), despite the higher doping and larger electric field inside the collector of the former device. Second, both the offset voltage ($V_{CE,offset}$) and saturation voltage ($V_{CE, sat}$) of the DHBT are smaller than those of the SHBT. The measured offset voltages (at $I_B = 0.1$ mA) were 40 and 110 mV for the DHBT and SHBT, respectively. The $V_{CE, offset}$ and $V_{CE, sat}$ of both devices reduced even further after alloying, due to the reduction of $R_E$ and $R_C$. The measured values for the DHBT after alloying were $V_{CE, offset} = 32$ mV and $V_{CE, sat} < 0.5$ V. The larger breakdown voltage and
smaller offset voltage of the DHBTs make them more suitable for high power and high efficiency microwave power amplifiers [100].

11.3.2. High Frequency Performance

Small-geometry HBTs were fabricated using the new mask set (see chapter 8) on both wafers, MR1106 and MR1108. The measured common-emitter current gain, $h_{21}$, and unilateral power gain, $U$, on devices of $6 \times 10$ μm$^2$ emitter size are shown as a function of frequency in Fig. 11.8. Both devices show similar cutoff frequencies of 50 GHz, while the $f_{\text{max}}$ of the DHBT (47 GHz) is only slightly lower than that of the SHBT (50 GHz).

![Figure 11.8- Variation of common-emitter current gain, $h_{21}$, and unilateral power gain, $U$, with frequency for (a) InGaP/GaAs SHBT (MR1106) and (b) InGaP doping spike DHBT (MR1108). Both devices have B-E and B-C areas of $6 \times 10$ and $14 \times 10$ μm$^2$, respectively.](image)

Parameter extraction technique of chapter 10 was then applied to both HBTs to extract their small-signal equivalent circuit elements. The results corresponding to the peak $f_T$ performance of the two HBTs (i.e., the same biasing condition as in Fig. 11.8) are summarised in Table 11.4. Majority of the parameters are more or less similar in the two HBTs. Even the parasitic pad capacitances and interconnect inductances have similar magnitudes, since the two HBTs have been fabricated in the same run with an identical mask set. The observed differences are mainly related to the collector parameters. Since the collector is more highly doped in the case of the DHBT, depletion width is narrower, leading to a higher $C_{bc}$ but smaller $\tau_C$. As a result, $\tau_C$ is 0.11 ps smaller in the DHBT than in the SHBT, although electrons move slower inside InGaP than GaAs. Another factor in favour of the DHBT is the collector current at the peak $f_T$. Again the higher collector doping in the DHBT allows larger collector currents before the onset of the Kirk effect. Therefore, $R_{\text{be}}$ and the emitter charging time are smaller in this case. The smaller $R_{\text{b}bi}$ of the DHBT, though, is a bit unexpected, since the measured base sheet resistance of this wafer was actually slightly larger than the SHBT wafer.
## Table 11.4 - Small-signal parameters of Fig. 10.4 for InGaP/GaAs single- and double HBTs as specified in Table 11.3 with B-E and B-C areas of 6×10 and 14×10 μm², respectively.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>SHBT (MR1106)</th>
<th>DHBT (MR1108)</th>
<th>Parameter</th>
<th>SHBT (MR1106)</th>
<th>DHBT (MR1108)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ic [mA]</td>
<td>28</td>
<td>35</td>
<td>V_CB [V]</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>C_Pbe [F]</td>
<td>22</td>
<td>20</td>
<td>Lb [pH]</td>
<td>39</td>
<td>45</td>
</tr>
<tr>
<td>C_Pbe [F]</td>
<td>13</td>
<td>13</td>
<td>Lc [pH]</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td>C_Pee [F]</td>
<td>52</td>
<td>55</td>
<td>L_e [pH]</td>
<td>50</td>
<td>60</td>
</tr>
<tr>
<td>R_bb [Ω]</td>
<td>1.0</td>
<td>0.8</td>
<td>R_ce [Ω]</td>
<td>7.6</td>
<td>6.8</td>
</tr>
<tr>
<td>R_bbb [Ω]</td>
<td>8.9</td>
<td>8.5</td>
<td>R_bbb [kΩ]</td>
<td>~20</td>
<td>~15</td>
</tr>
<tr>
<td>R_cc [Ω]</td>
<td>9.8</td>
<td>9.9</td>
<td>R_be [kΩ]</td>
<td>~20</td>
<td>~15</td>
</tr>
<tr>
<td>C_ibe [F]</td>
<td>107</td>
<td>112</td>
<td>C_bc [F]</td>
<td>60</td>
<td>75</td>
</tr>
<tr>
<td>r</td>
<td>0.43</td>
<td>0.43</td>
<td>α₀</td>
<td>0.967</td>
<td>0.958</td>
</tr>
<tr>
<td>τ_B [ps]</td>
<td>1.02</td>
<td>1.00</td>
<td>τ_C [ps]</td>
<td>0.90</td>
<td>0.79</td>
</tr>
<tr>
<td>fₚ [GHz]</td>
<td>50</td>
<td>50</td>
<td>f_max [GHz]</td>
<td>50</td>
<td>47</td>
</tr>
</tbody>
</table>

