INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6” x 9” black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
DESIGN AND FABRICATION
OF AlGaAs/GaAs HETERO-JUNCTION BIPOLAR TRANSISTORS
WITH SUB-MICRON PASSIVATION LEDGES

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF
MATERIALS SCIENCE AND ENGINEERING
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Changhong Dai
February 1996
I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Professor James S. Harris, Jr.
(Principal Advisor)

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Professor Richard H. Bube

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Professor Robert W. Dutton

Approved for the University Committee
on Graduate Studies:
Abstract

AlGaAs/GaAs Hetero-junction Bipolar Transistors (HBTs) have attracted significant interest because of their potential in microwave, communication, and high speed digital applications. Due to lack of a stable native oxide to passivate the surface around the emitter perimeter, the fabrication of high reliability, high current gain, low 1/f noise HBTs was difficult until it was discovered that a fully depleted, thin AlGaAs passivation ledge over the otherwise exposed base region could eliminate base surface recombination. However, the use of such passivation ledges increases the base-collector junction capacitance and the extrinsic base resistance, both of which reduce the operational speed of HBTs. For sub-micron emitter size HBTs, which are desired for low power and high integration density applications, the speed is degraded even more by these parasitic components. Moreover, when the device size is scaled to sub-micron dimensions, the scalability of HBTs will be limited by the minimum dimension of these passivation ledges. To optimize HBT device design and study the scalability of HBTs, it is very important to determine the minimum ledge length required to produce full passivation, to develop processing approaches for fabricating ultra-short passivation ledges, and to demonstrate the feasibility of HBTs with sub-micron emitters and passivation ledges.

This dissertation studies the minimum passivation ledge length required to fully passivate AlGaAs/GaAs HBTs. The experimental and theoretical results show that this minimum ledge length is about three times the base thickness when there is no base grading and shorter when the base is graded. Three processing approaches for fabricating sub-micron passivation ledges as small as 0.1 μm were developed. In order to fabricate sub-micron HBTs, InGaAs/AlGaAs/GaAs plasma etching techniques, in particular, a silicon-tetrachloride based plasma etching process suitable for forming sub-micron emitters was developed. These processing approaches were used to fabricate specially designed RF HBTs with sub-micron passivation ledges and they yield a cut-off frequency and a maximum oscillation frequency greater than 60 GHz, which is adequate for most radio frequency and millimeter wave applications. In addition, this dissertation examines scaling rules in the miniaturization of HBTs and design tradeoffs in epitaxial layer design and device layout design.

The contributions discussed in this dissertation have not only produced an optimized HBT performance, but also resulted in a process technique which is more appropriate to commercial fabrication lines.
Acknowledgments

I am very grateful for help and support from a large number of people during my studies at Stanford. First and foremost I would like to thank my advisor, Professor James Harris for providing an outstanding research environment that gives students the academic freedom to pursue the research topics interesting to them and to collaborate with each other. In addition to his "coaching" on technical issues, such as device physics and MBE techniques, he is always very supportive and generous in offering his students industrial opportunities to make them more prepared for real world activities.

I would also like to thank Professors Richard Bube and Robert Dutton for reading my dissertation, Professor Shan X. Wang for serving on my orals committee, and Professor Charles Marcus for chairing my orals committee.

This work has benefited from many fellow students in the Harris Group. First of all, I would like to thank all the previous students working on HBT research -- William Liu, Damian Costa, Tony Ma, Won-Seong Lee, and Darrell Hill. Their excellent research results and very informative dissertations have provided me a starting point and saved me a great amount of time figuring all the device designs and processing conditions from scratch. Dr. William Liu encouraged me to work on this topic at the beginning of my dissertation research and has given me a lot of technical help throughout the project. I would also like to thank Alan Massengale for his collaboration in developing the high-frequency HBT process and measurement techniques. Through the years we have worked together on processing, measurement, and simulation as well. I have learned a lot from him. (Not to mention English slang, though I am not sure how much Chinese he learned from me.) I would like to thank the other members of the Harris Group as a whole for their support. I have enjoyed the technical discussions, the chats (or debates) in the computer room, the lunches outside McCullough, and the softball games even though I always had to play catcher.

I would like to thank Gail Chun-Creech, Susie Burns, and Lu Miao for their outstanding administrative support. They deserve all the credit for keeping this research group running smoothly and I can hardly imagine how the group could have functioned without their help.
This work has also benefited greatly from collaborations with several people outside the Harris Group. I would like to thank Edwin Kan, Zhiping Yu and Richard Williams of Professor Robert Dutton’s group for assistance in doing SEDAN, PISCES and SEMI-CAD simulations, Marek Mierzwinski and Mary Weybright of Professor James Plummer’s group for help in setting-up the high frequency measurement, Brent Boyer of Professor Fabian Pease’s group for showing me how to use image reversal resist, Chris Knorr of CIS for assistance in using the Hitachi e-beam lithography system. I would like to thank Dr. Jerry Erickson of Dawn Technologies for providing us Semi-Cad simulation tools as well as technical help. I would like to thank Drs. Thomas Low and Don D’Avanzo of Hewlett-Packard for useful discussions about dry etching and making the dry etching system available for my use. I would also like to thank Drs. Sandeep Bahl and Stretch Camnitz of Hewlett-Packard for useful discussions and making the measurement system with hot chuck available for my use.

Finally, I would like to thank my friends and my family. I owe a tremendous thank you to Peng Cheng for introducing me to the Harris group and giving me help in all aspects during my years at Stanford. Thanks to my wife, Gina Yuhong Yao, for listening and understanding. Her positive and forward-looking attitude is always an encouragement to me.
# Table of Contents

Abstract .................................................................................................................. iii  
Acknowledgments .................................................................................................. iv  
List of Tables ......................................................................................................... x  
List of Figures ........................................................................................................ xi  

Chapter 1 Introduction ............................................................................................ 1  
1.1 Introduction ...................................................................................................... 1  
1.2 Motivation and outline of the dissertation ...................................................... 2  

Chapter 2 Background ............................................................................................ 5  
2.1 Molecular Beam Epitaxy (MBE) ...................................................................... 5  
2.1.1 MBE System ............................................................................................ 5  
2.1.2 MBE Growth .......................................................................................... 9  
2.1.3 Growth of HBTs ....................................................................................... 11  
2.2 Heterojunction Bipolar Transistors (HBTs) .................................................... 14  
2.2.1 Operation of HBTs .................................................................................. 14  
2.2.2 Applications of HBTs ............................................................................. 19  
2.2.3 Present State of HBT Research ................................................................ 20  

Chapter 3 HBT Device Design ............................................................................... 23  
3.1 Current Gain ..................................................................................................... 23  
3.1.1 Overview .................................................................................................. 23  
3.1.2 Beryllium Diffusion and Emitter Injection Efficiency .............................. 25  
3.1.3 Emitter Quality and Space Charge Region Recombination ...................... 29  
3.1.4 Base Doping and Bulk Base Recombination ............................................. 32  
3.1.5 AlGaAs Passivation Ledge and Surface Recombination ......................... 35  
3.2 The Cutoff Frequency, $f_t$ ............................................................................ 40  
3.2.1 The Total Emitter-Collector Transit Time .............................................. 40  
3.2.2 Ohmic Contacts to GaAs ......................................................................... 45  
3.2.3 Emitter and Collector Resistances ......................................................... 48  
3.2.4 Base-Collector Capacitance .................................................................... 52  
3.3 The Maximum Oscillation Frequency, $f_{\text{max}}$ .......................................... 55  
3.3.1 Derivation of $f_{\text{max}}$ ............................................................................. 55  
3.3.2 Base Resistance ...................................................................................... 56  
3.4 HBT Reliability ............................................................................................... 61  
3.4.1 Overview ................................................................................................. 61
5.3.2 Emitter Size Effect ......................................................... 133
5.3.3 Gummel Plots and Ideality Factors .................................. 135
5.3.4 Base Pushout Effect ..................................................... 136
5.3.5 Temperature Dependence ............................................... 137

Chapter 6 RF HBT Performance with and without Sub-Micron Passivation Ledges .... 140

6.1 Design of Experiments .................................................... 140
  6.1.1 Process Integration .................................................... 140
  6.1.2 Layout Design ......................................................... 142

6.2 High Frequency Measurement ......................................... 145
  6.2.1 Measurement Set-up and Procedure ................................ 145
  6.2.2 Data Processing and Performance Extraction .................... 147

6.3 Experimental Results .................................................... 148
  6.3.1 f_t and f_{max} ......................................................... 148
  6.3.2 Discussion ............................................................ 152

Chapter 7 Conclusions and Future Work .................................. 153

7.1 Conclusions ............................................................... 153
7.2 Future Work .............................................................. 154

Chapter 8 References .......................................................... 156

Appendix A The Processing Flow for DC HBTs with Passivation Ledges .......... 165
Appendix B The Processing Flow for RF HBTs without and with Passivation Ledges 169
List of Tables

Table 2-1   A typical HBT epitaxial structure grown by MBE. ........................................... 12
Table 3-1   Emitter contact resistance and mesa resistance. ............................................. 49
Table 3-2   DC HBT structure for ledge passivation study. .............................................. 69
Table 3-3   RF HBT structure for device optimization study. ............................................. 70
Table 3-4   Range of scalable quantities and fixed values of the non-scalable
             quantities during the scaling in HBT width dimension. ................................. 73
Table 4-1   The overlay accuracy of the HL-700F e-beam lithography system. ............... 89
Table 4-2   The main WM and CM search and scan specifications under the condition
             set #3. ............................................................................................................. 89
Table 4-3   Volatility and forming energy of possible etch products. ................................. 95
Table 4-4   The values of the input and output variables in the DOE test performed for
             optimizing SiCl\textsubscript{4}-based GaAs and AlGaAs dry etch. ...................... 97
Table 4-5   The values of the input and output variables in the DOE test performed for
             developing SiCl\textsubscript{4}-based InGaAs dry etch. ..................................... 102
Table 4-6   The conditions developed for dry etching sputtered W, PECVD silicon
             nitride and oxide, and polyimide. ................................................................. 105
Table 4-7   A summary of RIE recipes used in the Selective Undercut Etching
             approach ........................................................................................................ 109
List of Figures

Figure 2-1 A schematic diagram of the growth chamber I of Varian GEN-II Molecular Beam Epitaxy System. ......................................................... 6

Figure 2-2 A typical RHEED intensity oscillation plot at the beginning of GaAs MBE growth. ................................................................. 9

Figure 2-3 The band diagram of an n-p-n homojunction bipolar transistor under forward bias condition. \( j_{ne}, j_{pe}, \) and \( j_{eb} \) are the current components due to the electrons injected from emitter to base, the holes injected from base to emitter and electron-hole recombination in the base regions, respectively. \( j_{c} \) is the collector current. .............................................. 15

Figure 2-4 The band diagram of an N-p-n hetero-junction bipolar transistor (HBT) under the forward bias condition. ........................................ 17

Figure 3-1 The origins of the various base current components. .................. 25

Figure 3-2 (a) The hole back-injection determined current gain, \( \beta_{c} \), as a function of the p-n junction shift for different base doping concentrations as well as composition grading. (b) The hole back-injection determined current gain, \( \beta_{c} \), as a function of the p-n junction shift for different temperatures 28

Figure 3-3 Gummel plot of a 2x10 \( \mu \text{m}^2 \) passivated HBT device and the simulated space charge recombination current with different recombination lifetimes. The solid lines are the measured \( I_{b} \) and \( I_{c} \). The dotted lines are the simulated \( I_{scr} \). ................................................................. 31

Figure 3-4 Bulk base recombination determined current gain \( \beta_{bb} \) as a function of base doping concentrations for different base width. ................. 33

Figure 3-5 SEDAN simulation results for HBTs with different base doping concentrations. (a) HBTs with graded base-emitter junctions, (b) HBTs with abrupt base-emitter junctions. ......................................................... 34

Figure 3-6 (a) A schematic cross section of the emitter and base layers of AlGaAs/GaAs HBT with an AlGaAs ledge, (b) the corresponding energy band diagram of a section through the AlGaAs ledge layer and base layer. ................................................................. 39

Figure 3-7 An idealized HBT cross section for illustrating the various resistive elements, capacitive elements, and transit regions, which comprise the total emitter-collector transit time, \( t_{ec} \). ......................................................... 44
Figure 3-8 Two kinds of current flow configurations: (a) current flows perpendicularly through an contact pad; (b) current flows from one pad to another through a thin conducting layer. ........................................ 47

Figure 3-9 A schematic drawing of the emitter mesa structure. ........................................ 48

Figure 3-10 A schematic drawing of HBT cross section for illustrating the collector resistance components. ................................................................. 51

Figure 3-11 The specific collector contact resistance as a function of RTA temperature for two different sub-collector doping levels. ......................... 52

Figure 3-12 The specific contact resistance as a function of RTA temperature for different length of time. ................................................................. 52

Figure 3-13 A schematic drawing of HBT cross section illustrating the different components of base-collector capacitance associated with the intrinsic and extrinsic base areas. ................................................................. 53

Figure 3-14 A schematic drawing of HBT cross section for illustrating the three components in total base resistance. The width of the structure along the dimension perpendicular to the cross section is W. ........................................ 57

Figure 3-15 The calculated and experimentally measured sheet resistance values for p-type GaAs base layers with different base width. .......................... 58

Figure 3-16 The measured specific contact resistances of p-type GaAs base layers doped at different concentrations and contacted by different metal films. 59

Figure 3-17 The base contact transfer lengths for different base width and base doping concentrations. ................................................................. 60

Figure 3-18 The reliability data published by Texas Instruments Central Research Laboratories. ................................................................. 64

Figure 3-19 (a) The simulated average emitter-base junction temperature rise as a function of emitter perimeter area ratio for different emitter sizes; (b) the junction temperature rise as a function of emitter width for different emitter sizes. ................................................................. 65

Figure 3-20 The $f_i$ and $f_{max}$ as a function of scaling factor defined as the scaled emitter size divided by 2 µm. The device is only scaled in width dimension and the extrinsic base area is assumed to be non-scalable. .... 73

Figure 4-1 The SEM pictures of the metal profiles evaporated without and with the collimate alignment. ................................................................. 76

Figure 4-2 The SEM picture of the photoresist overhang processed by using the chlorobenzene treatment. ................................................................. 77
Figure 4-13  The Dimension of the ledge-to-base-metal spacing as a function of the BOE undercut etching time. ................................................................. 111

Figure 4-14  The Silicon Dioxide Sidewalls approach. ........................................ 112

Figure 5-1  The schematic drawing of the base region where the analytical simulation is performed to investigate the different components of the base current. The width of the structure along the dimension perpendicular to the cross section is W. ............................................ 114

Figure 5-2  The calculated current gain of HBTs as a function of passivation ledge length with different base widths. The emitter size is 2 x 10 \( \mu \text{m}^2 \), the recombination velocity is 1 x 10^6 cm/sec. for an unpassivated surface and 1 x 10^3 cm/sec. for passivated interfaces. There is no base electric field. ..................................................................................... 119

Figure 5-3  The calculated current gain of HBTs as a function of passivation ledge length with different base grading. The emitter size is 2 x 10 \( \mu \text{m}^2 \), the recombination velocity is 1 x 10^6 cm/sec. for unpassivated surface and 1 x 10^3 cm/sec. for passivated interface. The base width is 1000\( \AA \). .......... 120

Figure 5-4  The current gain dependence on surface recombination velocity, \( s_a \), underneath the passivation ledge. .................................................. 121

Figure 5-5  The current gain dependence on the surface recombination velocity \( s_{\text{max}} \) at the unpassivated base surface. .............................................................. 122

Figure 5-6  The Semi-Cad simulation results for current gain dependence on emitter-base bias for HBTs with different passivation ledge lengths. ........... 123

Figure 5-7  Comparison between Semi-Cad simulation and analytical calculation for current gain versus passivation ledge length. ........................................ 124

Figure 5-8  Contour map of electron concentration for a passivated HBT device under normal operation. The emitter size is 2 x 10 \( \mu \text{m}^2 \) and only half of the device is shown. The base width is 1000\( \AA \) and the passivation ledge length is 0.3 \( \mu \text{m} \). The HBT device is biased at \( V_{bc} = 1.4 \) volts and \( V_{ce} = 1.4 \) volts. ........................................................................................................ 125

Figure 5-9  A zoom view of the contour map in Figure 5-8 and the electron concentration plots in the extrinsic base. Top: zoom view of the contour map in Figure 5-8. Bottom Left: electron concentration right underneath the extrinsic base surface (0 < x < 0.3 passivated, and 0.3 < x unpassivated). Bottom Right: electron concentration 100\( \AA \) beneath the extrinsic base surface. ........................................................................ 126

xiv
Figure 6-7  The highest measured values of $f_t$ and $f_{max}$ for HBTs without and with passivation ledges. (a) without passivation ledges. (b) with 0.3 $\mu$m passivation ledges. ................................................................. 149

Figure 6-8  Measured dependence of $f_t$ and $f_{max}$ on collector current for HBTs without and with passivation ledges. (a) without passivation ledges. (b) with 0.3 $\mu$m passivation ledges. ................................................................. 151
Chapter 1
Introduction

1.1 Introduction

Since Integrated Circuits (ICs) based on silicon devices were invented by Kilby and Noyce in 1959, silicon technology has made remarkable advances and dominated electronic applications for the computer and communication industries. To achieve higher integration density and faster information processing speed, silicon devices have been scaled down by 100 times from 35 μm to 0.35 μm over the past thirty-five years. This trend will probably continue for another ten years to push the minimum feature size down to 0.1 μm, which is claimed to be the fundamental limit of silicon devices. [1-2] Parallel to the development of silicon technology, ICs based on III-V compound semiconductor devices have been developed for applications which require either higher operation speed or manufacturing compatibility with opto-electronic devices. Over the past five years, because of the maturity of epitaxial technologies, such as molecular beam epitaxy (MBE) and metal organic chemical vapor deposition (MOCVD), and because of the commercial demand created by wireless communication and high speed computing applications, compound semiconductor based ICs have become more and more competitive.