The slightly smaller \( f_{\text{max}} \) in the case of the DHBT is mainly related to the larger \( C_{bc} \) in this device. Another point worthy of mentioning is that the base transit time of both devices are similar. This once again proves that the carrier blocking effect of the B-C heterojunction spike of the DHBT is effectively eliminated. The thinner base (800Å) of the devices studied in this section has resulted in an almost 0.5 ps reduction in \( \tau_B \) compared to the SHBT in Table 11.1 (which has a 1000Å base). Also the use of the InGaAs cap layer has reduced the emitter specific contact resistance to \( \sim 2\times 10^{-6}\ \Omega \cdot \text{cm}^2 \) from \( \sim 6.4\times 10^{-6}\ \Omega \cdot \text{cm}^2 \) for devices with GaAs cap layers (see subsection 10.5.2).

Variation of the measured cutoff frequencies with collector current for the single- and double-HBTs studied in this section is shown in Fig. 11.9. Reduction of \( f_T \) at low current levels is mainly due to the enlargement of the B-E dynamic resistance, \( R_{be} \), and thus the emitter charging time. The fall-off of \( f_T \) at higher currents is due to a combination of Kirk effect and device self-heating, as discussed in subsection 10.5.1. Since the collector doping is higher in the DHBT, the onset of the Kirk effect occurs at larger current levels in this device. Figure 11.10 shows the variation of \( f_{\text{max}} \) with \( V_{CB} \) for both devices at the current levels corresponding to the peak \( f_T \). As \( V_{CB} \) increases, the B-C depletion region expands. Consequently, \( \tau_C \) increases, but \( C_{bc} \) gets smaller. Since \( f_{\text{max}} \) is mostly sensitive to \( C_{bc} \), it continues increasing with \( V_{CB} \) until the device self-heating degrades the high frequency performance. As a result, \( f_{\text{max}} \) goes through a broad peak in the medium range of \( V_{CB} \).

The results in this section have demonstrated that with careful design of the B-C heterojunction and the collector region, the high frequency performance of the InGaP/GaAs DHBTs can be made similar to that of the SHBTs, while the former still benefit from a higher breakdown voltage and a lower offset/saturation voltage. DHBTs with doping spike design have been experimentally tested in this section, and the use of the InGaAsP step-grading is expected (from numerical calculations) to increase the peak \( f_T \) by another 5-10 GHz. With an

287
optimised growth system, the base doping concentration can be increased to \( \sim 1 \times 10^{20} \text{ cm}^{-3} \), while keeping a useful DC current gain of 20-30 [66]. The higher base doping is in fact shown [111,246,263] to increase the minority electron mobility inside the base, as discussed in subsection 5.4.2. Therefore, both \( f_T \) and \( f_{\text{max}} \) of the HBTs would benefit from higher base doping levels, although reliability issues of highly doped base regions are of some concern. Finally, the InGaAsP step-graded DHBT proposed in section 11.2 may be easily combined with the graded base (using In\(_{1-x}\)Ga\(_x\)As) design analysed in subsection 7.5.1 to further improve the base transit time.

![Graphs showing variation of \( f_T \) and \( f_{\text{max}} \) with collector current and collector bias voltage.](image)

**Figure 11.9**- Variation of \( f_T \) with \( I_C \) for the single- and double-HBTs specified in Table 11.3.

**Figure 11.10**- Variation of \( f_{\text{max}} \) with \( V_{CB} \) for the single- and double-HBTs specified in Table 11.3.
CHAPTER 12. SUMMARY AND SUGGESTIONS FOR FUTURE WORK

This thesis can be distinguished from similar studies of heterojunction bipolar transistors in the following areas:

- A new small-signal parameter extraction method is introduced which accurately determines the equivalent circuit elements of HBTs, without reference to numerical optimisation. Very reasonable assumptions are used in the course of parameter extraction, which make the method applicable to a wide range of HBTs (Chap. 10).

- Novel techniques are presented to accurately determine the total delay time and its components in HBTs from the measured s-parameter data at relatively low frequencies (section 10.4).

- Since the new parameter extraction method produces physically accurate values of elements, it was utilised in an study to observe the effects of some high current phenomena (such as Kirk effect, device self-heating, and current crowding) on HBT parameters (subsection 10.5.1).

- Physical properties of a wide range of III-V binary, ternary, and quaternary compounds are extensively studied. Very useful empirical relations are provided to model the variations of important parameters such as energy bandgaps, carrier mobilities, minority carrier lifetimes, electron affinity, and the amount of bandgap narrowing as a function of doping concentration, temperature, electric field, etc. This is, to the best of author's knowledge, the most comprehensive compilation of electrical parameters for various types of III-V semiconductors (Chap. 5).

- A numerical device simulation program is developed in this work which takes into account many of the important physical phenomena for HBT, such as bandgap narrowing, Fermi-Dirac statistics, and thermionic-field emission (TFE) of carriers across an abrupt heterojunction interface. The latter is mainly emphasised in this work, where a very general boundary condition for all shapes of band structures near an abrupt heterojunction is developed. This TFE boundary condition even can be used for graded heterojunctions as well as homojunctions (Chap. 4, Appendix B, and section 7.2).

- The advantages of InGaP/GaAs DHBTs over both InGaP and AlGaAs SHBTs in terms of their DC characteristics are demonstrated using both the experimental data and numerical simulations (Chap. 9).

- Based on the new small-signal parameter extraction technique and the numerical simulation results, the B-C heterojunction of InGaP/GaAs DHBTs is optimised for improved high frequency performance. A new step-graded heterojunction using the quaternary material InGaAsP lattice-matched to GaAs is proposed to improve the carrier blocking effect of the B-C heterojunction. The DC and high frequency performance of fabricated InGaP/GaAs single- and double-HBTs are directly compared to demonstrate that
the DHBT has a high frequency performance almost as good as the SHBT, while the former benefits from a larger breakdown voltage and a smaller offset voltage (Chap. 12).

- Variation of HBT DC parameters with temperature are thoroughly studied. Temperature dependence of current gain in various types of HBTs is explained with the aid of numerical simulation. Very good agreement can be obtained between the numerical results and the measured data. In particular, the less temperature sensitivity of current gain in heterostructure-emitter bipolar transistors (HEBTs), which is recently reported in the literature [155-156], is properly explained in the present work (sections 7.4 and 9.2.6).
- The effect of base grading, using both Al\(_x\)Ga\(_{1-x}\)As and In\(_{1-x}\)Ga\(_x\)As, on the DC current gain and cutoff frequency of GaAs HBTs is studied in details using the numerical model. All the factors contributing to the improvement or degradation of performance are clearly identified. It is shown that using a small percentage of In or Al in the base layer of abrupt-emitter InGaP/GaAs or graded-emitter AlGaAs/GaAs HBTs can significantly improve the DC and RF performance. Using a graded AlGaAs base for abrupt-emitter AlGaAs/GaAs HBTs, on the other hand, does not seem to cause much improvement. The InGaP/(In)GaAs graded-base HBT certainly seems to be a promising device (section 7.5).
- The author has also contributed to the development of a planar fabrication process and design of a new mask set for microwave HBTs (section 8.4).
- A new geometry dependence of current gain in H\(^+\)-implanted HBTs is observed and explained in terms of formation of C-H complexes in selective areas of the device. The use of He\(^+\) ions instead of the conventional protons is shown to resolve this unwanted phenomenon (subsection 8.4.1).
- The sharp rise in the base current (i.e., the \textit{kink}) observed in the Gummel plot of some HBTs is properly explained in terms of the series resistances of the device. Then, its occurrence is used in a novel technique to extract the collector and base series resistances of the device using only the measured DC data (subsection 9.1.4).
- Various types of forward and reverse Gummel plots and their relation with the design of collector-up HBTs are studied in details, and a special new type of reverse Gummel plot is observed and explained here (subsection 9.1.5).
- The observed dependence of offset voltage on bias, vertical structure, and horizontal layout of HBTs are explained using a simple theory. Although the theory employed here is not new, a good agreement between this theory and experimental data for the variation offset voltage with I\(_B\) has not been demonstrated in the original work [408] (subsection 9.1.3).

This work, therefore, significantly contributes towards further understanding and improvement of HBTs for microwave power applications. The following are suggestions for future work in this area of research:

**Large-Signal Modelling and Characterisation:**

- The small-signal model and its associated parameter extraction technique, as discussed in
chapter 10, should be modified to a large-signal model [477-479].

- Large-signal parameters such as maximum output power, power-added efficiency, and intermodulation distortion should be measured. Appropriate thermal management would be a critical issue in obtaining good power performance.

**Modifications to the Numerical Simulation Program:**

- Impact ionisation has to be included as one of the sources of carrier generation in order to simulate the HBTs in the avalanche breakdown regime [480-482].
- We have already noticed that improper design of the doping spike in the collector of DHBTs may result in an early Zener breakdown (see Fig. 3.3). Zener breakdown, which is due to the band-to-band tunnelling of carriers, can be regarded as a recombination-generation component. Methods to include this mechanism into numerical simulations are suggested in [483-484].
- The effects of strain on the band structure [485-486], carrier effective masses, and mobilities [387] have to be included in the program in order to accurately simulate graded In$_{1-x}$Ga$_x$As base HBTs.
- Self-heating is an important phenomenon for power HBTs. It can be included in the calculations by solving the heat conduction equation [48,279,329,382]. However, for a first order analysis, the total power dissipation (i.e., $V_{CE}I_C + V_{BE}I_B$) together with an estimate of the thermal resistance can be used at the end of each loop of iteration to update the lattice temperature for next iteration run.