GaAs (which is the most widely used III-V material) ICs have several advantages when compared to silicon ICs. First, because of their lighter effective mass, the electrons in GaAs have significantly higher mobility and saturation velocity, which allows GaAs based devices to operate at higher speeds. Second, the availability of semi-insulating substrates minimizes parasitic capacitances and makes GaAs ICs suitable for high frequency applications. Finally, the GaAs material system has a direct bandgap, which is required for opto-electronic devices, thus GaAs ICs and opto-electronic devices can be monolithically integrated on the same chip to form opto-electronic integrated circuits (OEICs).

Metal Semiconductor Field Effect Transistors (MESFETs), Modulation Doped Field Effect Transistors (MODFETs), and Hetero-junction Bipolar Transistors (HBTs) are the three types of transistors upon which GaAs ICs are based. All of these devices have demonstrated impressive high frequency and high speed digital performance. Among them, HBTs have attracted attention as potentially the highest speed three
terminal device suitable for digital and high linearity analog applications. When compared to GaAs MESFETs or MODFETs, AlGaAs/GaAs HBTs have better threshold voltage control, higher integration density, lower 1/f noise, and no anomalies. When compared to silicon bipolar devices, AlGaAs/GaAs HBTs have a much higher cut-off frequency, larger power efficiency, better linearity, and less substrate parasitic capacitance. Moreover, silicon bipolar, MESFETs and MODFETs require very small critical dimension lithography, while HBTs do not have this requirement for most applications.

With progress in the AlGaAs/GaAs materials system, characterization and HBT fabrication process development, AlGaAs/GaAs HBTs have moved into the IC manufacturing stage. Having examined performance feasibility of HBTs, the compound semiconductor IC industry is now focused on device reliability and performance optimization issues. This dissertation investigates how to optimize the design and fabrication of high performance AlGaAs/GaAs HBTs with sub-micron features.

1.2 Motivation and Outline of the Dissertation

Bipolar transistors are minority carrier devices and suffer from high surface recombination rates of minority carriers. Fortunately, silicon bipolar transistors could solve the high recombination problem by passivating the surface with its native oxide, SiO₂. GaAs has a very high surface recombination velocity; however, unlike silicon, GaAs does not have a native oxide with a low surface state which passivates the surface. When the device geometry is scaled down to realize high frequency performance, the increase in the ratio of emitter perimeter to area in smaller devices results in an even larger peripheral recombination current, which has limited the optimization and scalability of AlGaAs/GaAs HBTs.

Although there are several different approaches to solving this surface passivation problem, such as applying a chemical passivation material to the GaAs surface, the real manufacturable solution came from the work reported in 1985 by H. H. Lin, who developed a ledge structure and used the passivating properties of AlGaAs to improve the performance of AlGaAs/GaAs HBTs. Since then, fully depleted AlGaAs passivation ledges have been widely used to reduce surface recombination around the emitter perimeter of AlGaAs/GaAs HBTs. Over the past decade, research on HBTs has shown
that this passivation ledge not only increases the current gain, but also reduces the low frequency noise and improves the reliability of HBTs. [1-4]

Unfortunately, this passivation ledge increases the base-collector junction capacitance and the base resistance, so it decreases the cut-off frequency, $f_t$, and the maximum oscillation frequency, $f_{max}$, of HBTs. To optimize the device performance and reliability, these passivation ledges have to be minimized. In this dissertation, I theoretically and experimentally studied HBT devices with different passivation ledge lengths ranging from 0.1 $\mu$m to 2.0 $\mu$m and found the minimum sufficient passivation ledge length required to fully passivate the AlGaAs/GaAs HBTs. To apply this ultra-short passivation ledge structure to high frequency HBT devices, I developed three new ledge processing approaches which can be used under different conditions and for different applications.

Sub-micron HBT design, sub-micron HBT fabrication techniques, and the scalability of AlGaAs/GaAs HBTs are the other topics studied and presented in this dissertation. Wireless communication and high speed computing applications require that GaAs ICs have lower power consumption and higher integration density. Scaling HBTs down to sub-micron sizes allows the transistors to achieve the desired high frequency performance at a relatively low current level, which reduces the total power consumption of integrated circuits. The smaller device size increases the integration density, which not only makes AlGaAs/GaAs HBT-ICs suitable for more applications but also reduces the manufacturing cost of such units. In addition, sub-micron HBTs will be more reliable because the self-heating problem in sub-micron HBTs is less serious than in larger HBTs.

An outline of this dissertation is as follows: Chapter 2 gives a brief summary of the molecular beam epitaxy system and the MBE growth techniques for HBT structures. It also describes the operation of HBTs, the application of HBTs and the present state of HBT research. Chapter 3 discusses HBT design principles and tradeoffs. The focus of my design is on current gain, cutoff frequency, maximum oscillation frequency, and optimization of device performance to meet the requirements of these applications. The reliability issues and the scalability of HBTs are also discussed in this chapter. Chapter 4 describes conventional HBT processing techniques and my developmental work on fabricating HBTs with sub-micron features. Electron beam lithography and dry etching III-V compound semiconductors, which are two necessary techniques to fabricate sub-micron HBTs, have been extensively studied and tested. The optimized processing
conditions and recipes are listed together with step-by-step instructions. Four approaches to form sub-micron AlGaAs passivation ledges are described and compared in terms of advantages and disadvantages. Chapter 5 presents the performance of Npn AlGaAs/GaAs HBTs used to investigate the minimum passivation ledge lengths. It includes both simulation and experimental results of HBTs with different passivation ledge lengths. The passivation ledge design rules established by these experimental observations and the temperature dependence of HBT performance are also discussed. Chapter 6 covers the measured high frequency performance of Npn AlGaAs/GaAs HBT devices fabricated using the design of chapter 3 and the processing approaches of chapter 4. In addition, the high frequency measurement setup and device parameter extraction techniques are described. Chapter 7 gives conclusions of this dissertation research and suggestions for future work. All the references referred to in this dissertation are listed in chapter 8. The complete DC and RF HBT processing flows are included in the appendices.
Chapter 2
Background

This chapter provides the background necessary for understanding the growth of HBT structures and the mechanism of HBT operation. Section 2.1 introduces the MBE system, MBE growth techniques, and typical MBE growth conditions for HBT structures. MBE allows extremely precise control of layer thickness and composition, which provides a tool to modify the band-structure within a single atomic layer of the material. Although HBTs do not require atomic layer control, compared to MOCVD, an alternative growth method, MBE provides better control in material composition and layer thickness. Section 2.2 introduces the operation of hetero-junction bipolar transistors (HBTs), the applications of HBTs, and the present state of HBT research.

2.1 Molecular Beam Epitaxy (MBE)

Molecular Beam Epitaxy (MBE) is a versatile growth technique by which a specific number of atomic layers can be precisely grown and by which precisely controlled composition profiles and doping levels can be produced. Moreover, because growth occurs in an ultrahigh vacuum environment, the growth can be monitored in-situ by surface sensitive diagnostic methods such as Reflection High Energy Electron Diffraction (RHEED), Auger Electron Spectroscopy (AES), and Mass Spectrometry. These powerful facilities for control and analysis eliminate much of the guesswork in MBE and allow the growth of sophisticated device epitaxial structures, which are required to fabricate HBTs with optimal performance. All the device epitaxial layers used for this dissertation were grown by Molecular Beam Epitaxy.

2.1.1 MBE System

The MBE System used for this dissertation research was a Varian GEN-II molecular beam epitaxy system. This MBE system consists of four chambers: load chamber, transition tube, growth chamber I, and growth chamber II. The function of the load chamber is to introduce cleaned wafers into the system and bake out the wafer and the wafer holder to remove water vapor. The load chamber, which is pumped by both a cryopump and an ion pump, achieves a pressure of ~ $10^{-6}$ torr range because of the periodic exposure to air. The function of the transition tube is to store wafers before transferring them into growth chamber I or growth chamber II, or removing them back.
into the load chamber after growth. Because the transition tube is at a pressure of \( \sim 10^9 \) torr, it is also utilized to isolate the UHV growth chambers from the high pressure load chamber during sample introduction. Growth chamber I and growth chamber II contain the effusion cells, substrate heaters, and \textit{in-situ} analysis tools. They are the location where all epitaxial growth takes places. The background pressure of the growth chambers is at low \( 10^{-10} \) torr range.

Figure 2-1 shows a schematic diagram of the Varian GEN-II MBE machine's growth chamber I. Growth chamber II is identical to growth chamber I except that novel silicon filament and carbon filament sources replace the traditional silicon and beryllium furnaces which serve as the n and p-type doping sources.

Figure 2-1  A schematic diagram of the growth chamber I of Varian GEN-II Molecular Beam Epitaxy System.

In growth chamber I, the six commonly used source materials, As, Ga, Al, In, Si (an n-type dopant), Be (a p-type dopant), are contained in pyrolytic boron nitride (PBN) crucibles which are heated by furnaces surrounding them. The temperatures of the
crucibles are monitored with thermocouples around the center or at the bottom of the crucibles. The regions surrounding the furnaces are constantly cooled by a chilled water/isopropyl-alcohol solution which flows through a shroud to provide a constant thermal environment which stabilizes the crucible temperatures and provides thermal isolation between the different furnaces. Each furnace is equipped with a mechanically controlled shutter in front of the crucible to turn on or off the beam flux rapidly. In growth chamber I, there is a total of eight furnaces, two for As, two for Ga, and the other four for Al, In, Si, and Be, respectively. On one of the As furnaces, a valved cracker is equipped to crack $\text{As}_4$ into $2\text{As}_2$ molecules which matches the natural vapor As species over GaAs.

In growth chamber II, there are no Si and Be furnaces. A resistively-heated silicon filament source supplies n-type dopants, and a resistively-heated graphite filament source supplies p-type dopants. Both of these filament sources are controlled by input current levels instead of furnace temperatures. The use of a Si filament source allows the growth of dopant concentration graded structures as well as highly doped n-type GaAs layers, which are very useful in HBT epitaxial growth. The use of a graphite filament source provides carbon doped HBTs which eliminate the Be diffusion problem in more conventional HBTs. One problem which has occurred is that heavily doped p-type GaAs from a graphite-filament source exhibits very short minority carrier lifetimes. Since the implementation of these filament sources into growth chamber II is fairly recent, all the HBT devices used in this dissertation research were grown in growth chamber I.

The substrate heater assembly has a substrate heater and a beam flux monitor on opposite sides of the heater. A 2-inch or 3-inch GaAs substrate is bound by a wire to a molybdenum holder and placed on the substrate heater, which radiatively heats the substrate to approximate by 600°C. The growth temperature is chosen to provide sufficient surface mobility for the evaporated elements to achieve good crystalline quality with respect to the deposited materials. As shown in Figure 2-1, the flux of each source material arrives at a different angle for different parts of the substrate. Therefore, the substrate is rotated at a speed of approximately 8 cycles per minute during the growth to reduce this flux non-uniformity across the wafer. The beam flux monitor is an ion gauge and is mounted on the backside of the heater. During the beam flux calibration and before each growth, the beam flux monitor is rotated to face the furnaces in order to measure the beam-flux-equivalent-pressure of each source (except the dopant sources
which have fluxes too low to be measured). The growth rates of the desired epitaxial materials can be calibrated using this flux measurement because when the substrate is facing the furnaces during the growth, the substrate is at exactly the same location where the beam flux was measured.

The cryo-shroud surrounding the interior of the growth chamber is constantly cooled with a flow of liquid nitrogen. Any source material molecules that miss the substrate hit the cryo-shroud and stick to it. Therefore, no re-evaporated molecules will get incorporated into the epitaxial layer. This cryo-shroud also traps foreign molecules in the chamber to make the epitaxial growth less susceptible to background contaminants. Whenever the growth chamber has been opened for maintenance, such as replacing the source materials or fixing the substrate heater assembly, the furnaces and the stainless steel chamber are extensively baked and outgassed for about a week at a temperature of ~200°C. After baking, a mass spectrometer is used to ensure that traces of foreign species, such as H₂O and O₂, are within tolerable limits. The vacuum in the growth chamber is maintained by a combination of a cryo-pump, ion pump, titanium sublimation pump, and liquid nitrogen cooled cryo-shroud.

The most frequently used analytical equipment includes the mass spectrometer and the Reflection High Energy Electron Diffraction (RHEED). [2-1] The mass spectrometer is a quadruple-type mass spectrum analyzer for probing the amount and species of all residual gases in the chamber. It is frequently used for leak checking and residual gas analysis. The RHEED system consists of an electron gun and a phosphor screen on opposite sides of the chamber. The electron beam strikes the substrate at grazing incidence and is diffracted onto the phosphor screen, providing information on the sample surface reconstruction. The diffracted beams undergo and oscillation which is used for calibrating growth rates and alloy compositions.

Figure 2-2 shows a typical RHEED intensity oscillation plot. The working mechanism of RHEED can be explained as follow. The central spot in the RHEED pattern is imaged onto a photo detector placed close to the phosphor screen. When the wafer surface is smooth, the maximum reflection signal is detected, and when half of a monolayer of Ga is deposited, the RHEED intensity decreases to its minimum because the electrons diffracted from different parts of the wafer surface having one monolayer of difference in height interfere destructively with one another. As a full monolayer is nearly completed, the intensity recovers to its original peak value. Therefore, the period
recorded on the RHEED plot exactly corresponds to the growth rate of a single monolayer of GaAs, which can be converted into a general growth rate in units of μm/hr. using the thickness of one GaAs monolayer. Since the alloy composition is proportional to the ratio of the individual group-III growth rates over the total growth rate, the alloy composition can be derived from two or more RHEED oscillation plots. One notable feature of Figure 2-2 is that the maximum intensity decreases after each successive layer of GaAs has been grown. This is due to two factors: 1) non-uniform growth rate because the substrate is not rotating during RHEED measurement and 2) a slight increase in surface roughness as more and more epitaxial layers are grown. [2-2\] If the Ga shutter is closed for a short period of time, the reflection intensity increases back to its initial value. It is thus conceivable to grow GaAs layers by alternately opening and closing the Ga shutter so that near perfect smoothness is achieved during the entire growth. But, such a process is usually not necessary for growing HBT wafers and was not used in this dissertation research.

![Intensity graph]

Figure 2-2  A typical RHEED intensity oscillation plot at the beginning of GaAs MBE growth.

2.1.2 MBE Growth

Performing a MBE growth requires three steps: substrate preparation, source calibration, and epitaxial growth.
To introduce substrates into the growth chamber, the starting GaAs wafers were attached by wire to a molybdenum holder and loaded into the load chamber. Then the load chamber was pumped down to approximately $1 \times 10^{-6}$ Torr before baking each wafer at 400°C for one hour to remove water vapor. After the bake was finished and the load chamber was cooled down, the chamber usually achieved $4 \times 10^{-7}$ Torr, which allowed us to open the valve between the load chamber and the transition tube and to transfer all the baked substrates into the transition tube. All the prepared substrates sat on the transition tube trolley for two hours or longer before they were transferred into the growth chamber for epitaxial growth. This period in the transition tube enables the pressure to go down to its stand-by value after being opened to the load chamber. Since the starting material delivered by vendors is very clean and ready for direct loading, no cleaning or etching processes were used before loading substrates into the system.

The group-III source calibration of growth rates is done after each source loading using the RHEED technique which was previously discussed. Dopant source calibration was performed by growing a few test wafers with different dopant source temperatures and doing Hall measurements. Because the doping concentrations were determined by the Hall effect which stems from transport properties of the carriers, all the doping concentrations cited in this dissertation are the electrically activate carrier densities. When comparing experimental results to device simulations, the impurity activation factors for n and p-type dopants were always set to one to match the experimental carrier concentrations.

Because the amount of source material in the crucibles decreases after each growth, the actual growth rates decrease slightly during a sequence of runs. Theoretically, the RHEED calibration could be done before each growth, but practically the beam flux measurement was employed to monitor the growth rates and adjust the source temperatures so that the desired growth rates and alloy compositions were achieved. The validity of this calibration approach comes from the fact that the relationship between the growth rate and the beam flux still holds while the correspondence between the beam flux and the source temperature varies. Because the group III beam flux can only be measured while the As furnace stays idle, the beam flux measurement data was obtained a few hours before the first epitaxial growth and this value was used for two days during which the As source stayed at an elevated temperature and a number of growth runs were completed.
Once the calibration had been done and the As source (As$_2$ most of the time) had reached its target vapor pressure, a substrate was transferred from the transition tube into the growth chamber. Then the substrate was heated up to 700°C for 10 minutes, during which time the thin native oxide on the wafer surface decomposed, exposing a fresh GaAs surface ready for epitaxial growth. The substrate temperature was then usually set at 600°C for the growth of GaAs and AlGaAs and at 480°C for the growth of InGaAs.

The MBE epitaxial growth process involves the reaction of one or more thermal beams of atoms or molecules with a crystalline surface under ultra-high vacuum conditions. The growth of GaAs is commonly done in an arsenic over-pressure environment, under which the surface is always covered by a single atomic layer of arsenic and the sticking coefficient for any additional impinging arsenic on this surface is zero. When the shutter(s) of group III source(s) are opened, impinging group III atoms stick on this arsenic surface and form a fresh location which will be covered immediately by new arsenic. Because the adsorbing of As atoms on a fresh group-III atom surface takes negligible time, the group-III flux(es) determine the growth rate and alloy composition. When both Ga and Al shutters are open to grow Al$_x$Ga$_{1-x}$As, the total growth rate $G_{\text{Total}} = G_{\text{Al}} + G_{\text{Ga}}$, and the aluminum composition $x = G_{\text{Al}} / G_{\text{Total}}$.

All furnace temperatures and shutter positions were controlled by computer, which allowed the MBE growth to be programmable, providing reproducible growth and precise layer compositions and thicknesses. For the epitaxial layers grown in our MBE system, the thickness and composition can be controlled with an error of less than 1%, and the non-uniformity across a 2-inch wafer is less than 2%. The doping concentrations are reproducible, although the absolute values may have an error of 20% - 50% since the doping concentration calibrations have not been cross-checked by other measurement techniques.

2.1.3 Growth of HBTs

Table 2-1 shows a typical HBT epitaxial structure grown by MBE. For the growth of Npn AlGaAs/GaAs HBTs, the GaAs substrates were either silicon doped, n$^+$, or chromium doped, which is semi-insulating(SI). High frequency HBT wafers were always grown on SI substrates, so that the parasitic capacitances of the HBTs were small. DC HBT wafers were usually grown on n$^+$ substrates because processing with a backside ohmic collector contact was much simpler.
Table 2-1  A typical HBT epitaxial structure grown by MBE.

Typically, the growth rate of one Ga source (Ga1) was set to 0.5 μm/hr., and the growth rate of the other Ga source (Ga2) was set to 0.214 μm/hr. The growth rate of the Al source was set to 0.214 μm/hr. as was the growth rate of the In source. The sub-collector and collector layers were grown with both Ga1 and Ga2 sources open. With a total growth rate of 0.714 μm/hr., the growth of these two thicker layers did not take a long time. The base layer was usually grown with only Ga2 open so that the growth rate was 0.214 μm/hr., with which a high base doping concentration was achieved without raising the Be doping source temperature to too high a value.

The emitter region consisted of three layers: emitter-base grading, bulk emitter, and emitter-cap grading. The bulk emitter layer was grown with both Ga1 and Al shutters open, so the total growth rate was 0.714 μm/hr. and the Al composition was x = 0.3 under the normal conditions described at the beginning of this section. The emitter-base and emitter-cap grading layers were grown by the “chirp grading” technique. In this technique, the Ga1 shutter remained open at all times while the Al shutter and Ga2 shutter were opened and closed alternatively. The grading was accomplished by changing the percentage of Al open time over the total time of each sequential cycle. Whenever the Al shutter was closed, the Ga2 shutter was open. Thus, the total growth rate was a constant. This constant growth rate made the doping profiles in the grading layers uniform; otherwise, the discontinuity in doping profiles might affect the transport
of electrons in the graded region. The electron mobility might also be reduced by the conduction band discontinuity, so a grading layer was usually divided into 20 periods to make all AlGaAs barriers thin.

The GaAs cap layer was grown with both Ga1 and Ga2 open. The InGaAs cap-contact grading layer was also grown by the “chirp grading” technique similar to that used for grading the emitter-base. When both Ga2 and In shutters were open, In0.5Ga0.5As was grown with a growth rate of 0.428 μm/hr. When only Ga1 was open, the GaAs growth rate was 0.5 μm/hr., which is close to In0.5Ga0.5As growth rate so that a fairly uniform doping profile could be achieved in the cap-contact layer. Finally, the epitaxial growth was completed with an InAs contact layer. The lattice mismatch between InGaAs and GaAs would generate misfit dislocations in the InGaAs grading layer and InAs contact layer. However, these dislocations only propagate to the wafer surface, thus the active device layers are not affected. The use of this InAs cap dramatically reduces the emitter contact resistance, which is discussed extensively in chapter 3.
2.2 Heterojunction Bipolar Transistors (HBTs)

The bipolar transistor, so named to emphasize the importance of both electrons and holes in the base to its operation, was invented by William Shockley in 1949.\cite{2-4} The HBT is a bipolar transistor with a wider bandgap material for the emitter and a narrower bandgap material for the base. The collector is usually made of the same material as either the base or the emitter. The latter is called Double Heterojunction Bipolar Transistors (DHBTs). Depending the material system used, the emitter-base heterojunction can be formed by AlGaAs/GaAs, InGaP/GaAs, InAlAs/InGaAs, InP/InGaAs, or Si/SiGe. The Npn AlGaAs/GaAs HBT has been most extensively studied and is now available in IC products. However, there are still some remaining issues, which are the main subjects of this dissertation.

2.2.1 Operation of HBTs

An n-p-n bipolar junction transistor consists of two n-type semiconductor layers separated by a thin, p-type layer, which forms two pn junctions, connected back-to-back. Under normal operation, the base-emitter pn junction is forward biased and the base-collector pn junction is reverse biased. Figure 2-3 shows the energy band diagram of an n-p-n homojunction bipolar transistor under this active bias condition.

Because electrons always flow from a region with a high Fermi level to one of a lower Fermi level, forward biasing the base-emitter junction causes electrons in the emitter to flow or be injected into the base. Since the base layer is designed to be sufficiently thin, most of these electrons reach the base-collector junction edge before recombining with the holes inside the base. Once reaching the reverse biased base-collector junction, electrons are swept into the collector by the large electric field within the junction and constitute the collector current of the transistor.

Useful figures of merit for bipolar transistors include the emitter injection efficiency, $\gamma$, base transport factor, $\alpha$, and common emitter current gain, $\beta$, which are defined as:

$$\gamma = \frac{\text{electron current from the emitter}}{\text{total emitter current}} = \frac{j_{ne}}{j_{ne} + j_{pe}},$$

(2.1)
\[ \alpha = \frac{\text{electron current reaching the collector}}{\text{electron current from the emitter}} = \frac{j_{ne} - j_{rb}}{j_{ne}}, \]  
\[ \beta = \frac{j_c}{j_b} = \frac{j_e}{j_e - j_c} = \frac{j_{ne}}{j_{ne} + j_{pe}} \frac{j_{ne} - j_{rb}}{j_{ne}} = \frac{\gamma \alpha}{1 - \gamma \alpha}, \]  

where \( j_e \) is the total emitter current and all other current components are defined in the diagram above. The goal is to have \( j_{ne} \gg j_{pe} \) and \( j_c \approx j_{ne} \), thus achieving a maximum common emitter current gain.

Figure 2-3 The band diagram of an n-p-n homojunction bipolar transistor under forward bias condition. \( j_{ne}, j_{pe}, \) and \( j_{rb} \) are the current components due to the electrons injected from emitter to base, the holes injected from base to emitter and electron-hole recombination in the base regions, respectively. \( j_c \) is the collector current.

In a homojunction bipolar transistor (BJT), the emitter and base regions are composed of the same energy bandgap material. Therefore, when the conduction-band potential barrier is reduced by the forward bias to inject electrons from the emitter into the base, the valence-band potential barrier for holes is simultaneously reduced. This reduced barrier allows holes in the base to “back-inject” into the emitter and reduces the emitter injection efficiency. Since the potential barriers for electrons and holes are the
same in a homojunction device, the ratio of injected electron current to injected hole current \((j_{ne}/j_{pe})\) is controlled by the relative doping densities in the emitter and base. To achieve an appreciable current gain \(\beta\) in a BJT, the emitter must be far more heavily doped than the base. This inherent restriction on the base doping concentration limits the high frequency performance of the transistor and has motivated development of heterojunction bipolar transistors (HBTs) since their inception by Shockley in 1953. \cite{2-5}

Figure 2-4 shows the energy band diagram of an Npn HBT with a graded heterojunction. (The impact of heterojunction grading is discussed in several references.\cite{2-6}) In contrast to the BJT, in a HBT, the emitter is made of a larger bandgap material than the base. Consequently, with proper grading, the potential barriers for electron and hole injection, \(\Delta V_n\) and \(\Delta V_p\), differ by the bandgap potential difference between the emitter and base, \(\Delta E_g/q\). Under this condition, the injected electron current density from the emitter to base can be expressed as \cite{2-7}

\[
i_{ne} = qn_b(x_p) \frac{D_{nb}}{W_b} = q \frac{N_c \text{base} \cdot N_v \text{base}}{N_A} \exp \left( -\frac{q \Delta V_n}{kT} \right) \frac{D_{nb}}{W_b}, \tag{2.4}
\]

and the injected hole current density from base to emitter can be expressed as

\[
i_{pe} = qp_e (-x_0) \frac{D_{pe}}{L_{pe}} = q \frac{N_c \text{emitter} \cdot N_v \text{emitter}}{N_D} \exp \left( -\frac{q \Delta V_p}{kT} \right) \frac{D_{pe}}{L_{pe}}, \tag{2.5}
\]

where \(q\) is the electronic charge, \(k\) is the Boltzmann constant, \(T\) is the temperature, \(n_b(x_p)\) is the electron concentration at the base edge of the base-emitter junction, \(p_e(-x_0)\) is the hole concentration at the emitter edge of the base-emitter junction, \(D_{nb}\) is the electron diffusivity in the base, \(D_{pe}\) is the hole diffusivity in the emitter, \(W_b\) is the base width, \(L_c\) is the hole diffusion length in the emitter, \(N_D\) is the emitter doping, \(N_A\) is the base doping, \(N_{\text{base}}\) and \(N_{\text{emitter}}\) are the electron effective densities of states in the base and emitter, respectively, and \(N_{\text{vbase}}\) and \(N_{\text{ vemitter}}\) are the hole effective densities of states in the base and emitter, respectively. In these equations, Boltzmann statistics are used and the current flow is assumed to be by diffusion only. Dividing equation (2.4) by (2.5), the ratio of electron injection to hole injection currents can be written as

\[
\frac{j_{ne}}{j_{pe}} = \frac{D_{nb} L_{pe} N_D N_c \text{base} \cdot N_v \text{base}}{D_{pe} W_b N_A N_c \text{emitter} \cdot N_v \text{emitter}} \exp \left( \frac{\Delta E_g}{kT} \right). \tag{2.6}
\]
where $\Delta E_g = q(\Delta V_p - \Delta V_n)$. As a result of the electron and hole barrier differences in HBTs, the hole injection current is reduced by a factor of $\exp(-\Delta E_g/kT)$ compared to the electron injection current. For example, in an Al$_{0.3}$Ga$_{0.7}$As/GaAs graded junction, $\Delta E_g$ is 0.37 eV, which decreases the hole injection current by a factor of $1.9 \times 10^6$ at room temperature. Therefore, the base region can be doped much higher than the emitter region without seriously degrading the current gain.

---

**Figure 2-4** The band diagram of an N-p-n hetero-junction bipolar transistor (HBT) under the forward bias condition.

A high doping in the base lowers the base sheet and contact resistances, and thus improves the maximum oscillation frequency of the transistor. Simultaneously, low emitter doping reduces the base-emitter junction capacitance because the junction depletion width is wider with lower emitter doping. The lower base-emitter junction capacitance is desirable because it allows the emitter small signal resistance, $r_e$, to be larger without increasing the emitter delay time ($\tau_e = C_w \cdot r_e$). Because $r_e$ is inversely proportional to the collector current, reducing the base emitter junction capacitance enables HBTs to operate at a lower collector current and still achieve the necessary high frequency performance. To meet the requirement of low power applications, the emitter doping level is sometimes designed to be a few orders of magnitude lower than the base doping. [2-8]
In terms of applications, the HBT is targeted for frequencies higher than 10 GHz. The major high frequency figures-of-merit are the unity current gain cutoff frequency, $f_t$, and maximum oscillation frequency, $f_{\text{max}}$, for microwave analog applications, and a gate propagation delay time, $\tau_d$, for digital applications.

$f_t$ is defined as the frequency at which the incremental current gain drops to unity. The cutoff frequency is closely related to the carrier transit time through the vertical device structure from the emitter to the collector. This delay results from the base-emitter and base-collector junction charging times plus the finite time required for electrons to traverse the neutral base and depleted collector regions. The total delay can be written as

$$\tau = \frac{1}{2 \eta} = \tau_e + \tau_b + \tau_{\text{sc}} + \tau_c,$$  \hspace{1cm} (2.7)$$

where $\tau_e$ is the base-emitter capacitance charging time, $\tau_b$ is the base transit time, $\tau_{\text{sc}}$ is the collector region transit time, and $\tau_c$ is the base-collector capacitance charging time. To minimize the total transit time, all four components of $\tau$ must be reduced. The detailed equations describing each of these four components and how to minimize them are discussed extensively in chapter 3.

$f_{\text{max}}$ is defined as the frequency at which the unilateral power gain drops to unity (or zero if the conventional unit in dB is used). $f_{\text{max}}$ can be expressed as

$$f_{\text{max}} = \sqrt{\frac{f_t}{\delta n R_b C_{\text{bc}}}},$$  \hspace{1cm} (2.8)$$

where $f_t$ was defined in equation (2.7), $R_b$ is the total base resistance, and $C_{\text{bc}}$ is the base-collector junction capacitance, respectively. To design for optimal power gain, it is not necessarily desirable to achieve the highest current gain cutoff frequency, $f_t$. $f_{\text{max}}$ is very sensitive to the parasitic components of HBTs. The design tradeoffs are discussed later in chapter 3.

The gate propagation delay time of the HBT has attracted considerable attention recently because the accurate threshold voltage control of HBTs makes HBT based integrated circuits promising in high speed digital applications. The approximate gate delay in ECL circuits can be expressed as
\[ \tau_c = \sqrt{\frac{K(R_g + \frac{1}{g_m})(C_{ac} + \frac{1}{g_m})}{f_t}} \]  

where \( R_L \) and \( C_L \) are the load resistance and capacitance, respectively and \( K \) is a constant. The propagation delay differs from \( f_t \) and \( f_{\text{max}} \) because it depends on not only on device parameters, but also external load conditions.

### 2.2.2 Applications of HBTs

HBT technology has made rapid strides in maturity and has become more and more competitive in addressing a diversity of applications. [2-10] In terms of the market, HBT applications include communications systems, [2-11] high speed computing, [2-12] avionics and weapon systems, [2-13] industrial systems and high speed test equipment. [2-14] Among these applications, the communication market, which covers wireless communications (cellular phones, ground stations, and satellites), fiber optical communications, and local area networks (LANs), has exploded and is expected to keep growing well into the next decade. The “information highway” plan funded by the US government and the merging of computer and communications are generating great demand to drive this growth to greater bandwidth and higher speed.

Applications can also be categorized into four different types, based on the signal type and power handled by HBT-ICs: high power microwave applications, low power microwave applications, analog applications, and ultrahigh speed digital applications.

HBTs have unique advantages in high power microwave applications, such as power amplifiers, synthesizers, and voltage-controlled oscillators (VCOs) for radar and communications systems. [2-15] The main advantages include high RF power density, high power-added efficiency, large breakdown voltages, and low 1/f noise for synthesizers and oscillators. In addition to these performance advantages, an HBT with 2.0 \( \mu \)m minimum feature size, provides excellent performance combined with good yield, uniformity, and relatively low cost wafer processing. This combination makes power HBTs very attractive for low-cost high-volume manufacturing and applications.

HBTs have unique advantages in low power microwave circuits because of their high output resistance and low leakage conductance, provided devices are scaled properly so parasitic capacitance is minimized (which is the main subject of this dissertation).
HBT amplifiers have demonstrated record linearity at low dc power dissipation and as a result, they are of major interest in many receiver applications. Moreover, in the context of low power RF circuits, reproducible control of the knee voltage as well as reducing it are absolutely critical. The knee voltage in HBTs is not sensitive to processing because it depends largely upon the bandgap of the base region. Thus AlGaAs/GaAs HBTs have a larger knee voltage than Si BJTs; however, by using InGaAs as base material, the turn-on voltage can be reduced below 0.5 V (which is lower than a Si BJT).

HBTs have unique advantages in analog applications, such as broadband amplifiers, because of their extremely high transconductance, high output resistance, high breakdown voltage, excellent turn-on voltage uniformity, and low 1/f noise. At a given breakdown voltage, the current gain cutoff frequency of AlGaAs/GaAs HBTs is usually two to three times higher than that of state-of-the-art Si bipolar transistors. As a result, the bandwidth of broadband amplifiers made with HBTs is two to three times higher than that of Si bipolar transistors.

Finally, HBTs have unique advantages in ultrahigh speed digital applications and high speed analog/digital and digital/analog circuits, which require that both ultrahigh $f_t$ and ultrahigh $f_{max}$ be achieved, and that the turn-on voltages are extremely uniform across the chip. All multiplexers, demultiplexers, and decision circuits for fiber optical communications fall into this category, and HBTs can best serve these circuit applications.

2.2.3 Present State of HBT Research

Since HBT technology, especially in the AlGaAs/GaAs material system, has become very mature, the focus of HBT research has switched to manufacturability, cost, and reliability of the technology, as well as to circuit design and circuit performance. The research activities of both university groups and industrial laboratories are now mainly focused on achieving the highest performance of HBTs, new materials systems suitable for optimal HBT performance and manufacturing compatibility, the miniaturization of HBTs, and HBT device modeling.

Carbon-doped base InP/InGaAs HBTs have yielded a $f_t$ of 200 GHz, which is the highest value reported for any bipolar transistor at the time this dissertation was written.
The reported highest maximum oscillation frequencies \( f_{\text{max},u} \) (extracted from unilateral power gain measurement), and \( f_{\text{max},G} \) (extracted from maximum stable gain and maximum available gain measurement), are 236 GHz and 160 GHz, respectively. These high oscillation frequencies are achieved in InAlAs/InGaAs HBTs.\(^{[2-20]}\) While the circuit performance requirements can mostly be met using HBTs with \( f_t \) and \( f_{\text{max}} \) less than 100 GHz, pushing these high frequency figures of merit to higher frequencies is still attracting research interest.

In addition to the work on InP/InGaAs and InAlAs/InGaAs HBTs, InGaP/GaAs HBTs have attracted considerable interest in the recent two to three years.\(^{[2-21]}\) There are two advantages to using InGaP as the emitter. First, material quality of InGaP can be much better than AlGaAs, which is very susceptible to the incorporation of oxygen. As a result, the base-emitter space-charge region recombination in InGaP/GaAs HBTs is low enough to produce a base current ideality factor close to 1 and nearly-ideal I-V characteristics with constant current gain are obtained for five decades of collector current. Second, GaInP/GaAs HBTs are expected to result in lower 1/f noise. Replacing AlGaAs as the emitter, GaInP has far fewer traps and recombination centers associated with oxygen complexes. DX centers associated with n-type AlGaAs degrade the 1/f noise performance in these HBTs, but are not present in GaInP.\(^{[2-22]}\) The current research effort in this field is focused on material growth issues, such as grading the base-emitter junction and device processing issues, such as dry etching the InGaP emitter.