**Device Design and Fabrication:**

- The InGaAsP step-graded design for InGaP/GaAs DHBTs can be experimentally verified.
- Graded In$_{1-x}$Ga$_x$As base HBT is certainly a promising area. This idea can be easily combined with the step-graded collector of DHBTs.
- Fully self-aligned HBTs (i.e., with the emitter metal used as the mask for the emitter mesa etch) should be fabricated to take full advantage of the potentials of HBTs for high speed.
- Finally, wide bandgap GaN- and SiC-based HBTs are new and very promising candidates for microwave power applications. A recent special issue of the *Solid-State Electronics* journal (February 2000) on this topic can be a good starting point for research into this field.
APPENDIX A
NETWORK PARAMETERS AND RELATIONSHIPS

Linear networks, or non-linear networks operating with signals sufficiently small to cause the network to respond in a linear manner, can be completely characterised by parameters measured at the network terminals without considering the contents of the network. Once the parameters of a network have been determined, its behaviour in any external environment can be predicted, again regardless of the contents of the network.

The network parameters can be represented in several forms associated with the input and output variables being represented. Admittance parameters (y-parameters), for example, are associated with the currents being represented as a function of potentials; impedance parameters (z-parameters) are associated with the voltages being written as a function of currents; hybrid parameters (h-parameters) are a combination of these; and finally scattering parameters (s-parameters) relate power waves.

For a two-port linear network as shown in Fig. A.1 the voltages at the two ports can be represented as linear functions of the currents as:

\[ V_1 = z_{11}I_1 + z_{12}I_2 \]
\[ V_2 = z_{21}I_1 + z_{22}I_2 \]

or in a matrix form as:

\[ V = z \cdot I \quad \text{with} \quad V = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}, \quad I = \begin{bmatrix} I_1 \\ I_2 \end{bmatrix}, \quad z = \begin{bmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{bmatrix} \] (A.1)

Similarly, one can write:

\[ I = y \cdot V \quad \text{with} \quad y = \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix} \] (A.2)

\[ \begin{bmatrix} V_1 \\ I_2 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} I_1 \\ V_2 \end{bmatrix} \] (A.3)

In terms of the current vector \( I \), the voltage vector \( V \), and the characteristic impedance \( Z_0 \), the power waves are defined as:

\[ a = \frac{V + Z_0 I}{2\sqrt{Z_0}} \] (A.4)
\[ b = \frac{V - Z_0 I}{2\sqrt{Z_0}} \]
the \( s \)-parameters then follow from the definition:

\[
\mathbf{b} = \mathbf{s} \cdot \mathbf{a} \quad (A.5)
\]

The ease with which scattering parameters can be measured makes them especially well suited for describing transistors and other active devices. Measuring most other parameters calls for the input and output of the device to be successively opened and short circuited. This can be hard to do, especially at microwave frequencies where lead inductance and capacitance make short and open circuits difficult to obtain. \( s \)-parameters, on the other hand, are usually measured with the device embedded between a 50\( \Omega \) load and source, and there is no need to produce open or short circuits.

Experimental measurements usually provide \( s \)-parameter data, while device analysis is best carried out using parameters based on currents and voltages. The relationships between the \( s \)-parameters and other parameters can be directly derived from (A.1)-(A.5) and are summarised in Tables A.1-A.4.

The stability of a network can be described by the Rollet stability factor, \( K \) [457,487]:

\[
K = \frac{1 - |s_{11}|^2 - |s_{22}|^2 + |D|^2}{2|s_{12}s_{21}|} \quad (A.6)
\]

where

\[
D = s_{11}s_{22} - s_{12}s_{21} \quad (A.7)
\]

It can be shown that the necessary and sufficient conditions for inherent stability of the network are [457,487]:

\[
\begin{cases} 
K > 1 \\
B_1 = 1 + |s_{11}|^2 - |s_{22}|^2 - |D|^2 > 0
\end{cases} \quad (A.8)
\]

Under above conditions, the maximum available gain (MAG) of the network can be determined as [456-457]:

\[
\text{MAG} = \frac{|s_{21}|}{|s_{12}|} \sqrt{K^2 - 1} \quad (A.9)
\]

However, if any of the conditions (A.8) fails, the network is potentially unstable and a maximum stable gain (MSG) can be defined:

\[
\text{MSG} = \frac{|s_{21}|}{|s_{12}|} = \frac{|z_{21}|}{|z_{12}|} = \frac{|y_{21}|}{|y_{12}|} = \frac{|h_{21}|}{|h_{12}|} \quad (A.10)
\]

It is worth noting that always \( \text{MAG} \leq \text{MSG} \) and under the condition \(-1 \leq K \leq 1\), MAG will be equal to MSG.
Appendix A  
Network Parameters and Relationships

The maximum power gain when a two-port device has been made unilateral using only lossless reciprocal elements is called the unilateral power gain \((U)\). The unilateral power gain is originally defined and expressed by Mason \([463]\) as:

\[
U = \frac{|z_{21} - z_{12}|^2}{4[\text{Re}(z_{11}) \cdot \text{Re}(z_{22}) - \text{Re}(z_{12}) \cdot \text{Re}(z_{21})]}
\]

\[
= \frac{|y_{21} - y_{12}|^2}{4[\text{Re}(y_{11}) \cdot \text{Re}(y_{22}) - \text{Re}(y_{12}) \cdot \text{Re}(y_{21})]}
\]

The maximum frequency of oscillation \((f_{\text{max}})\) is defined as the frequency at which either \(\text{MAG}\) or \(U\) becomes unity. The cut-off frequency \((f_c)\) is the frequency where \(|h_{21}|\) reduces to one.

### Table A.1 - The relationships between \(s\)-parameters and \(z\)-parameters.

<table>
<thead>
<tr>
<th>(s)-parameters in terms of (z)-parameters</th>
<th>(z)-parameters in terms of (s)-parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>(s_{11} = \frac{(z_{11} - Z_0)(z_{22} + Z_0) - z_{12}z_{21}}{(z_{11} + Z_0)(z_{22} + Z_0) - z_{12}z_{21}})</td>
<td>(z_{11} = Z_0 \cdot \frac{(1 + s_{11})(1 - s_{22}) + s_{12}s_{21}}{(1 - s_{11})(1 - s_{22}) - s_{12}s_{21}})</td>
</tr>
<tr>
<td>(s_{12} = \frac{2z_{12}Z_0}{(z_{11} + Z_0)(z_{22} + Z_0) - z_{12}z_{21}})</td>
<td>(z_{12} = Z_0 \cdot \frac{2s_{12}}{(1 - s_{11})(1 - s_{22}) - s_{12}s_{21}})</td>
</tr>
<tr>
<td>(s_{21} = \frac{2z_{21}Z_0}{(z_{11} + Z_0)(z_{22} + Z_0) - z_{12}z_{21}})</td>
<td>(z_{21} = Z_0 \cdot \frac{2s_{21}}{(1 - s_{11})(1 - s_{22}) - s_{12}s_{21}})</td>
</tr>
<tr>
<td>(s_{22} = \frac{(z_{11} + Z_0)(z_{22} - Z_0) - z_{12}z_{21}}{(z_{11} + Z_0)(z_{22} + Z_0) - z_{12}z_{21}})</td>
<td>(z_{22} = Z_0 \cdot \frac{(1 + s_{22})(1 - s_{11}) + s_{12}s_{21}}{(1 - s_{11})(1 - s_{22}) - s_{12}s_{21}})</td>
</tr>
</tbody>
</table>

### Table A.2 - The relationships between \(s\)-parameters and \(y\)-parameters. \((Y_0 = 1/Z_0)\)

<table>
<thead>
<tr>
<th>(s)-parameters in terms of (y)-parameters</th>
<th>(y)-parameters in terms of (s)-parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>(s_{11} = \frac{(Y_0 - y_{11})(Y_0 + y_{22}) + y_{12}y_{21}}{(Y_0 + y_{11})(Y_0 + y_{22}) - y_{12}y_{21}})</td>
<td>(y_{11} = Y_0 \cdot \frac{(1 + s_{12})(1 - s_{11}) + s_{12}s_{21}}{(1 + s_{11})(1 + s_{22}) - s_{12}s_{21}})</td>
</tr>
<tr>
<td>(s_{12} = \frac{-2y_{12}Y_0}{(Y_0 + y_{11})(Y_0 + y_{22}) - y_{12}y_{21}})</td>
<td>(y_{12} = Y_0 \cdot \frac{-2s_{12}}{(1 + s_{11})(1 + s_{22}) - s_{12}s_{21}})</td>
</tr>
<tr>
<td>(s_{21} = \frac{-2y_{21}Y_0}{(Y_0 + y_{11})(Y_0 + y_{22}) - y_{12}y_{21}})</td>
<td>(y_{21} = Y_0 \cdot \frac{-2s_{21}}{(1 + s_{11})(1 + s_{22}) - s_{12}s_{21}})</td>
</tr>
<tr>
<td>(s_{22} = \frac{(Y_0 + y_{11})(Y_0 - y_{22}) + y_{12}y_{21}}{(Y_0 + y_{11})(Y_0 + y_{22}) - y_{12}y_{21}})</td>
<td>(y_{22} = Y_0 \cdot \frac{(1 + s_{11})(1 - s_{22}) + s_{12}s_{21}}{(1 + s_{11})(1 + s_{22}) - s_{12}s_{21}})</td>
</tr>
</tbody>
</table>
**Table A.3** - The relationships between z-parameters and y-parameters.

<table>
<thead>
<tr>
<th>z-parameters</th>
<th>y-parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>in terms of y-parameters</td>
<td>in terms of z-parameters</td>
</tr>
<tr>
<td>$z_{11} = \frac{y_{22}}{y_{11}y_{22} - y_{12}y_{21}}$</td>
<td>$y_{11} = \frac{z_{22}}{z_{11}z_{22} - z_{12}z_{21}}$</td>
</tr>
<tr>
<td>$z_{12} = \frac{-y_{12}}{y_{11}y_{22} - y_{12}y_{21}}$</td>
<td>$y_{12} = \frac{-z_{21}}{z_{11}z_{22} - z_{12}z_{21}}$</td>
</tr>
<tr>
<td>$z_{21} = \frac{-y_{21}}{y_{11}y_{22} - y_{12}y_{21}}$</td>
<td>$y_{21} = \frac{-z_{21}}{z_{11}z_{22} - z_{12}z_{21}}$</td>
</tr>
<tr>
<td>$z_{22} = \frac{y_{11}}{y_{11}y_{22} - y_{12}y_{21}}$</td>
<td>$y_{22} = \frac{z_{11}}{z_{11}z_{22} - z_{12}z_{21}}$</td>
</tr>
</tbody>
</table>