Another material system, which does not belong to the III-V compound semiconductor family, but has been receiving a great amount of interest for HBT applications, is Si/SiGe. The key advantage of Si/SiGe HBTs is that the manufacturing processes could be compatible with well-established Si processing techniques, although putting Ge into the epitaxial structures creates some limitations, such as, native oxide growth, passivation and limits on high temperature processes, all of which reduce this compatibility. The best reported Si/SiGe HBTs yield a \( f_t \) of 113 GHz and a \( f_{\text{max}} \) of 90 GHz, respectively.\(^{[2-23]}\) Furthermore, IBM has demonstrated the manufacturability of Si/SiGe HBTs using their standard CMOS processing lines with only an extra 10% in cost.\(^{[2-24]}\) However, to bring Si/SiGe HBTs up to a real manufacturing stage like many companies did with AlGaAs/GaAs HBTs, a lot of reliability studies have to be done. By the time this dissertation is written, no reliability work on Si/SiGe HBTs has been reported yet.
Most of the work on HBTs has concentrated on Npn transistors. This is largely because of their very high speed. However, Pnp HBTs have the potential for speed approaching that of the Npn HBT. Pnp HBTs, used in conjunction with Npn HBTs, would be very useful for many applications, such as complementary digital logic circuits, efficient microwave push-pull amplifiers and active loads in analog circuits. The development of complementary HBTs (Pnp/Npn pairs) has been part of the concentration of HBT research at Stanford. The Pnp AlGaAs/GaAs HBTs made at Stanford have yielded values of \( f_t \) of 37 GHz and \( f_{\text{max}} \) of 30 GHz. The 37 GHz \( f_t \) is the highest value yet reported for Pnp AlGaAs/GaAs HBTs. \[2-25\]

The miniaturization of HBTs and accurate HBT device modeling are still two important unsolved problems, which have limited HBT manufacturability, cost, and circuit applications. The motivation of this dissertation work, as described in chapter 1, is to address some of the issues related to HBT miniaturization and device modeling.
Chapter 3
HBT Device Design

This chapter discusses HBT design issues and reliability considerations as well as the scalability of HBTs. Section 3.1 concentrates on the factors which play important roles in improving HBT current gain. Sections 3.2 and 3.3 describe the $f_i$ and $f_{\text{max}}$ related design parameters and discuss possible approaches for minimizing parasitic components. The importance of ohmic contacts and the requirements on emitter, base and collector resistances are extensively discussed. Section 3.4 discusses reliability considerations with respect to AlGaAs/GaAs HBTs and the effect of surface passivation on HBT reliability. Section 3.5 summarizes this chapter and discusses the design tradeoffs and the scalability issues of HBTs.

3.1 Current Gain

One of the most important parameters in the characterization of HBTs is current gain. HBTs with a high current gain can be used to reduce the number of amplification stages required to achieve a desired circuit performance. High current gain also allows feedback amplifiers to be implemented with ease. The use of feedback provides high linearity at lower power level than are otherwise needed. Because the degradation of current gain, noise and reliability all stem from the same sources, such as surface recombination and traps in the emitter-base junction, high current gain is a good general index of device performance.

3.1.1 Overview

The current gain, $\beta$, is simply, $I_c/I_b$, where $I_c$ denotes the collector current and $I_b$ is the base current. The terminal current $I_c$ and $I_b$, instead of the terminal current density $j_c$ and $j_b$, are used here because a more comprehensive discussion on current gain will include the surface recombination component of $I_b$, which is not proportional to the junction area. In chapter 2, where the operation of HBTs was described, the common emitter current gain $\beta$ is expressed as

$$\beta = \frac{\gamma \alpha}{1 - \gamma \alpha}.$$  \hspace{1cm} (3.1)
By assuming $\alpha=1$, we can define a common emitter current gain, $\beta_e$, representing the hole back-injection effect. When the holes back-injected from the base to the emitter dominate the base current, the common emitter current gain is simply equal to the ratio of the injected electrons to the back-injected holes.

$$\beta = \frac{\gamma}{\gamma-1} = \beta_e = \frac{I_{ne}}{I_{pe}}.$$  \hspace{1cm} (3.2)

By assuming $\gamma = 1$, we can define a common emitter current gain, $\beta_b$, representing all the base current components except the hole back-injection current.

$$\beta = \frac{\alpha}{\alpha - 1} = \beta_b = \left(\frac{I_{bsc} + I_{bb} + I_{bs}}{I_c}\right)^{-1} = \left(\frac{1}{\beta_{bsc}} + \frac{1}{\beta_{bb}} + \frac{1}{\beta_{bs}}\right)^{-1},$$  \hspace{1cm} (3.3)

where $I_{bsc}$ is the base-emitter junction space-charge region recombination current, $I_{bb}$ is the base bulk recombination current, $I_{bs}$ is the base surface recombination current, and $\beta_{bsc}$, $\beta_{bb}$ and $\beta_{bs}$ are the common emitter current gains defined as $I_c/I_{bsc}$, $I_c/I_{bb}$, and $I_c/I_{bs}$, respectively.

Figure 3-1 shows the origins of the various base current components. When all base current components are present and $\beta_e, \beta_b \gg 1$, which generally holds, the equation (3.1) can be rewritten as

$$\frac{I}{\beta} = \frac{I}{\beta_e} + \frac{I}{\beta_{bsc}} + \frac{I}{\beta_{bb}} + \frac{I}{\beta_{bs}}.$$  \hspace{1cm} (3.4)

Equation (3.4) is very useful for estimating the current gain performance of HBTs in device design. Instead of solving Poisson’s equation and the current continuity equation, we can estimate each of these four base current components separately and identify the dominant component among them. By so doing, the design effort focuses on the most limiting factors. The following four sections discuss these four components separately.
Figure 3-1 The origins of the various base current components.

3.1.2 Beryllium Diffusion and Emitter Injection Efficiency

In AlGaAs/GaAs HBTs, because of the wider emitter bandgap, the hole back-injection is virtually eliminated and, $1/\beta_e$ can be assumed to be zero. However, the p-type doping species commonly used in MBE grown wafers is beryllium, which diffuses very fast at high concentrations, and as a result, the base-emitter p-n junction shifts into the AlGaAs grading layer, and the real bandgap is smaller than the designed value. In analyzing the emitter-base current, the effect of Be diffusion on current gain must be incorporated.

Parabolic grading is normally used to smooth the conduction band for electron transport in the AlGaAs grading layer. Under parabolic grading, the band discontinuity for preventing holes from back-injecting as a function of the distance by which the p-n junction has shifted into the AlGaAs layer can be expressed as \cite{3-1}

$$\Delta E_g(z) = 1.247 \cdot \left( x_{\text{emit}} - x_{\text{base}} \right) \left( 1 - \frac{z}{W_d} \right)^2,$$  \hspace{1cm} (3.5)

where $z$ is the p-n junction shift, $W_d$ is the AlGaAs grading layer thickness, and $x_{\text{emit}}$ and $x_{\text{base}}$ are the Al compositions of the emitter and base, respectively. Equation (2.6) can be used to calculate $\beta_e$, i.e., the ratio of electron injection current to hole back-injection.
current. For the AlGaAs emitter, the conduction band density of states, \( N_{c,\text{emitter}} \), and the valence band density of states, \( N_{v,\text{emitter}} \), can be taken as

\[
N_{c,\text{emitter}} = N_{c, GaAs} \left(1 + \frac{0.083 \cdot x_{\text{emitter}}}{0.067} \right)^{3/2},
\]

\[
N_{v,\text{emitter}} = N_{v, GaAs} \left(1 + \frac{0.13447 \cdot x_{\text{emitter}}}{0.0472} \right)^{3/2},
\]

where \( N_{c, GaAs} \) and \( N_{v, GaAs} \) are the densities of states of GaAs, and they are taken as \( 4.7 \times 10^{17} \text{ cm}^{-3} \) and \( 7 \times 10^{18} \text{ cm}^{-3} \), respectively. \[^{3-2}\] The electron minority carrier mobility in the p-type base can be approximated as \[^{3-3}\]

\[
\mu_{n}^{p} = 8300 \left(1 + \frac{N_{\text{base}}}{3.98 \times 10^{15} + \frac{N_{\text{base}}}{641}} \right)^{-1/3},
\]

where \( N_{\text{base}} \) is the base doping concentration and the unit for mobility is \( \text{cm}^{2}/\text{V} \cdot \text{sec} \). Because of a lack of data in the literature, the hole minority mobility in the n-type emitter is taken to be identical to the hole majority mobility. \[^{3-4}\]

\[
\mu_{p}^{N} = \mu_{p}^{P} = \frac{380}{\left(1 + 3.17 \times 10^{-17} \cdot N_{D} \right)^{0.266}}.
\]

The diffusivity \( D_{n} \) and \( D_{p} \) can be calculated from the Einstein relationships, \( \mu_{n}^{P} \cdot kT/q \) and \( \mu_{p}^{N} \cdot kT/q \), respectively. The hole diffusion length in the emitter, \( L_{ep} \), can be calculated from

\[
L_{ep} = \sqrt{D_{p} \tau_{p}},
\]

where \( \tau_{p} \) is the minority carrier lifetime of holes in n-type AlGaAs emitter. Because, the minority carrier lifetime in AlGaAs is very subject to the material quality, there is no good reference to take. By comparing the current gain obtained from our measurements
to that of Sedan simulations, $\tau_p = 5 \times 10^{-11}$ sec. is a good approximation, and this value is used to calculate $\beta_e$. [3-5]

Figure 3-2(a) shows the hole-back-injection-determined current gain, $\beta_e$, as functions of p-n junction shift for different grading conditions and base doping concentrations. The solid circle data points represent devices with a moderate base doping, $5 \times 10^{18}$ cm$^{-3}$, and a grading layer graded from Al$_{0.3}$Ga$_{0.7}$As to pure GaAs. As indicated by this curve, when the Be diffusion causes the base-emitter p-n junction to shift into AlGaAs by about 125Å, the $\beta_e$ drops to about 100, which sets a limit for the highest achievable current gain when more than one component of base current exists. And when the shift is larger than 250Å, the current gain will fall below unity. This degradation has been investigated experimentally by Won-Seong Lee, who intentionally left the Be shutter open during the wafer growth and doped the AlGaAs grading layer with Be to imitate the Be diffusion effect. [3-6] His experimental results shows that when the p-n junction moves into linear-graded AlGaAs by 150Å, the current gain drops to about 100, and when it extends 300Å, the current gain is below unity. Having considered the difference of the linear grading and the parabolic grading, the modeling in Figure 3-2(a) agrees with the experiments very well.

Figure 3-2(a) also shows that when the base doping concentration increases, and/or when the change in Al composition during AlGaAs grading layer decreases, the degradation caused by a p-n junction shift becomes even more severe. Figure 3-2(b) shows the temperature effect on this current gain degradation caused by Be diffusion. Since HBTs usually operate at fairly high current density, the junction temperature of an operating HBT is normally higher than the ambient temperature by 10 to 200°C depending on the device geometry and bias conditions. If the junction temperature is around 175°C, as indicated by the curve plotted in Figure 3-2(b), Be diffusing into AlGaAs by only 75Å can degrade the current gain to below 100.
Figure 3-2  (a) The hole back-injection determined current gain, $\beta_e$, as a function of the p-n junction shift for different base doping concentrations as well as composition grading. (b) The hole back-injection determined current gain, $\beta_e$, as a function of the p-n junction shift for different temperatures.
The diffusion coefficient of beryllium in GaAs is proportional to the square of the hole concentration during growth, and increases dramatically as the growth temperature increases.\textsuperscript{[3-7]} However, the beryllium diffusion during growth is a problem with a moving boundary, which makes simulation very difficult. In practice, the lowest growth temperature of HBTs is determined by the necessity of achieving good quality AlGaAs. This limit is typically 600 to 700°C. To minimize beryllium diffusion, we usually minimize the growth time required to grow the layers above the base. To grow high frequency HBT devices, which require base doping concentrations higher than \(2 \times 10^{19} \text{ cm}^{-3}\), undoped spacers are usually used to prevent beryllium from diffusing into AlGaAs grading layer.

It is very difficult to simulate the Be diffusion processes either during the growth or under current stress. Secondary Ion Mass Spectroscopy (SIMS) measurements do not have sufficient spatial resolution to reveal the diffused beryllium profile. A useful qualitative measurement is the turn-on voltage of AlGaAs/GaAs HBTs from the I-V characteristics of base-emitter junction.\textsuperscript{[3-8]} After beryllium diffuses into the graded AlGaAs, the energy bandgap at the p-n junction position is due to the AlGaAs material, thus the turn-on voltage of the HBT shifts to a larger value. However, in order to determine the turn-on voltage, a reference collector current level is needed, and since the collector current is junction area dependent, accurate turn-on voltage determination relies on accurate measurement of the emitter area or uniformity from wafer to wafer. HBTs processed by wet etch techniques usually do not offer such reproducibility. Therefore, this measurement approach is generally good for measuring the turn-on voltage shift after stressing HBT devices, but not for comparing the beryllium diffusion of differently doped wafer structures. In this dissertation research, a carefully designed base structure was used to fabricate high frequency HBTs; however, the degradation in current gain caused by beryllium diffusion is still not negligible.

3.1.3 Emitter Quality and Space Charge Region Recombination

The second component of base current is space-charge region recombination current, which results from the recombination of electrons and holes in the bulk and around the surface of the base-emitter space-charge region. To determine space-charge recombination current, the current continuity equation and the Poisson equation have to be solved simultaneously. One dimensional simulators, such as SEDAN, and two dimensional simulators, such as PISCES and SEMI-CAD, were extended to handle these
calculations comprehensively yet efficiently. However, even without these sophisticated computer codes, the most important feature of this space-charge recombination current, the ideality factor, can still be reasonably approximated through a general derivation.

To find expressions for generation-recombination in the space-charge region, we use the Shockley-Hall-Read theory. \[3-9\] Assuming equal hole and electron capture cross-sections and evaluating them in the space-charge region with an applied bias \(V_{be}\), the overall recombination-generation rate is

\[
U = \frac{n_i^2 e^{qV_{be}/kT} - 1}{p + n + 2n_i \cosh \left( \frac{E_t - E_i}{kT} \right) \cdot \tau_0} \tag{3.11}
\]

where \(n_i\) is the intrinsic carrier concentration, which is a function of aluminum composition, \(E_i\) is the intrinsic Fermi level, \(E_t\) is the energy level of traps, \(\tau_0\) is the hole and electron lifetime associated with the recombination through traps, and \(p\) and \(n\) are the hole and electron densities in the space-charge region, respectively. The total current arising from generation and recombination is given by

\[
J_r = q \cdot \int_{x_p}^{x_n} U \, dx \tag{3.12}
\]

where \(x_n\) and \(x_p\) are the emitter side edge and base side edge of the space-charge region, respectively. By assuming \(n\) and \(p\) are approximately equal and considering only the traps at the intrinsic Fermi level, the recombination current can be approximated as

\[
J_r = \frac{q(x_n + x_p)n_i}{2\tau_0} \exp \left( \frac{qV_{be}}{2kT} \right) \tag{3.13}
\]

As indicated by equation (3.13), the space-charge recombination current varies with base-emitter bias as \(\exp(qV_{be}/2kT)\). This dependence can be understood intuitively, as on average, each carrier is only required to overcome half of the barrier before recombination takes place. In practice, the general dependence is proportional to
exp(qV_{be}/nkT), where n is called the current ideality factor and, in this case, it is equal to 2.

Among the four components of base current, the space-charge recombination current is the most undesirable one because it not only reduces the achievable current gain but also makes the current gain bias dependent. The collector current normally varies with base-emitter bias as \( \exp(V_{be}/kT) \), so when space-charge recombination current dominates, the current gain, \( \beta_{bse} = I_c/I_{bse} \), has an \( \exp(V_{be}/2kT) \) dependence on bias, which causes signal distortion and non-linearity. In contrast, the other three base current components have an \( \exp(V_{be}/kT) \) dependence on the bias, thus the current gain is uniform as \( I_c \) varies over a few order of magnitude. In Silicon BJTs or GaInP/GaAs HBTs, at high bias, the space-charge recombination usually becomes negligible; however, in AlGaAs/GaAs HBTs, it is always a big portion of the total base current. Because aluminum is extremely reactive, the AlGaAs emitter incorporates a fair amount of foreign species during the growth to form mid-gap traps. Those traps and the presence of DX centers dramatically reduce the recombination lifetime in the AlGaAs space-charge region and generate a large recombination current.

![Gummel Plot](image)

**Figure 3-3** Gummel plot of a 2x10 \( \mu \text{m}^2 \) passivated HBT device and the simulated space charge recombination current with different recombination lifetimes. The solid lines are the measured \( I_b \) and \( I_c \). The dotted lines are the simulated \( I_{scr} \).
Figure 3-3 shows the measured Gummel plot of a 2 µm x 10 µm AlGaAs/GaAs HBT as well as the simulated space-charge recombination base current based on equation (3.13). The \( I_{scr} \) curves are generated by using the intrinsic carrier density of GaAs to approximate the entire space-charge region and taking the total depletion width as 250Å, which is about the value at 1.2 volts base-emitter bias. The recombination lifetimes are set to be 1.0 x 10^{-8} s for the top \( I_{scr} \) and 1.0 x 10^{-6} s for the bottom \( I_{scr} \), respectively. When collector current is below 10 µA (~ 50 A/cm²), base current is dominated by \( I_{scr} \), the base current ideality factor is equal to 2, and the recombination lifetime is shorter than 1.0 x 10^{-6} s. As \( I_c \) increases along with bias, other base current components get involved, and the base current ideality factor becomes smaller than 2. However, unlike what happens to silicon BJTs, the collector current curve and base current curve never get parallel with each other, and as a result, in a \( \beta \) versus \( I_c \) plot, there is a peak instead of a plateau for \( \beta \). As indicated by Figure 3-3, the current gain is higher than 1000, if \( I_{scr} \) is the only source of the base current.