**Table A.4** - The relationships between s-parameters and h-parameters. ($Y_0 = 1/Z_0$)

<table>
<thead>
<tr>
<th>s-parameters in terms of h-parameters</th>
<th>h-parameters in terms of s-parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_{11} = \frac{(h_{11} - Z_0)(h_{22} + Y_0) - h_{12}h_{21}}{(h_{11} - Z_0)(h_{22} + Y_0) - h_{12}h_{21}}$</td>
<td>$h_{11} = Z_0 \cdot \frac{(1 + s_{11})(1 + s_{22}) - s_{12}s_{21}}{(1 - s_{11})(1 + s_{22}) + s_{12}s_{21}}$</td>
</tr>
<tr>
<td>$s_{12} = \frac{2h_{12}}{(h_{11} + Z_0)(h_{22} + Y_0) - h_{12}h_{21}}$</td>
<td>$h_{12} = \frac{2s_{12}}{(1 - s_{11})(1 + s_{22}) + s_{12}s_{21}}$</td>
</tr>
<tr>
<td>$s_{21} = \frac{-2h_{21}}{(h_{11} + Z_0)(h_{22} + Y_0) - h_{12}h_{21}}$</td>
<td>$h_{21} = \frac{-2s_{21}}{(1 - s_{11})(1 + s_{22}) + s_{12}s_{21}}$</td>
</tr>
<tr>
<td>$s_{22} = \frac{(Z_0 + h_{11})(Y_0 - h_{22}) + h_{12}h_{21}}{(h_{11} + Z_0)(h_{22} + Y_0) - h_{12}h_{21}}$</td>
<td>$h_{22} = Y_0 \cdot \frac{(1 - s_{11})(1 - s_{22}) - s_{12}s_{21}}{(1 - s_{11})(1 + s_{22}) + s_{12}s_{21}}$</td>
</tr>
</tbody>
</table>
This appendix presents more details regarding the numerical evaluation of the tunnelling integral (4.30), also repeated here in a slightly more general form:

$$\delta_L = \frac{1}{kT} C_{\text{max}} \int_{E_{\text{min}}}^{E_{\text{max}}} \exp \left( \frac{E_{\text{max}} - E_z}{kT} \right) \cdot \exp \left( \frac{-4\pi}{\hbar} \int_{z_1}^{z_2} \left| V \right|^2 n(z) \cdot dz \right) \cdot dE_z \quad (B.1)$$

where $\delta_L$ is tunnelling factor for the interface between layers $L$ and $L+1$, and $z_1$ and $z_2$ are defined by $E_C(z_1) = E_C(z_2) = E_z$, $z_1 \leq z_j \leq z_2$. The band structure in the vicinity of an interface can have one of the four general shapes shown in Fig. B.1. In this figure, $N_{\text{layer}}(L)$ and $N_{\text{layer}}(L+1)$ are the number of mesh points in layers $L$ and $L+1$, respectively, and $..., j-1, j, j+1, ...$ are the indices of the discretised mesh points. Also:

$$j = \sum_{\ell=1}^{L} N_{\text{layer}}(\ell) \quad (B.2)$$

Figure B.1- Four possible variation of conduction band near a heterojunction interface and definition of parameters in the tunnelling integral (B.1).
Appendix B  

Calculation of Tunnelling Integral

In cases (b) and (d) the tunnelling parameter \( \delta(L) \) for carriers transporting from layer L to layer L+1 (and vice versa) is non-zero. We consider the case (d) for our discussion.

In Fig. B.2(a), a conduction band diagram similar to that in Fig. B.1(d) is repeated with some more details relevant to our following discussion. First the minimum points of the conduction band on both layers are found, namely \( J_L \) and \( J_R \) in Fig. B.2(a). In cases of thick bulk materials, the points \( J_L \) and \( J_R \) are chosen to have a maximum distance of 400\( \AA \) from the interface, since the tunnelling probability of electrons through a barrier thicker than 400\( \AA \) is effectively zero. Then \( E_{\text{min}} \) and \( E_{\text{C,max}} \) are defined as:

\[
\begin{align*}
E_{\text{min}} &= \text{Max}\{E_{C,J_L}, E_{C,J_R}\} \\
E_{\text{C,max}} &= \text{Max}\{E_{C,j}, E_{C,j+1}\}
\end{align*}
\]  

(B.3)

As shown in figure B.2(a), the energy interval \([E_{\text{min}}, E_{\text{C,max}}]\) is divided into a finite number of energy steps. For each discretised energy level \( E_z \) the mesh points \( J_1 \) and \( J_2 \) are defined as:

\[
\begin{align*}
E_C(J_1) &\leq E_z < E_C(J_1 +1) \quad J_1 < j \\
E_C(J_2) &\leq E_z < E_C(J_2 -1) \quad J_2 > j
\end{align*}
\]  

(B.4)

Now the internal integral in (B.1) can be rewritten as:

\[
\begin{align*}
\int_{z(1)}^{z(1+1)} \sqrt{\bar{m}_n}(E_C(z) - E_z) \cdot dz &= \int_{z(1)}^{z(1+1)} \sqrt{\bar{m}_n}(J_1) \cdot (E_C(z) - E_z) \cdot dz + \\
&+ \int_{z(1)}^{z(2)} \sqrt{\bar{m}_n}(J_2 -1) \cdot (E_C(z) - E_z) \cdot dz \\
&+ \int_{z(1)}^{z(2)} \sqrt{\bar{m}_n}(J_2) \cdot (E_C(z) - E_z) \cdot dz = \\
&+ \sum_{i=J_1+1}^{J_2-2} \int_{z(i)}^{z(i+1)} \sqrt{\bar{m}_n}(i) \cdot (E_C(z) - E_z) \cdot dz + \\
&+ \int_{z(1)}^{z(2)} \sqrt{\bar{m}_n}(J_2 -2) \cdot (E_C(z) - E_z) \cdot dz \\
&+ \int_{z(2)}^{z(1+1)} \sqrt{\bar{m}_n}(J_1) \cdot (E_C(z) - E_z) \cdot dz
\end{align*}
\]  

(B.5)

where a constant effective mass is assumed for each interval between two neighbouring mesh points with the following definitions:

\[
\begin{align*}
\bar{m}_n(J_1) &= \left[m^*_n(J_1) + 3m^*_n(J_1 +1)\right]/4 \\
\bar{m}_n(i) &= \left[m^*_n(i) + m^*_n(i +1)\right]/2 \quad \text{for} \quad i = J_1 +1, \cdots, J_2 - 2 \\
\bar{m}_n(J_2 -1) &= \left[3m^*_n(J_2 -1) + m^*_n(J_2)\right]/4
\end{align*}
\]  

(B.6)

297
Appendix B

Calculation of Tunnelling Integral

Figure B.2- Details of the parameters and integration method for the (a) inner and (b) outer integral in equation (B.1).

If we further assume that the conduction band edges vary linearly between neighbouring mesh points (as is already assumed for the discretisation of the basic semiconductor equations in chapter 6), then it would be a straightforward task to prove that:

\[
\int_{z_1}^{z_2} \sqrt{\frac{m_n(i)}{m_n(i)}} \cdot (E_C(z) - E_z) \cdot dz = \frac{2}{3} \cdot \frac{z(i+1) - z(i)}{E_C(i+1) - E_C(i)} \cdot \sqrt{m_n(i)} \cdot \left[ (E_C(i+1) - E_z)^{3/2} - (E_C(i) - E_z)^{3/2} \right] \tag{B.7}
\]

and thus:

\[
\begin{align*}
\int_{z_1}^{z_2} \sqrt{m_n(i)} \cdot (E_C(z) - E_z) \cdot dz & = \frac{2}{3} \cdot \frac{z(J_1 + 1) - z(J_1)}{E_C(J_1 + 1) - E_C(J_1)} \cdot \sqrt{m_n(J_1)} \cdot \left[ (E_C(J_1 + 1) - E_z)^{3/2} - 0 \right] \\
+ \sum_{i=J_1+1}^{J_2-2} & \cdot \frac{2}{3} \cdot \frac{z(i+1) - z(i)}{E_C(i+1) - E_C(i)} \cdot \sqrt{m_n(i)} \cdot \left[ (E_C(i+1) - E_z)^{3/2} - (E_C(i) - E_z)^{3/2} \right] \\
+ \frac{2}{3} \cdot \frac{z(J_2) - z(J_2 - 1)}{E_C(J_2) - E_C(J_2 - 1)} \cdot \sqrt{m_n(J_2 - 1)} \cdot \left[ 0 - (E_C(J_2 - 1) - E_z)^{3/2} \right] \tag{B.8}
\end{align*}
\]

The sigma term in (B.8) is evaluated only if \( J_2 \geq J_1 + 3 \). The reader will notice that applying the above method of integration has let us avoid the troublesome process of locating the integration limits, \( z_1 \) and \( z_2 \), for each discretised energy level, \( E_z \).

Using the above method to calculate the result of the inner integral in (B.1), the integrand of the outer integral would be evaluated for each energy level \( E_z \), and denoted as \( f(E_z) \). Then the problem of calculating the tunnelling parameter in (B.1) will be reduced to finding the area under the curve \( f(E_z) \) versus \( E_z \) (shown in Fig. B.4(b)) using, for example, a standard trapezoidal integration method.