3.1.4 Base Doping and Bulk Base Recombination

The bulk base recombination current is the third component of the base current in a bipolar transistor. If this component dominates, the current gain is simply expressed as

\[
\beta_{bb} = \frac{\tau_n}{\tau_b},
\]

(3.14)

where \( \tau_n \) is the electron life time in bulk base and \( \tau_b \) is the base transit time of electrons, which can be calculated by

\[
\tau_b = \frac{W_b^2}{2D_{nb}}.
\]

(3.15)

By using the dependence of \( \mu_n^p \) on base doping given in equation (3.8) and the dependence of \( \tau_n \) given by [3-10]

\[
\tau_n = \left( \frac{N_{base}}{1 \times 10^{10}} + \frac{N_{base}^2}{1.6 \times 10^{29}} \right)^{-1},
\]

(3.16)

32
the current gain, $\beta_{bb}$, can be plotted as a function of base doping concentration and base width, as shown in Figure 3-4. When the base is moderately doped at a $5 \times 10^{18}$ cm$^{-3}$ level, depending on the base width, $\beta_{bb}$ ranges from 900 to 3800, which is probably higher than the limit set by the other base current components. However when the base is heavily doped at a $3 \sim 5 \times 10^{19}$ cm$^{-3}$, the current gain is almost exclusively determined by bulk base recombination and ranges from 20 to 50 depending on the base width. In practice, to achieve high speed performance, the base is always heavily doped to meet the requirements on base sheet and contact resistances. The design tradeoffs are discussed in section 3.5.1.

![Figure 3-4](image)

Figure 3-4 Bulk base recombination determined current gain $\beta_{bb}$ as a function of base doping concentrations for different base width.
Figure 3-5 SEDAN simulation results for HBTs with different base doping concentrations. (a) HBTs with graded base-emitter junctions, (b) HBTs with abrupt base-emitter junctions.
Figure 3-5(a) and (b) show the SEDAN simulation results on current gain for HBTs with different base doping concentrations. The HBTs in plot (a) have linearly graded base-emitter junctions, while the HBTs in plot (b) have abrupt base-emitter junctions. These SEDAN simulations include both space-charge region and bulk base recombination currents. As we discussed earlier, in AlGaAs/GaAs HBTs, current gain has a strong dependence on the base-emitter bias, even at high base doping levels. All these SEDAN simulations demonstrate the expected current dependence and confirm that bulk base recombination dominates the total base current and that the current gain is about 20 when the base is doped at $4 \times 10^{19} \text{ cm}^{-3}$.

3.1.5 AlGaAs Passivation Ledge and Surface Recombination

The discussion so far has proceeded under assumption that the extrinsic base surface is perfect and that there is no surface recombination current; this is not necessarily true. Surface recombination current could be a major component of the total base current, especially for AlGaAs/GaAs HBTs. [3-11]

The unpassivated GaAs surface at the extrinsic base of HBTs contains a large number ($10^{13} - 10^{14} \text{ cm}^{-2}$) of localized states, probably due to the displaced Ga and As atoms as well as absorbents incorporated during processing. These states may act as charge traps or recombination centers. The high surface charge density requires a redistribution of charge near the semiconductor surface in order to preserve charge neutrality. As a result, the Fermi level is essentially pinned near mid-gap. In p-type GaAs, such as the extrinsic base of an Npn HBT, the Fermi-level is pinned approximately 0.5 eV above the valence-band edge, and in n-type GaAs, the Fermi level is pinned approximately 0.75 eV above the valence-band edge. [3-12] The band bending caused by the Fermi level pinning at the surface creates an electric field which drives minority carriers toward the surface, where they recombine with majority carriers at the surface recombination centers. This high density of surface states, coupled with the band bending near the surface, makes the extrinsic base surface recombination the most critical effect in AlGaAs/GaAs HBTs.

Similar to space-charge region recombination, surface recombination can be modeled by the Shockley-Read-Hall recombination theory. Assuming that there exists a single dominating trap level, the surface recombination rate is
\[ U = \frac{n_s p_s - n_i^2}{(n_s + n_{Ts}) \sigma_n + (p_s + p_{Ts}) \sigma_p} \sigma_n \sigma_p \nu_{th} N_{Ts}, \]  \hfill (3.17)  

where \( n_s \) and \( p_s \) are the surface carrier concentrations, \( n_{Ts} \) and \( p_{Ts} \) are the surface carrier concentrations if the Fermi level were at the trap level, \( \nu_{th} \) is the thermal velocity of the carriers, and \( \sigma_n \) and \( \sigma_p \) are the capture cross sections of the carriers. To simplify equation (3.17), we make several assumptions about the surface condition which lead to very different characteristics of surface recombination current.

Let us first assume that the Fermi level is pinned and that the thermal equilibrium surface concentrations are maintained. To meet the charge neutrality requirement, \( n_s \sigma_n \nu_{th} \) and \( p_s \sigma_p \nu_{th} \) have to be equal, and the recombination rate can be written as

\[ U = S_0 \sqrt{n_s p_s}, \]  \hfill (3.18)  

where \( S_0 \) is the intrinsic surface recombination velocity defined as

\[ S_0 = \frac{N_{Ts} \cdot \sqrt{\sigma_n \sigma_p \nu_{th}^2}}{2}. \]  \hfill (3.19)  

The recombination rate is proportional to the square root of the electron and hole concentration at the surface, and since the quasi-Fermi levels are flat because of the pinning, it is also proportional to the square root of the electron and hole concentration in the bulk. Consequently, the surface recombination current depends on the base-emitter bias with an ideality factor of 2 as the square root of \( n_s p_s \) has an \( \exp(qV_{be}/2kT) \) dependence on \( V_{be} \).

On the other hand, if we assume that thermal equilibrium is not maintained and that the excess minority carrier concentration limits the recombination process. On the p-type extrinsic base surface, the electron is a minority carrier, and the hole concentration is large. Thus, equation (3.17) can be simplified as

\[ U = \sigma_n \nu_{th} N_{Ts} n_s. \]  \hfill (3.20)
The rate of recombination of electrons, per unit time, per unit area, is given by the above expression. The surface recombination velocity generally used to specify the boundary conditions in solving continuity equations is defined as

$$\mu_n \cdot E + D_n \cdot \frac{\partial n}{\partial z} = U = S \cdot n_s,$$  \hspace{1cm} (3.21)

where $\mu_n$ and $D_n$ are the mobility and diffusivity of the electrons in the base, $E$ is the electric field in the base, and $z$ is the position coordinate perpendicular to the surface. Comparing equation (3.21) to equation (3.20), we can express the surface recombination velocity as

$$S = \sigma_n v_{th} N_{Ts}.$$  \hspace{1cm} (3.22)

On a GaAs surface, $N_{Ts} = 10^{14}$ cm$^{-2}$, $\sigma_n = 10^{-15}$ cm$^{2}$, and the thermal velocity $v_{th} = 10^{7}$ cm/s, so the GaAs surface recombination velocity is on the order of $10^6$ cm/s. For the extrinsic base of HBTs, because of the device processing, the surface recombination velocity could be higher or even reach the highest limit, the thermal velocity. Different from equation (3.18), in equation (3.21), the recombination rate is proportional to surface electron density. As a result, the surface recombination current depends on the base-emitter bias with an ideality factor of 1 because surface electron density is proportional to the injected electron concentration, which has an $\exp(qV_{be}/kT)$ dependence on bias.

All the experimental observations we have show that the ideality factors of extrinsic base surface recombination current are closer to 1 instead of 2, which implies that the second situation we assumed in the previous discussion more accurately describes the reality. [3-13] Intuitively, because the Fermi level pinning and high recombination centers make the GaAs surface recombination velocity higher than $10^6$ cm/s, the recombination rate is limited by the supply of carriers from the bulk. Therefore, the recombination rate is proportional to the minority carrier concentration instead of the square root of $n_s p_s$. Moreover, because all the excess electrons reaching the surface are immediately swept out by the high surface recombination, maintaining thermal equilibrium is very unlikely.

Whether the ideality factor of the extrinsic base surface recombination current should be 1 or 2 is very controversial; our experimental results suggest that this ideality factor is closer to 1. In terms of device modeling, we use equation (3.20) in our
simulation except when Poisson's equation and continuity equation are solved simultaneously.

To reduce and eventually eliminate this extrinsic base surface recombination, a thin AlGaAs passivation layer on top of the extrinsic base surface has been used since 1985, when H. H. Lin first invented this passivation ledge structure and demonstrated the improvement on current gain by fabricating such ledges surrounding the emitter mesa of AlGaAs/GaAs HBTs. While the processing approaches and difficulties to implement such passivation ledges into an HBT device are to be discussed in the next chapter, the working mechanism of the passivation ledge can be explained as follows.

Figure 3-6(a) shows a schematic cross section of the emitter and base layers of AlGaAs/GaAs HBT with an AlGaAs ledge covering the extrinsic base surface. Figure 3-6(b) shows the corresponding energy band diagram of a section through the AlGaAs ledge layer and base layer. For comparison, it also shows the extrinsic base region for the case where the base surface is left unpassivated. Under ideal conditions, this ~ 600 - 700 Å thick AlGaAs ledge layer is fully depleted due to the overlapping of the depletion regions resulted from the AlGaAs surface Fermi level pinning and the base-emitter junction bias. In this depleted ledge layer, the electric field, which drives injected minority carrier electrons toward the surface, is proportional to the slope of the conduction band. As shown by the band diagram, the passivated extrinsic base has a much smaller and less penetrating electric field than the unpassivated extrinsic base. By reducing the concentration of minority carrier electrons reaching the surface, the AlGaAs passivation ledge reduces the effective surface recombination velocity and the extrinsic base surface recombination current. The effective surface recombination velocity here means the velocity of electrons leaving AlGaAs/GaAs interface.

To effectively passivate the extrinsic base, achieving complete depletion of the AlGaAs ledge layer is very important. Otherwise, the parasitic conductance through the undepleted portion of the ledge makes the ledge just an extension of the emitter, and some of the electrons are actively injected into the base right next to the unpassivated part of the extrinsic base. As a result, the formerly improved current gain is now degraded.
Figure 3-6 (a) A schematic cross section of the emitter and base layers of AlGaAs/GaAs HBT with an AlGaAs ledge, (b) the corresponding energy band diagram of a section through the AlGaAs ledge layer and base layer.

Since the depleted AlGaAs ledge eliminates the extrinsic base surface recombination in general, this passivation structure can not only increase the current gain but also reduce 1/f noise as well as improve the HBT reliability. [3-14] According to the research work done by Damian, the 1/f noise is mainly associated with the extrinsic base surface. By using an AlGaAs surface passivation ledge, the reduction of 1/f noise involved in AlGaAs/GaAs HBTs has been demonstrated. The improvement in HBT reliability by using an AlGaAs passivation ledge structure has been reported by Hewlett Packard in 1991 and by Texas Instruments in 1994. [3-15] More discussion on reliability considerations in HBT design is given in section 3.4 of this chapter.
3.2 The Cutoff Frequency, $f_t$

As introduced in section 2.2.1, the high frequency performance of a transistor can be characterized by its cutoff frequency, $f_t$, maximum oscillation frequency, $f_{max}$, and gate propagation delay time, $t_d$. In this section, in particular, we consider how the design of the HBT epitaxial structure and structural layout effects $f_t$.

3.2.1 The total emitter-collector transit time

The cutoff frequency is inversely proportional to the total emitter-collector transit time, which includes four terms as shown by equation (2.7)

$$\tau = \frac{1}{2 \pi f_t} = \tau_e + \tau_b + \tau_{sc} + \tau_c.$$  \hspace{1cm} (3.23)

This equation can be derived either by the charge control model or the equivalent circuit model.\textsuperscript{[3-16]} The first term, i.e., the emitter charging time, is the time required to charge the base-emitter junction capacitance, $C_{be}$, when the base-emitter junction bias ($V_{be}$) is changed to the desired operating value.

$$\tau_e = \frac{C_{be}}{g_m} = r_e' C_{be} = \frac{kT}{qI_c} C_{be},$$  \hspace{1cm} (3.24)

where $g_m$ is the transconductance of the transistor, $r_e'$ is the differential base-emitter junction resistance, and $I_c$ is the collector current. Equation (3.24) assumes that the delay associated with holes that are stored in the quasi-neutral emitter is negligible compared to the other delay components. This assumption is generally true for HBTs because of very low hole back-injection, but not for BJTs, in which the storage of holes in the quasi-neutral emitter can constitute a significant portion of the total delay.

To reduce emitter charging time, we have to either increase the collector current density or decrease the base-emitter junction capacitance. RF HBT devices usually operate under high collector current density, around $6 \sim 10 \times 10^4$ A/cm$^2$, which reduces the emitter charging time below the other terms in equation (3.23). However, high collector current causes two serious problems: self-heating and high power consumption, both of which affect the reliability and the integration density of HBT-ICs. Scaling HBT
devices down to sub-micron sizes can reduce the heat generation and the total power consumption while maintaining the same high collector current density. Decreasing the base-emitter junction capacitance is the other alternative that has been demonstrated by fabricating HBTs with a lightly doped emitter. The more lightly the emitter is doped, the wider the base-emitter depletion region becomes, and $C_{be}$ is smaller. The potential drawback of this approach is that the emitter injection efficiency is reduced; however, with sizable emitter-base heterojunction bandgap differences, a considerable range of tradeoff exists.

The second term, $\tau_b$, is the base transit time, which is the base delay due to the storage of holes in the quasi-neutral base in order to neutralize the charge associated with electrons crossing the base region. It is given as

$$\tau_b = \frac{W_b^2}{2 \cdot D_{nb}} f(\kappa), \quad (3.25)$$

where

$$f(\kappa) = \frac{2}{\kappa} \cdot \left(1 - \frac{1}{\kappa} + \frac{1}{\kappa} \cdot e^{-\kappa}\right), \quad (3.26)$$

$$\kappa = \frac{1}{kT} \Delta E_g, \quad (3.27)$$

$\Delta E_g$ is the bandgap difference at the two ends of the base, $kT$ is the thermal energy, $W_b$ is the base thickness, and $D_{nb}$ is the electron minority-carrier diffusivity. The factor $\kappa$ represents the base quasi-electric field strength. With an aluminum grading from 0.1 to 0.0 across the base, $\kappa$ ~ 4.83, and $f(\kappa) = 0.329$. If the base is purely GaAs, $\kappa = 0$, $f(\kappa) = 1$ by taking the limit as $\kappa$ approaches zero, and equation (3.25) becomes identical to equation (3.15).

To reduce the base transit time, we can either reduce the base width or add a quasi-electric field in the base by grading the base layer. Because of the square dependence, the base transit time can be dramatically reduced by thinning the base. However, thinning the base increases the base sheet resistance and thus decreases $f_{max}$. In addition, fabricating a base thinner than 500Å would be very challenging and causes yield problems. The base width used for Npn RF HBTs usually varies from 500 to 1000Å. The other approach grades the base region and drives the carrier drifting through
the base. The most commonly used grading method changes the aluminum composition across the base and builds a quasi-electric field due to the bandgap difference. The aluminum composition of the bulk emitter is usually chosen to be 0.3 or lower in order to minimize the concentration of the donor-aluminum DX recombination centers. [3-18] Therefore, the largest composition difference available for base grading is about 0.1 because at least a 0.2 aluminum composition difference has to be used for the base-emitter junction to maintain the junction bandgap discontinuity and high injection efficiency. Fortunately, this 0.1 to 0.0 composition grading can reduce the base transit time by a factor of 3, thus making the base transit time insignificant in the total delay. The drawbacks of this approach, such as degrading the base quality slightly by having a certain amount of aluminum in the base or complicating the epitaxial growth, are minor when compared to its advantages.

The third term \( \tau_{sc} \) is the transit time across the base-collector space charge region, which is the time delay due to hole charge stored in the base-collector space charge region. It is given as

\[
\tau_{sc} = \frac{W_{bc}}{2v_{sat}},
\]  

(3.28)

where \( W_{bc} \) is the base-collector depletion width and \( v_{sat} \) is the saturation velocity of electrons. \( W_{bc} \) usually takes the value of the collector thickness because the collector is lightly doped and fully depleted under normal base-collector reverse bias. In this derivation, it is assumed that the electrons travel at a constant saturation velocity, \( v_{sat} \), because the electric field is large throughout the entire reverse-biased base-collector junction.

To reduce base-collector space charge region transit time, \( \tau_{sc} \), we either have to use a relatively thin collector layer or have to change the dependence of the transit time on the saturation velocity. Since the collector thickness is usually determined by the breakdown voltage required for certain circuit applications, we usually do not have the freedom to reduce the collector thickness. However, the dependence of the transit time on the saturation velocity can be changed by design. For Npn AlGaAs/GaAs HBTs, it is possible to design the collector layer such that the field strength in the base-collector space charge region is reduced such that electrons travel at a velocity determined completely by the \( \Gamma \) valley. [3-19] Without being scattered to the L valley, the electrons
continue to travel at a overshoot velocity that is much larger than $v_{sat}$. Therefore, $\tau_{sc}$ can be significantly reduced by careful design.

Equation (3.28) is not strictly valid at high collector currents with the onset of the Kirk effect. When the collector doping concentration is designed to be low in order to reduce the field strength, the collector current density can be high enough that the ionized donors in the depletion region are neutralized by the injected electrons. The onset condition is $J_c \geq q \cdot v_{sat} \cdot N_D = 8 \times 10^4 \text{ A/cm}^2$ for a collector doping of $N_D = 5 \times 10^{16} \text{ cm}^{-3}$. When the collector current density is larger than this onset value, the electric field inside the base-collector space charge region is dramatically reduced and eventually disappears, and the electron transport becomes partially diffusion rather than purely drift as assumed in equation (3.28). This diffusion transport on the part of the collector makes the base equivalently wider, so the Kirk effect is also called the Base Push-out effect. As a result, the current gain and cutoff frequency both decrease rather dramatically when the collector current density becomes higher than the onset value.

The last term, $\tau_c$, is the collector charging time, which is a RC time constant for charging the base-collector junction capacitance in response to an incremental input voltage with the output terminals short-circuited. It is given as

$$
\tau_c = \left( R_E + R_c + r_e \right) C_{bc} = \left( R_E + R_c + \frac{kT}{q I_c} \right) C_{bc} ,
$$

where $R_E$ is the series emitter resistance, $R_c$ is the series collector resistance, and $C_{bc}$ is the base-collector junction capacitance, which includes intrinsic and extrinsic base regions. The collector charging time can often dominate other terms in equation (3.29) at high collector current levels.

To reduce the collector charging time, we have to reduce these resistive and capacitive elements: $R_E$, $R_c$, and $C_{bc}$. The series emitter resistance, $R_E$, is almost always dominated by the emitter contact resistance; the series collector resistance, $R_c$, is determined by both sub-collector contact resistance and sub-collector sheet resistance. Approaches to make good ohmic contact and reduce these series resistances are discussed in section 3.2.2 and section 3.2.3. The base-collector junction capacitance is proportional to the junction area (including both the intrinsic and extrinsic bases) and inversely proportional to the collector thickness provided that the collector is completely depleted.
during operation. While the collector thickness is usually predetermined by breakdown voltage requirement, reducing the base-collector junction area, in particular the extrinsic base-collector junction area becomes very important, especially when HBTs are scaled down to sub-micron emitter sizes. The design consideration of $C_{be}$ is discussed in section 3.2.4.

The overall transit time is the sum of these four terms and can be expressed as

$$\tau_{ec} = \frac{kT}{qI_c} \left( C_{be} + C_{bc} \right) + \frac{W_b^2}{2 \cdot D_{nb}} f(\kappa) + \frac{W_{bc}}{2v_{sat}} + \left( R_E + R_c \right) C_{bc}. \tag{3.30}$$

Figure 3-7 can be used to visualize the various charging and transit times across each region. The equation (3.30) and the previous discussion give a brief yet complete description of design considerations with respect to HBT high frequency performance. From the description, HBTs have several advantages in achieving high speed operation when compared to Si BJT. First, the emitter can be doped lighter, so $C_{be}$ decreases, and thus the emitter charging time is significantly reduced. Second, the possibility of base bandgap grading enables one to add a large base quasi-electric field, which reduces the base transit time to negligible values. Third, the high overshoot velocity in GaAs makes the base-collector space charge region transit time much smaller than $W_{bc}/2v_{sat}$ provided that proper collector design is used. Finally, the availability of a semi-insulating GaAs substrate can greatly reduce the parasitic pad and interconnect capacitances and thus improve overall GaAs based IC performance.

![Diagram of HBT cross section](image)

Figure 3-7 An idealized HBT cross section for illustrating the various resistive elements, capacitive elements, and transit regions, which comprise the total emitter-collector transit time, $\tau_{ec}$.
3.2.2 Ohmic Contacts to GaAs

The previous discussions have shown how the series emitter and collector resistances would affect the cutoff frequency of HBTs. In all these parasitic resistive elements, contacts to the emitter, base or collector contribute a big portion of the total resistances. Therefore, achieving a stable, low-resistance ohmic contact is essential for obtaining superior high frequency performance. In spite of being as much a technical art as a science, achieving stable, low resistance ohmic contacts involves a lot of design considerations as well as processing principles.

An ohmic contact means that the contact has a linear I-V characteristic, that the contact is stable over time and temperature, and that the contact contributes as little resistance as possible. Theoretically, if the work function of the contact metal is smaller than that of an n-type semiconductor, or larger than that of a p-type semiconductor, this metal could make a perfect ohmic contact with the n-type or p-type semiconductor, respectively. However, when metals contact GaAs, the metal Fermi level is always aligned with the middle of the GaAs bandgap because of surface Fermi level pinning caused by the high density of surface traps. Therefore, in practice, all contacts to GaAs are rectifying contacts instead of ohmic ones, although some are called ohmic contacts when the I-V characteristics are relatively linear and contact resistances are small.

Across a rectifying contact, the current conduction processes are thermionic emission at light semiconductor doping levels, field emission at high doping levels, or thermionic-field emission at intermediate doping levels. To make rectifying contacts to GaAs similar to an ohmic contact, field emission has to be the dominant conduction process, where reducing the contact barrier height or thickness can greatly decrease the contact resistance.

In order to reduce the barrier height, narrower bandgap materials have to be used because in GaAs, the contact barrier height is almost a constant, which is controlled by the surface traps. A suitable candidate is InGaAs or ins because the Fermi level is pinned above the conduction band in ins. However, because of the lattice mismatch between GaAs and InGaAs, this InGaAs contact can only be placed on the top of the wafer so that the device’s epitaxial structure and material quality would not be affected. For the Npn
AlGaAs/GaAs HBTs used in this dissertation research, this extra InGaAs contact layer was used to reduce emitter contact resistance; the details are given in next section.

In order to reduce the contact barrier thickness, the semiconductor has to be heavily doped near the surface to make the surface depletion width thinner than a few nanometers such that electrons might tunnel through this barrier. It is relatively straightforward for Be doped p-type GaAs because the Be doping concentration can be as high as $1 \times 10^{20} \text{ cm}^{-3}$, provided that the Be diffusion has been properly handled and does not cause performance degradation. For the RF Npn AlGaAs/GaAs HBTs used in this dissertation research, the p-type GaAs base is usually doped as high as $5 \times 10^{19} \text{ cm}^{-3}$ to make the base contacts ohmic.

In Si doped n-type GaAs, however, the surface layer could not be doped high enough. At a relatively low doping level, Si usually takes the Ga site and becomes a donor. But at a high doping level, Si can either take the Ga site serving as a donor or take the As site serving as an acceptor, and these donors and acceptors compensate each other. The highest doping concentration in Si doped n-type GaAs is about $5 \times 10^{18} \text{ cm}^{-3}$. For RF Npn AlGaAs/GaAs HBTs, the sub-collector doping concentration is $5 \times 10^{18} \text{ cm}^{-3}$. To make ohmic contacts to the sub-collector, an alloy contact has to be used. The most widely used alloy contact is Au/Ge/Ni. During the rapid thermal annealing process, Ge is driven into and effectively dopes the GaAs surface layer so that the low contact resistance becomes achievable. The processing details are left for the next chapter.

The generally used parameter for measuring an ohmic contact is the specific contact resistance, $r_c$, which is defined as the voltage drop divided by the current flowing perpendicularly through a unit area. The specific contact resistance has units of $\Omega \cdot \text{cm}^2$. There are two kinds of current flowing configurations, as shown in Figure 3-8. If the current flows perpendicularly through an contact area, the contact resistance is simply expressed as

$$R_c = \frac{r_c}{A}.$$  \hspace{1cm} (3.31)

If the current flows from one contact pad to another contact pad through a thin conducting layer, the contact resistance for each pad can be expressed as [3-22]
\[ R_c = \sqrt{\frac{R_{sh} r_c}{W}} \cdot \cosh \left( \frac{d}{L_t} \right) , \quad (3.32) \]

and

\[ L_t = \sqrt{\frac{r_c}{R_{sh}}} , \quad (3.33) \]

where \( R_{sh} \) is the sheet resistance of the thin conducting layer, \( W \) is the contact pad width, \( d \) is the contact pad length, and \( L_t \) is called the transfer length, which is a quantity related to the lateral distance required for current to flow into or out of the ohmic contact. When the contact pad length is larger than the transfer length, \( \cosh(d/L_t) \) approaches unity. Therefore, the contact pad length -- the lateral base and collector dimension in the case of an HBT -- is usually designed to be equal or a little longer than the transfer length.

![Diagram of current flow configurations](image)

**Figure 3-8** Two kinds of current flow configurations: (a) current flows perpendicularly through an contact pad; (b) current flows from one pad to another through a thin conducting layer.

In the second configuration, the total resistance between the two contact pads is

\[ R = 2R_c + \frac{R_{sh2} L}{W} = 2\sqrt{\frac{R_{sh} r_c}{W}} \cdot \cosh \left( \frac{d}{L_t} \right) + \frac{R_{sh2} L}{W} , \quad (3.34) \]

where \( R_{sh1} \) is the sheet resistance of the conducting layer underneath the metal pads, \( R_{sh2} \) is the sheet resistance of the conducting layer between the metal pads, and \( L \) is the space between the metal pads. \( R_{sh1} \) and \( R_{sh2} \) have the same values if the contacts are non-
alloyed so that the conducting layer underneath the pads has not been affected. However, $R_{sh1}$ could be very different and even non-uniform if the contacts have been alloyed.

In this dissertation research, we assumed that $R_{sh1}$ and $R_{sh2}$ have the same values. By assuming $R_{sh1} = R_{sh2}$, we could use Transmission Line Model (TLM) to measure the sheet resistance and contact resistance. In our TLM measurement, the spaces between contact pads were 80, 40, 20, 10, 5 microns, respectively. For each space, a 3 mA current was forced through the two contact pads and the total voltage drop over the contacts and conducting layer underneath the space was measured. According to equation (3.34), when R versus L is plotted, the slope of the line is proportional to the sheet resistance and the intercept with y-axis is $2R_c$, so both sheet resistance and specific contact resistance are measured.

3.2.3 Emitter and Collector Resistances

As discussed in section 3.2.1, smaller series emitter and collector resistances can dramatically reduce the collector charging time and increase $f_t$. When the emitter size is small, the emitter resistance usually dominates the sum of these two resistive components. Theoretically, the emitter resistance includes two parts - the emitter contact resistance and the emitter mesa resistance. Figure 3-9 shows a schematic drawing of the emitter mesa structure used for all HBT devices in this dissertation research. An InAs contact layer was always used to make ohmic contact achievable, and an InGaAs grading layer was used to smooth the bandgap discontinuity. Because the cap layer, the InGaAs grading layer, and the InAs contact layer were all heavily doped, the conduction band was flat over all these layers because of Fermi level alignment.

Figure 3-9 A schematic drawing of the emitter mesa structure.
Because the current flow is perpendicular to the emitter contact interface, the contact resistance is simply equal to the specific contact resistance divided by the emitter contact area. The lowest measured specific contact resistance for our emitter mesa structure was $7.9 \times 10^{-8} \ \Omega \cdot \text{cm}^2$. The specific contact resistance measured from our real RF HBT wafers lay between the lower limit $1.5 \times 10^{-7} \ \Omega \cdot \text{cm}^2$ and the higher limit $3.0 \times 10^{-6} \ \Omega \cdot \text{cm}^2$. The wide data range and larger values of $r_c$ on our real wafers probably resulted from the ion implant masking process; more discussion on implantation process is given in next chapter. Table 3-1 lists the emitter contact resistances for different size emitters with higher and lower limits of specific contact resistances.

For the devices with 2 $\mu$m by 10 $\mu$m emitter size, the emitter contact resistances ranged from a few ohms to about fifteen ohms. If the emitter size is scaled down to a half-micron or even smaller, the emitter contact resistance could become larger than 60 ohms, which would certainly degrade the high frequency response of these submicron devices. Among other challenges, reducing emitter contact resistance is one that has to be solved in order to realize sub-micron AlGaAs/GaAs HBTs. The normal approach is to reduce the specific emitter contact resistance which involves the surface material growth, surface cleanliness and metalization processes as well as the overall processing sequence. Although the implantation lithography was the first step in our RF HBT processing, it is usually better to deposit the emitter metal on the fresh epitaxial surface so that the surface is the cleanest and with the best morphology.

<table>
<thead>
<tr>
<th>Emitter Size</th>
<th>Contact Resistance ($\Omega$)</th>
<th>Mesa Resistance ($\Omega$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>($\mu$m x $\mu$m)</td>
<td>$r_c = 1.5 \times 10^{-7} \ \Omega \cdot \text{cm}^2$ lower limit of measured specific contact resistances</td>
<td>$r_c = 6.5 \times 10^{-8} \ \Omega \cdot \text{cm}^2$ calculated equivalent specific mesa resistance</td>
</tr>
<tr>
<td>2 x 10</td>
<td>0.75</td>
<td>0.325</td>
</tr>
<tr>
<td>1 x 10</td>
<td>1.50</td>
<td>0.650</td>
</tr>
<tr>
<td>0.5 x 10</td>
<td>3.00</td>
<td>1.300</td>
</tr>
</tbody>
</table>

Table 3-1 Emitter contact resistance and mesa resistance.

Table 3-1 also shows the calculated mesa resistances for the emitter mesa structure shown in Figure 3-9. Because the current flows vertically down through the emitter mesa, an equivalent specific mesa resistance can be defined as the voltage drop across the mesa divided by the emitter current density. This specific mesa resistance can be estimated provided that the carrier densities and mobilities of all layers are known.
For the specific mesa resistance of our mesa structure, InAs and InGaAs layers are assumed to have the same electron mobility as GaAs; AlGaAs electron mobilities are assumed to be the GaAs electron mobility multiplied by the Ga composition. Although this calculated result may have a ten to fifty percent uncertainty, the specific mesa resistance is certainly small compared to the specific emitter contact resistance. Thus, the emitter mesa resistance is generally negligible unless the mesa area is many times smaller than the emitter contact area.

The series collector resistance includes three parts: intrinsic collector resistance, $R_{\text{cint}}$, extrinsic collector resistance, $R_{\text{ext}}$, and collector contact resistance, $R_{\text{cc}}$. Figure 3-10 shows a schematic diagram of the HBT cross-section. The total collector resistance is given by

$$R_c = \frac{1}{2} \left( R_{\text{cint}} + R_{\text{ext}} + R_{\text{cc}} \right) = \frac{1}{2} \left( \frac{1}{3} \cdot \frac{w_e}{W} \cdot R_{\text{shc}} + \frac{L_{\text{ec}} R_{\text{shc}}}{W} + \frac{r_c R_{\text{shc}}}{W} \right), \quad (3.35)$$

where $W_e$ and $W$ are the width and length of emitter, respectively, $L_{\text{ec}}$ is the lateral distance from emitter edge to collector metal, $R_{\text{shc}}$ is the sub-collector sheet resistance, and $r_c$ is the specific collector contact resistance. In equation (3.35), it is assumed that the transfer length of the collector contact is smaller than the collector metal. The factor of 1/3 in the expression of $R_{\text{cint}}$ results from the distributed current flow from the sub-collector to the collector. \[3-24\] The factor of 1/2 in the expression of total $R_c$ is because of the two symmetric collector metal contacts.

The sheet resistance of a thin uniformly-doped semiconductor layer can be calculated by

$$R_{\text{sh}} = \frac{I}{nq\mu d}, \quad (3.36)$$

where $n$ is the majority carrier concentration, $\mu$ is the carrier mobility, and $d$ is the layer thickness. When a 3000Å thick n-type GaAs layer is doped at $2 \times 10^{18}$ cm$^{-3}$ or $5 \times 10^{18}$ cm$^{-3}$, the sheet resistances calculated by using equation (3.36) are 43.36 $\Omega$ and 21.45 $\Omega$, respectively. However, our measured sheet resistance from a 3000Å thick, $5 \times 10^{18}$ cm$^{-3}$ Si-doped GaAs layer is 32 $\Omega$, which suggests that the electron concentration is actually lower than the doping level. Our RF HBTs have a sub-collector of 6000Å, and the...
collector sheet resistance measured from the processed wafers is 16.7 $\Omega$ and consistent with the 32 $\Omega$ from a 3000Å thick layer. For sub-collector layers with different thicknesses, the sheet resistance could always be predicted accordingly during the device design.

![Diagram of HBT cross section](image)

Figure 3-10 A schematic drawing of HBT cross section for illustrating the collector resistance components.

The specific collector contact resistance, $r_c$, depends on the collector doping concentration and contact process. For the Npn AlGaAs/GaAs HBTs in this dissertation, the sub-collector was always Si-doped at $5 \times 10^{18}$ cm$^{-3}$ and Au/Ge/Ni/Au was usually used and alloyed to form ohmic contacts. Figure 3-11 shows the specific collector contact resistance as a function of rapid thermal annealing (RTA) temperature for two different sub-collector doping levels. The results indicate that a higher sub-collector doping concentration yields a lower specific contact resistance although the difference is small. Figure 3-12 shows the specific contact resistance as a function of RTA temperature for different length of time. While 30 seconds of RTA was generally believed to be adequate for forming ohmic contacts, our experimental results suggests that an RTA time as short as one second is long enough, and over annealing would result in a higher contact resistance.

The RTA conditions used for real HBT wafers were 5 seconds at 420°C under nitrogen ambient. The specific contact resistance was $2 \times 10^{-6}$ $\Omega$ cm$^2$ or lower. Therefore, the collector transfer length was about 3.5μm, which was shorter than the designed collector metal width 5 μm so that equation (3.35) could be used to calculate the total collector series resistance.
Figure 3-11 The specific collector contact resistance as a function of RTA temperature for two different sub-collector doping levels.

Figure 3-12 The specific contact resistance as a function of RTA temperature for different length of time.

3.2.4 Base-Collector Capacitance

The total base-collector capacitance is equal to base-collector junction capacitance per unit area, $C_j$, multiplied by the total junction area, $A_{bc}$, provided that the base mesa
fringing capacitance is negligible. \(C_j\) is inversely proportional to the base-collector junction depletion width. Because collectors are lightly doped and HBTs are usually operating under relatively high base-collector bias, the collector region is almost always fully depleted, so that in practice \(C_j\) becomes inversely proportional to the collector thickness. For a 1 \(\mu\)m thick GaAs collector, \(C_j\) is 0.114 \(\text{fF} / \mu\text{m}^2\). In order to achieve high speed operation, fully depleting the lightly doped collector layer is very important because any undepleted portion of collector would contribute a large series resistance.

![Figure 3-13](image-url) A schematic drawing of HBT cross section illustrating the different components of base-collector capacitance associated with the intrinsic and extrinsic base areas.