298
APPENDIX C. ANALYTICAL APPROXIMATIONS FOR FERMI-DIRAC INTEGRALS

The Fermi-Dirac integral of order $j$ is defined as:

$$F_j(\eta) = \frac{1}{\Gamma(j+1)} \int_0^\infty \frac{x^j}{1 + \exp(x - \eta)} \, dx$$ \hspace{1cm} (C.1)

where $\Gamma$ is the Gamma function with the following properties:

$$\Gamma(1/2) = \sqrt{\pi}, \quad \Gamma(j+1) = j \cdot \Gamma(j)$$

Numerical calculation of the integral in (C.1) is a time consuming process, bearing in mind that this process needs to be repeated thousands of times during the numerical calculations for just one bias point. Therefore, analytical approximations have been suggested for evaluation of this integral with small percentages of error in the whole range of $\eta$. The most useful orders of this integral are $(1/2)$ and $(-1/2)$, the former appears directly in Eqs. (4.11) and (4.12), while the application of the latter is in the calculation of the elements of the Jacobian matrix (see section 6.? for details) through the following property of the Fermi-Dirac integrals:

$$\frac{d}{d\eta} F_j(\eta) = F_{j-1}(\eta)$$ \hspace{1cm} (C.2)

The analytical approximations for these two integrals are in the form [161,488]:

$$F_{\pm 1/2} \approx \left[ \exp(-\eta) + \xi_{\pm 1/2}(\eta) \right]^{-1}$$ \hspace{1cm} (C.3)

where:

$$\xi_{+1/2} = 3\sqrt{\pi} / 2 \cdot \left[ (\eta + 2.13) + (|\eta - 2.13|^2 + 9.6)^{5/12} \right]^{-3/2}$$ \hspace{1cm} (C.4)

with a maximum error of ±0.5% [161], and:

$$\xi_{-1/2} = \sqrt{\pi} / 2 \cdot \left[ (\eta + 1.63) + (|\eta - 1.63|^2 + 9.74)^{1/4.05} \right]^{-1/2}$$ \hspace{1cm} (C.5)

with a maximum error of ±0.76% [488].

In addition to the Fermi-Dirac integrals, we need to have an analytical approximation for the inverse function of $F_{1/2}$. The following approximation for $F_{1/2}^{-1}$ exists in the literature [161]:

$$F_{1/2}^{-1}(u) \approx \frac{\ln(u)}{1 - u^2} + \frac{(3\sqrt{\pi} \cdot u / 4)^{2/3}}{1 + \left[ 0.24 + 1.08 \left( 3\sqrt{\pi} \cdot u / 4 \right)^{2/3} \right]^{-2}}$$ \hspace{1cm} (C.6)
with a maximum error of ±0.5%. In the vicinity of $u = 1$, $\frac{\ln(u)}{1-u^2}$ can be approximated as $\frac{u - 3}{2(u + 1)}$. 
LIST OF PUBLICATIONS

Journal Articles


Conference Publications


301
REFERENCES


References


References


[76] Q. J. Hartmann, H. Hwangbo, A. Yung, D. A. Ahmari, M. T. Fresina, J. E. Baker, and G. E. Stillman, "Removal of hydrogen from the base of carbon-doped In_{0.49}Ga_{0.51}P/GaAs


References


References


[150] K. H. Bachem, T. Lauterbach, M. Maier, W. Pletschen, and K. Winkler, “MOVPE growth, technology and characterization of Ga0.5In0.5P/GaAs heterojunction bipolar transistors,” Inst. of Phys. Conf. Ser., no. 120, chapter 6, pp. 293-298, 1992.


References


References


[210] A. B. Chen, and A. Sher, “Electronic structure of pseudobinary semiconductor alloys AlxGa1-xAs, GaP0.5As1-0.5, and Ga0.5In0.5P,” Phys. Rev. B, vol. 23, pp. 5360-5374, 1981.


References


[221] Z. Hang, H. Shen, and F. H. Pollak, “Temperature dependence of the E_{0} and E_{0} + \Delta_{0} gaps of InP up to 600°C,” Solid State Comm., vol. 73, pp. 15-18, 1990.


[258] R. Dittrich, and W. Schroeder, “Empirical pseudopotential band structure of In$_{0.53}$Ga$_{0.47}$As and In$_{0.52}$Al$_{0.48}$As,” Solid-State Electron., vol. 43, pp. 403-407, 1999.


References


[294] III-V Semiconductor Centre, University of Sheffield, UK (private communication).

[295] Epitaxial Products International Ltd., Cardiff, UK (private communication).


[298] M. Ikeda, and K. Kaneko, “Selenium and zinc doping in Ga0.5In0.5P and (Al0.5Ga0.5)0.5In0.5P grown by metalorganic chemical vapor deposition,” *J. Appl. Phys.*, vol. 66, pp. 5285-5289, 1989.
References


[301] J. B. Lee, S. D. Kwon, I. Kim, Y. H. Cho, and B-D. Choe, “The characteristics of an In0.5Ga0.5P and In0.52Ga0.48P/GaAs heterojunction grown on a (100) GaAs substrate by liquid-phase epitaxy,” J. Appl. Phys., vol. 71, pp. 5016-5021, 1992.


References


References


References


322


References


324
References