The total base-collector junction area \(A_{bc}\) consists of two parts: intrinsic junction area, which is the area underneath the emitter mesa, and the extrinsic junction area, which is the area underneath the passivation ledge, base metal contacts, and metal-to-ledge spacing. Figure 3-13 shows the different components of base-collector capacitance associated with the intrinsic and extrinsic base area. Because a larger \(C_{bc}\) will cause both \(f_t\) and \(f_{\text{max}}\) to drop and propagation delay \(\tau_d\) to increase, the mission of design optimization is to reduce the base-collector junction area, especially the extrinsic part. Given the device length, i.e., the dimension in the direction perpendicular to the cross-section drawing, the extrinsic base-collector junction area can be expressed as

\[
A_{bc\ extr} = W \left( L + S + W_{bm} \right),
\]

where \(W\) is the device length, \(L\) is the passivation ledge length, \(S\) is the base-metal-to-ledge spacing, and \(W_{bm}\) is the base-metal width. In order to minimize \(A_{bc\ extr}\), we have to minimize the dimensions of passivation ledges, the spacing between the ledge and base.
contact and the base contact. Chapter 5 is devoted to the investigation of minimum passivation ledge lengths. $W_{bm}$ is usually designed to be equal to or wider than the transfer length of the base contact, which is about 1 µm or less provided that base doping concentration is higher than $2 \times 10^{19}$ cm$^{-3}$. (See section 3.3.2 for more details.)
3.3 The maximum oscillation frequency, $f_{\text{max}}$

As introduced in section 2.2.1, the maximum oscillation frequency, $f_{\text{max}}$, is the second commonly quoted figure of merit, which is associated with the power gain of the device and very important for microwave applications. The previous section has shown that the cutoff frequency does not depend on the series base resistance. This fact can be understood because the current which charges the capacitances does not come from the base, but is injected through the emitter. However, the derivation in section 3.3.1 shows that the maximum oscillation frequency strongly depends on the base resistance. Section 3.3.2 discusses the different components of base resistance and how to reduce them.

3.3.1 Derivation of $f_{\text{max}}$

Unlike the current gain, the power gain of the transistor is quite sensitive to the source and load impedances. In order to deliver the maximum power to the load, the input and output of the transistor must be conjugately matched to the source and load, respectively. When the two-port has been simultaneously matched and the feedback parameter neutralized to zero, the maximum available power gain is then defined as the unilateral power gain, $U$. This gain is the highest possible gain that an active two-port network can deliver. The maximum oscillation frequency is the frequency at which this unilateral power gain drops to unity. In terms of $y$-parameters, the unilateral power gain is given by $^{[3-26]}$

$$
U \equiv \frac{|y_{21} - y_{12}|^2}{4 \cdot \left( Re(y_{11}) \cdot Re(y_{22}) - Re(y_{12}) \cdot Re(y_{21}) \right)}.
$$

(3.38)

Experimentally, $U$ is calculated from the measured $s$-parameters which can be converted into $y$-parameters, thus allowing equation (3.38) to be used.

However, because the input and output impedances of the transistor are complex at high frequencies, the simultaneous conjugate-matching condition at both ports may cause the device to be unstable. When the device is conditionally stable, the maximum power gain is defined as the maximum stable gain (MSG). In real applications, MSG tends to be more useful than $U$ because most transistors are not unconditionally stable over the entire frequency range. In terms of $h$-parameters, the MSG is given by $^{[3-27]}$

55
\[ MSG = \frac{|h_{21}|^2}{4 \cdot \text{Re}(h_{11}) \cdot 	ext{Re}(h_{22}) - 2 \cdot \text{Re}(h_{12} \cdot h_{21})}. \quad (3.39) \]

The difference between MSG and U is less than or equal to 3 dB and when the operating frequency is much larger than the critical frequency above which the device becomes unconditionally stable they approach the same value. Because of the simplicity of equation (3.39), it can be used to identify the physical parameters that limit the maximum oscillation frequency if we assume that the difference between MSG and U is negligible in the high frequency range.

In a first order analysis, neglecting the effects of emitter and collector series resistances on power gain, the input resistance \( \text{Re}(h_{11}) \) is \( \sim R_b \), and the output conductance \( \text{Re}(h_{22}) \sim 2\pi f_c C_{bc} \). At high frequency, assuming the reverse-current gain \( (h_{12}) \) is zero and taking the magnitude of the forward-current gain at high frequencies to be \( |h_{21}| = f_l/f_c \), the MSG can be expressed as

\[ MSG = \frac{1}{8\pi \cdot R_b C_{bc}} \frac{f_l}{f^2}. \quad (3.40) \]

By setting MSG equal to unity, the maximum oscillation frequency \( f_{\text{max}} \) is therefore given by the classical expression

\[ f_{\text{max}} = \sqrt{\frac{f_l}{8\pi R_b C_{bc}}}. \quad (3.41) \]

This expression of \( f_{\text{max}} \) illustrates that in order to achieve the highest \( f_{\text{max}} \) both the base-collector junction capacitance and base resistance must be minimized. Section 3.2.4 discussed the base-collector junction capacitance and the next section discusses the base resistance.

### 3.3.2 Base resistance

Lower base resistance, which yields higher maximum oscillation frequencies, is one of the most important advantages HBTs have over BJTs. Because of the emitter-base
heterojunction energy band differences, the base of HBTs can be doped much higher compared to BJTs and still achieve useful emitter injection efficiency. This high base doping not only reduces the base sheet resistance but also decreases the base contact resistance, thus much lower total base resistances can be achieved in HBTs.

The total base resistance includes three parts: intrinsic base resistance, $R_{\text{bintr}}$, extrinsic base resistance, $R_{\text{bext}}$, and base contact resistance, $R_{\text{bc}}$. Figure 3-14 shows a schematic drawing of these three components of total base resistance, which can be expressed as

$$
R_b = \frac{1}{2} \left( R_{\text{bintr}} + R_{\text{bext}} + R_{\text{bc}} \right) = \frac{1}{2} \left( \frac{w_b}{W} \cdot R_{\text{shb}} + \frac{(L+S)}{W} \cdot R_{\text{shb}} + \frac{\sqrt{r_c R_{\text{shb}}}}{W} \right).
$$

(3.42)

where $R_{\text{shb}}$ is the sheet resistance of base layer and $r_c$ is the specific contact resistance of base metal contacts.

![Figure 3-14 A schematic drawing of HBT cross section for illustrating the three components in total base resistance. The width of the structure along the dimension perpendicular to the cross section is W.](image)

Base sheet resistances can be calculated from equation (3.36) if we use $n$ as the hole concentration, $\mu$ as the hole mobility and $d$ as the base width. Base sheet resistances can also be measured by fabricating transmission line patterns on top of the base layer. Figure 3-15 shows the calculated as well as the experimentally measured sheet resistance values for base layers with different base widths. The solid and open squares represent the measured and calculated values of 1500Å thick base layers, respectively. Considering that there are no fitting parameters in equation (3.36), the differences between the
experimental and theoretical values are small (less than 25%) and probably result from the accuracy of the mobility model given by equation (3.9). Because the errors are relatively small, the calculated values can be used for base structure design.

![Graph](image)

Figure 3-15 The calculated and experimentally measured sheet resistance values for p-type GaAs base layers with different base width.

The specific contact resistance of base metal contacts strongly depends on the base layer doping concentration as well as the metal used. Figure 3-16 shows the measured specific contact resistances of base layers doped at different concentrations and contacted by different metal films. The squares are for 200Å-Ti/200Å-Pt/2000Å-Au base metal contacts, and the circles are for 100Å-Pt/200Å-Ti/100Å-Pt/2000Å-Au base metal contacts. The Pt/Ti/Pt/Au contacts yield lower specific contact resistances than the Ti/Pt/Au contacts. However, our base metalization process shows that Ti/Pt/Au contacts have better adhesion. Moreover, after being annealed at 420°C for 30 seconds as shown by Figure 3-16, the Ti/Pt/Au specific contact resistances decrease while the Pt/Ti/Pt/Au contact resistance increases. As a result, the improvement in the contact resistances by using Pt/Ti/Pt/Au becomes smaller after the annealing process, which is not required for the base contacts but inevitable because the collector contacts must be alloyed. Considering that good adhesion is essential for small size HBTs, the Ti/Pt/Au contact was chosen for our RF HBT fabrication.
Figure 3-16 The measured specific contact resistances of p-type GaAs base layers doped at different concentrations and contacted by different metal films.

Unlike the base sheet resistances, the specific contact resistances are very sensitive to surface conditions. Because the surface of the test wafers was very different from the real base surface, the measured base specific contact resistances from the real HBT wafers were much higher than the values plotted in Figure 3-16. The test wafers we used were MBE grown, and the metalization process was done directly on the top of the epitaxial surface. However, for HBT wafers, the base layer is underneath the emitter layer, and the base surface was exposed by a wet chemical etch process. Because of photoresist and wet etching, there is some surface contamination that is avoided on the test wafers. Thus, the wet etched base surface was worse than the epitaxial surface, and consequently the base specific contact resistances were higher. The measured values from the HBT wafer with base doping concentration of $5 \times 10^{19}$ cm$^{-3}$ are about $8 \times 10^{-6}$ Ω-cm$^2$, compared to about $5 \times 10^{-7}$ Ω-cm$^2$ on the test structures.

After knowing the device geometry, $R_{shb}$ and $r_e$, the total base resistance $R_b$ can be calculated by using equation (3.42) if the base metal width $W_{bm}$ is designed to be larger than the base contact transfer length. Figure 3-17 shows the base contact transfer lengths for different base width and base doping concentrations. The calculated sheet resistances
and the measured specific contact resistances (with Ti/Pt/Au after annealing) are used to plot Figure 3-17. As shown by the figure, the thinner the base layer, the narrower the base metal can be because of current crowding. If the base metal width is designed to be 1 μm, then the base doping concentration has to be higher than $2 \times 10^{19}$ cm$^{-3}$. Considering that the actual $r_c$ could be higher, $5 \times 10^{19}$ cm$^{-3}$ is a safer base doping concentration in order to guarantee the actual base transfer length not larger than 1 μm. As shown by Figure 3-17, the transfer length starts to increase at very high base doping concentration. This is because the decrease in $r_c$ becomes slower than the decrease in $R_{shb}$ when the base doping becomes very high.

![Graph showing base contact transfer lengths for different base width and base doping concentrations.]

Figure 3-17 The base contact transfer lengths for different base width and base doping concentrations.
3.4 HBT Reliability

3.4.1 Overview

A high level of reliability is key to the use of HBTs in almost any applications. HBT reliability is usually characterized by mean time to failure (MTTF). The degradation of HBT devices is normally monitored by continuously measuring the typical device performance parameters, such as current gain, turn-on voltage and base resistance. Depending on the parameters under monitoring, failure is usually defined as a 20% change in the monitored performance parameter. In order to speed up reliability tests, devices under test are always thermally stressed and/or bias-stressed. By measuring MTTF at different stress temperatures, an activation energy of the degradation process can be extrapolated and used for predicting the MTTF under normal circuit operation conditions.

By the time this dissertation was written, several degradation mechanisms were known to cause HBT reliability problems. These mechanisms are: 1) Be diffusion, 2) base strain, 3) hydrogen redistribution, 4) extrinsic base surface recombination, and 5) self-heating.

Before Be was known to cause HBT reliability problems, MBE had been the primary method of producing AlGaAs/GaAs HBTs with Be as the preferred p-type dopant. However, Be-doped HBTs have a very uncontrollable doping profile, which diffuses rapidly at elevated temperatures or under bias stress and causes current gain decrease and turn-on voltage shift. In order to improve the reliability of Be-doped HBTs, TRW used a combination of reduced substrate temperature and increased As/Ga flux ratio during MBE growth to realize a stable doping profile. TRW HBTs demonstrated an extrapolated MTTF of $1.5 \times 10^8$ hr. at 125°C junction temperature. Considering that TRW's reliability test used a fairly low Be doping concentration of $1 \times 10^{19}$ cm$^{-3}$ in the base and a very low stress current of $6.7 \times 10^3$ A/cm$^2$ during the bias-stressing, TRW’s approach did not provide a universal solution to the reliability problem caused by Be diffusion. Instead of trying to solve the problem, most HBT production lines have switched their starting material to MOCVD grown material which uses carbon-doped base regions. In this dissertation research, all HBT wafers were still MBE grown and Be-doped because MBE was more available for us.
Because the base layer of AlGaAs/GaAs HBTs are always heavily doped to lower the base resistance, the lattice constant in the base layer decreases slightly from high carbon doping. This lattice constant shrinking results in lattice strain and causes reliability problems. With base strain, conventional HBTs show an increase in base current during reliability testing at elevated temperatures and current stress. The ideality factor for the excess base current is considerably larger than 2, indicating that this increase in base current is due to a leakage through certain defects instead of a simple generation-recombination process. This kind of base current increase is followed by a catastrophic reduction in current gain. NTT has tried to reduce the defect generation rate by relaxing the lattice strain occurring in the base by introducing indium to carbon-doped GaAs.\textsuperscript{3-35} Because InAs has a much larger lattice constant than GaAs, incorporating a very small amount of indium into the carbon-doped GaAs releases the base strain. This strain relaxation has been reported to improve the life of carbon-doped base AlGaAs/GaAs HBTs by a factor of approximately 100 times.

After carbon doped, MOCVD wafers had replaced beryllium doped MBE wafers for AlGaAs/GaAs HBTs, hydrogen redistribution was found to cause reliability problems in carbon-doped base HBTs. During the MOCVD growth process, a large amount of hydrogen atoms are incorporated into the HBT epitaxial layers. Because hydrogen atoms are very mobile, current stress under elevated temperatures results in hydrogen redistribution. It is well known that hydrogen can passivate or neutralize p-type dopants in GaAs.\textsuperscript{3-36} As a result, some carbon atoms in the base are passivated by diffusing hydrogen atoms and the base resistance increases. In order to solve this problem, hydrogen incorporation has to be minimized during the epitaxial growth and proton isolation implant should be avoided.

3.4.2 HBT reliability and surface passivation

The current gains and turn-on voltages of AlGaAs/GaAs HBTs have a tendency to drift under bias. Diffusion of the Be base dopant toward the emitter is one source of this drift. And another source of current gain drift is the extrinsic base surface recombination.

Hewlett-Packard Laboratories have experimentally demonstrated that the reliability of carbon-doped HBTs can be improved by using an AlGaAs surface passivation ledge.\textsuperscript{3-37} In the experiment, non-self-aligned HBTs were fabricated on MOCVD grown layers with $4 \times 10^{19}$ cm$^{-3}$ carbon-doped bases. The spacing from the
emitter to the base contact was 2 μm. Approximately 700Å of AlGaAs remained after etching the emitter layer on some devices, forming the surface passivation between the emitter and base, while this region was etched completely down to the base on others. A PECVD nitride layer was deposited after etching. In the device tests, the passivated HBTs with small emitters exhibited higher current gain than unpassivated ones - 18 vs. 12 @ $J_c = 6.5 \times 10^3$ A/cm². (We have discussed how an AlGaAs passivation ledge can increase the current gain in section 3.1.5 and this is what has been expected.) In the reliability test, the devices were stressed under $J_c = 2.6 \times 10^4$ A/cm² and $V_{ce} = 5$ volts with junction temperature at about 200°C. The current gain of the unpassivated C-doped HBTs degraded relatively quickly - current gain @ $J_c = 6.5 \times 10^3$ A/cm² dropped 40% in 10 hours; however, for the passivated carbon-doped HBTs, the current gain under the same current density dropped 4% in the first hour, then stabilized. The improvement in current gain stability demonstrates the importance of AlGaAs passivation ledges on HBT reliability.

Texas Instruments Central Research Laboratories also reported the performance of carbon-doped HBTs bias stressed at elevated temperatures, showing that the MTTF for the passivated HBTs is much longer than for unpassivated devices. [3-38] Figure 3-18 shows $I_c$ versus time for passivated and unpassivated devices at 238°C junction temperature with $V_{ce} = 6$ volts and constant base current.

The current gain degradation mechanism for HBTs with an unpassivated base surface is not well understood yet. It is suspected that the energy released by electron-hole recombination events at or near the exposed surface of the base-emitter junction stimulates the generation of more recombination-centers there, and the positive loop in trap generation causes current gain drift. As discussed in section 3.1.5, the use of an AlGaAs surface passivation ledge can reduce the surface recombination. Thus, this passivation ledge can also slow the trap creation process as well as make the current gain less sensitive to the trap concentration. Thus HBT reliability improves with an AlGaAs passivation ledge around the base-emitter junction.

Because an AlGaAs passivation ledge improves current gain and current gain stability simultaneously by reducing the surface recombination, if HBT devices have a high current gain and do not show a surface dependence, then the reliability of these HBTs should not depend on the surface effect. Therefore, in our theoretical and experimental study of the minimum passivation ledge length required to fully passivate the extrinsic
base, only Gummel plot measurements were used to study the dependence of current gain on passivation ledge length. Reliability tests were not performed on these Be-doped HBTs because Be diffusion kills the device first; however, based on the discussion in the previous section, the minimum passivation ledge length extracted from our Gummel plot measurements should be equally valid for achieving high HBT reliability.

![Graph showing Ic vs. time for passivated and unpassivated devices at 238°C junction temperature, Vce = 6 V, Ib constant.](image)

Fig. 1. Ic vs. time for passivated and unpassivated devices at 238°C junction temperature. Vce = 6 V, Ib constant.

Figure 3-18 Reliability data published by Texas Instruments Central Research Laboratories. [3-38]

3.4.3 HBT reliability and self-heating

For all integrated circuits, keeping device temperatures within reasonable bounds is critical to long-term reliability. In general, GaAs ICs suffer more from all thermal effects than silicon ICs because the thermal conductivity of GaAs is only one third that of silicon. For HBT-ICs, the self-heating problem becomes even worse because the speed and gain of AlGaAs/GaAs HBTs is best at very high current densities and with relatively short and wide emitters, both of which increase the peak device junction temperatures.
Figure 3-19 (a) The simulated average emitter-base junction temperature rise as a function of emitter perimeter area ratio for different emitter sizes; (b) the junction temperature rise as a function of emitter width for different emitter sizes.

Figure 3-19(a) shows the simulated average emitter-base junction temperature rise as a function of emitter perimeter area ratio for different emitter sizes. This simulation
was done using a thermal simulation tool called ThCalc, which calculates the temperature profile of an IC and runs fast enough to allow calculations on a whole chip. \[^{3-39}\] As shown in Figure 3-19(a), for a single 4 \(\mu\)m x 4 \(\mu\)m HBT running under \(J_e = 1 \times 10^5\) A/cm\(^2\) and \(V_{ce} = 5\) volts, the average junction temperature rise can be as high as 233°C. If the activation energy for MTTF is taken as 1.1 eV, \[^{3-40}\] the increase of 233°C in junction temperature leads to a reduction in device lifetime by a factor of \(1.3 \times 10^8\). In practice, a structure called Backside Thermal Via \[^{3-41}\] is always fabricated to reduce the operating temperature of AlGaAs/GaAs HBTs to acceptable levels; otherwise, the lifetime of HBT-ICs would never be long enough for any practical applications.

Figure 3-19(a) also shows that the junction temperature rise becomes smaller if HBTs have a smaller emitter size or a larger emitter perimeter area ratio. Without extrinsic base surface passivation, HBTs with a larger emitter P/A ratio tend to suffer more from surface recombination and this is why a square emitter is generally preferred. However, if the emitter periphery is fully passivated, then a long and narrow emitter shape will improve the heat conducting and reduce junction temperature. Figure 3-19(b) shows the junction temperature rise as a function of emitter width for different emitter sizes. If a 4 \(\mu\)m x 4 \(\mu\)m square emitter is re-shaped to 0.5\(\mu\)m x 32 \(\mu\)m finger, the junction temperature rise is only 71°C. Thus, no Backside Thermal Via structure is needed.

Therefore, fully passivated sub-micron width HBTs have less problem in self-heating and yield better reliability. But in order to fabricate fully passivated submicron width HBTs, three steps of research have to be done: 1) investigate the minimum passivation ledge length required to fully passivate extrinsic base surface; 2) develop processing approaches to implement sub-micron ledges in RF HBTs; 3) develop processing techniques suitable for submicron emitter HBTs.
3.5 Design Tradeoffs and HBT Scalability

3.5.1 Design tradeoffs

There are several important design tradeoffs in AlGaAs/GaAs HBT epitaxial structures. The optimization should always be based upon the needs of the circuit applications.

Breakdown voltage versus $f_t$: a higher breakdown voltage usually requires a thicker collector layer, which increases the collector transit time and thus reduces cut-off frequency, $f_t$. A more lightly doped collector may also increase the breakdown voltage, however the highest achievable collector current density without triggering the Kirk effect is reduced, as a result, the highest achievable $f_t$ turns out to be lower. In practice, breakdown voltages are defined by the circuit operating conditions, and $f_t$ is optimized within the limits by minimizing the base transit time and parasitic components, such as emitter and collector resistances as well as junction capacitances.

Current gain versus $f_{max}$: a higher current gain usually requires a relatively lower base doping concentration, which yields relatively higher base sheet and contact resistances and thus reduces maximum oscillation frequency, $f_{max}$. However, if the material quality is good and the extrinsic base surface is fully passivated, a reasonable current gain, like 50, will not set a low limit on the base doping level. In addition, because the base contact resistance also strongly depends on the processing technology, some optimization can be done by using superior contact techniques and layout design.

$f_{max}$ and $f_t$: a higher $f_{max}$ normally requires a thicker base layer which allows a lower base sheet resistance; however, a thicker base layer increases the base transit time and thus reduces the cutoff frequency, $f_t$. Because $f_{max}$ is also proportional to the square root of $f_t$, the compromise between $f_{max}$ and $f_t$ or even maximizing $f_{max}$ itself is not a simple matter. As discussed in section 3.2.1, a built in base electric field created by Al composition grading can increase the electron transport in the base, thus reducing the base transit time. Thus when a thick base has to be used to lower the base sheet resistance, using a graded thick base will minimize the reduction in $f_t$.

Most parasitic components related to performance optimization not only depend on epitaxial structures but also rely on processing techniques as well as device layouts.
Therefore, in order to optimize device performance, the processing techniques and procedures have to be well known so that all these parasitic components are within an expected range. Moreover, because reduction in these parasitic components would normally bring no disadvantages, but only improvements, a superior processing approach or layout design can benefit the device performance with very few drawbacks. For example, scaling the passivation ledge length down to sub-micron sizes will reduce the base-collector junction capacitance $C_{bc}$, thus resulting in a higher $f_i$ and $f_{max}$ simultaneously. There are no tradeoffs as long as this submicron ledge can fully passivate the device; and the only challenge left is the processing difficulties we need to overcome.

3.5.2 Our final designs

In this dissertation research, three kinds of HBT epitaxial structures were designed for three different research purposes.

The first type of HBT was used for investigating the minimum passivation ledge length required to fully passivate the extrinsic base surface recombination. Table 3-2 shows the basic epitaxial structure of these DC HBTs. Because there were four components in base current - hole back-injected current, base-emitter junction space-charge region recombination current, bulk base recombination current, and base surface recombination current, in order to study the base surface recombination and passivation effect, the design principle was to minimize the other base current components and make the base surface recombination current dominate the total base current. Therefore, in the final design, 1) the Al composition in the emitter was relatively high so that the hole back-injection would be prevented; 2) the growth temperature for the base-emitter junction was 600°C so that good material quality would minimize the density of recombination centers in the base-emitter space-charge region; 3) the base layer was moderately doped at $5 \times 10^{18}$ cm$^{-3}$ level so that there would not be much Be out-diffusion and the bulk base recombination current was small. The base thickness was chosen to be 1000Å. Because these HBTs were to be measured only at low frequency, $n^+$ substrates were chosen and there were no sub-collector layers in the epitaxial structures. The processing procedures and device performance results for these HBTs are discussed in chapter 5.
The second type of HBT was used for demonstrating optimized high frequency performance of HBTs with and without sub-micron passivation ledges. Table 3-3 shows the epitaxial structure of these RF HBTs. For RF HBTs, Semi-Insulating GaAs substrates were always used to minimize the substrate parasitics, and in order to make collector metal contact, a 6000Å thick and $5 \times 10^{18}$ cm$^{-3}$ Si doped sub-collector was grown beneath the collector layer. In order to make the device performance more sensitive to the base-collector junction area, the collector thickness was chosen to be relatively thin - 3000Å. In order to reduce the base transit time, an Al composition grading was used in the 400Å thick base layer. In order to realize a reasonably low sheet resistance in this thin base layer, the Be base doping concentration was increased to $5 \times 10^{19}$ cm$^{-3}$. Because the Be tended to diffuse into the emitter, thus reducing the emitter efficiency, a 200Å undoped spacer was grown between the emitter and base. An InAs contact layer was grown on top of the whole device in order to reduce the emitter contact resistance. The MBE growth conditions have been discussed in section 2.1.3.

<table>
<thead>
<tr>
<th>Layer</th>
<th>Thickness (Å)</th>
<th>Doping (cm$^{-3}$)</th>
<th>Al(In) Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contact</td>
<td>n+ InAs</td>
<td>200</td>
<td>2E+19</td>
</tr>
<tr>
<td></td>
<td>n+ InGaAs</td>
<td>300</td>
<td>1E+19</td>
</tr>
<tr>
<td>Cap</td>
<td>n+ GaAs</td>
<td>1000</td>
<td>2E+18</td>
</tr>
<tr>
<td>Emitter</td>
<td>N GaAlAs</td>
<td>300</td>
<td>5E+17</td>
</tr>
<tr>
<td></td>
<td>N GaAlAs</td>
<td>300</td>
<td>5E+17</td>
</tr>
<tr>
<td></td>
<td>N GaAlAs</td>
<td>329</td>
<td>5E+17</td>
</tr>
<tr>
<td>Base</td>
<td>p+ GaAs</td>
<td>1000</td>
<td>5E+18</td>
</tr>
<tr>
<td>Collector</td>
<td>n- GaAs</td>
<td>10000</td>
<td>2.00E+16</td>
</tr>
<tr>
<td>Substrate</td>
<td>GaAs</td>
<td>n+</td>
<td></td>
</tr>
</tbody>
</table>

Table 3-2 DC HBT structure for ledge passivation study.
### EPI STRUCTURE OF RF HBTs

<table>
<thead>
<tr>
<th>Layer</th>
<th>Thickness (Å)</th>
<th>Doping (cm⁻³)</th>
<th>Al(In) Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contact</td>
<td>n⁺ InAs</td>
<td>200</td>
<td>2E⁺19</td>
</tr>
<tr>
<td></td>
<td>n⁺ InGaAs</td>
<td>300</td>
<td>1E⁺19</td>
</tr>
<tr>
<td>Cap</td>
<td>n⁺ GaAs</td>
<td>1000</td>
<td>2E⁺18</td>
</tr>
<tr>
<td>Emitter</td>
<td>N GaAlAs</td>
<td>300</td>
<td>5E⁺17</td>
</tr>
<tr>
<td></td>
<td>N GaAlAs</td>
<td>300</td>
<td>5E⁺17</td>
</tr>
<tr>
<td></td>
<td>N GaAlAs</td>
<td>269</td>
<td>5E⁺17</td>
</tr>
<tr>
<td>Base</td>
<td>i GaAs</td>
<td>200</td>
<td>0.1 -&gt; .07</td>
</tr>
<tr>
<td></td>
<td>p⁺ GaAs</td>
<td>400</td>
<td>5E⁺19</td>
</tr>
<tr>
<td>Collector</td>
<td>n⁻ GaAs</td>
<td>3000</td>
<td>2E⁺16</td>
</tr>
<tr>
<td>Sub-coll</td>
<td>n⁺ GaAs</td>
<td>6000</td>
<td>5E⁺18</td>
</tr>
<tr>
<td>Substrate</td>
<td>GaAs</td>
<td></td>
<td>Semi-Insulating</td>
</tr>
</tbody>
</table>

Table 3-3 RF HBT structure for device optimization study.

The third type of HBT was used for demonstrating the feasibility of using an Undercut-Emitter-Metal approach to fabricate sub-micron HBTs. These HBTs had a similar epitaxial structure to that shown in Table 3-3 except that the GaAs cap layer thickness was 3000Å instead of 1000Å in order to improve the undercut process.

#### 3.5.3 HBT scalability

To increase the number of components per integrated circuit chip and to reduce the device junction temperature rise due to self-heating, the HBT device dimensions must be scaled down. Commercial HBTs with minimum feature width (e.g. emitter width) of 1.5 to 2 μm are now available, and we expect that this minimum dimension will continue to shrink into the submicron region to allow LSI and VLSI HBT-IC fabrication.

Theoretically, the device scaling includes shrinkage in both vertical and lateral dimensions. A simple approach is to reduce all dimensions by a scaling factor \( \kappa > 1 \), and by using the notations introduced in section 3.2, the new dimensions are

\[
\begin{align*}
\text{Vertical dimension: } & \quad W'_b = \frac{W_b}{\kappa}, \quad W'_{bc} = \frac{W_{bc}}{\kappa}, \\
\text{Width dimension: } & \quad W'_e = \frac{W_e}{\kappa}, \quad L'_{eb} = \frac{L_{eb}}{\kappa}.
\end{align*}
\] (3.43a)
Length dimension: \[ W' = \frac{W}{\kappa}. \quad (3.43c) \]

In order to keep the base sheet resistance constant, we have to increase the base doping level by a factor of \( \kappa \) during the scaling. We assume that bulk base recombination lifetime is inversely proportional to base doping concentration, which is so high that bulk base recombination current will dominate the total base current. We also assume that the collector doping concentration is low and unchanged during the scaling so that the open emitter breakdown voltage will be simply proportional to the collector thickness. Then, the scalability of HBT DC performance can be expressed as follows:

\[ g_m' = \frac{g_m}{\kappa^2}, \quad \beta' = \beta \cdot \kappa, \quad BV_{CBO}' = \frac{BV_{CBO}}{\kappa}. \quad (3.44) \]

The turn-on voltage does not scale due to the constant band-gap of the GaAs base. As for the scalability of HBT high frequency performance, let us assume that all the metal contacts are ohmic and yield negligible contact resistances and that the base-collector space charge region transit time dominates the total time delay. Using equations (3.30), (3.37), (3.41) and (3.42), the scalability of HBT high frequency performance can be expressed as follows:

\[ C_{bc}' = \frac{C_{bc}}{\kappa}, \quad R_b' = R_b \quad (3.45) \]

\[ f_t' = f_t \cdot \kappa, \quad f_{max}' = f_{max} \cdot \kappa \quad (3.46) \]

The scalability of total DC power consumption is

\[ P_{dc}' = \frac{P_{dc}}{\kappa^3}. \quad (3.47) \]

Unfortunately, in practice this theoretical scaling approach is not realistic considering the following facts: 1) terminal contact resistances are not generally negligible. The emitter contact resistance, which is proportional to the square of the scaling factor, will increase dramatically during the scaling. As a result, the base-collector junction charging time will increase proportionally to the scaling factor. Second, the base contact resistance is usually large and does not scale with the total series
base resistance; 2) the base-collector junction capacitance normally does not scale as described by equation (3.45) because the minimum size extrinsic base area is limited by the minimum passivation ledge length and base metal width; 3) when the base-collector transit time decreases and the base-collector junction charging time increases as the device scaling proceeds, the cutoff frequency will eventually be determined by the latter instead of the former and start to decline.

In current HBT structures, the base doping concentration is already high (~ 5 x 10^{19} \text{ cm}^{-3}) and the base thickness is already thin (~ 400\text{Å}), so increasing the base doping level or decreasing base thickness is not very practical. In addition, because the turn-on voltage is a constant, most circuits operate under similar bias conditions. Thus, the breakdown voltage and collector thickness should not be scaled. Fortunately, unlike MOSFET device scaling where the junction depth has to be reduced together with the gate length in order to avoid short channel effects, HBT device scaling in lateral and vertical dimensions can be done independently.

Now, let us consider the HBT scaling in one lateral dimension - width dimension. During this scaling, all the geometric parameters in the vertical dimension and the other lateral dimension, such as \( W_b \), \( W_{bc} \) and \( W \), are not scaled. After this scaling, the HBT will become long and narrow so that the self-heating and high power consumption problems can be solved. Table 3-4 shows the range of the scalable quantities and the fixed values of the non-scalable quantities during the scaling in width dimension. The base metal width has to be larger than base contact transfer length which depends on base sheet and contact resistances. We take 1 \text{μm} as its minimum in this case. The minimum passivation ledge length is about three times the base width (chapter 5) and we choose 0.3 \text{μm} here. Both \( R_b \) and \( R_c \) do not strongly depend on emitter width, so we use the measured values we obtained from our real devices. The transit times and the base-emitter junction charging time are independent of emitter width and the experimentally extracted values are used.

Figure 3-20 shows the high frequency performance of the HBTs scaled in width dimension. Because the extrinsic base area does not scale, the total base-collector junction capacitance does not decrease as much as the emitter contact resistance increases. Thus, the base-collector junction charging time becomes longer and the cutoff frequency drops when emitter sizes are scaled down. As for the maximum oscillation frequency, the decrease in \( f_t \) can be compensated by the decrease in total \( C_{bc} \) at the early
stage of the scaling, so the curve climbs to a peak and then start to drop slowly. A key to solve this high frequency performance degradation problem is to lower the emitter contact resistance during the scaling. An undercut emitter metal structure is proposed and studied in chapter 4 in order to maintain a constant $R_e$ when HBTs are scaled down.

<table>
<thead>
<tr>
<th>scalable dimension</th>
<th>non-scalable or not-scaled dimensions</th>
<th>fixed parasitics</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W_e$ (µm)</td>
<td>$W_{bs}$ (µm) Ledge (µm) Spacing (µm) W (µm)</td>
<td>$R_e$ (Ω) $R_o$ (Ω)</td>
</tr>
<tr>
<td>2.0 -&gt; 0.2</td>
<td>1 0.3 0.3 10</td>
<td>8.8 25.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>scalable intrinsic base</th>
<th>non-scalable extrinsic base</th>
<th>fixed time delay</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 -&gt; 2 µm²</td>
<td>32 µm²</td>
<td>$C_{be}/g_m$</td>
</tr>
<tr>
<td>7.62 -&gt; 0.76 ff</td>
<td>12.18 ff</td>
<td>$l_b$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$l_c$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.64 ps</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.23 ps</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.35 ps</td>
</tr>
<tr>
<td>scalable emitter contact</td>
<td></td>
<td>$(R_e+R_c)-C_{bc}$</td>
</tr>
<tr>
<td>20 -&gt; 2 µm²</td>
<td></td>
<td>0.45 -&gt; 1.91 ps</td>
</tr>
<tr>
<td>13.9 -&gt; 139 Ω</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3-4 Range of scalable quantities and fixed values of the non-scalable quantities during the scaling in HBT width dimension.

Figure 3-20 The $f_i$ and $f_{max}$ as a function of scaling factor defined as the scaled emitter size divided by 2 µm. The device is only scaled in width dimension and the extrinsic base area is assumed to be non-scalable.
Chapter 4
Fabrication Process Development for HBTs with Submicron Features

4.1 Standard HBT Fabrication Processes

4.1.1 Overview

HBT processing is the most difficult among all devices because HBTs have three terminals and vertical current flow. For discrete device fabrication, six lithography levels are normally required for DC HBTs and eight for RF HBTs. While the processing sequence is sometimes customized, the generic processing steps can be listed as follow:

<table>
<thead>
<tr>
<th>Process</th>
<th>RF only</th>
<th>DC only</th>
<th>RF only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isolation implant</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alignment mark etch</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emitter metal formation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passivation ledge formation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base metal formation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base mesa formation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collector metal formation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Second metal VIA definition</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Second metal formation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dark field</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dark field</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bright field</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bright field</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dark field</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dark field</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

These processing steps can be categorized into several processing modules: lithography, metalization, InGaAs/GaAs/AlGaAs etch, dielectric deposition and etch, and isolation implant.

Optical lithography was done using a contact mask aligner, operating at 436 nm and offering a resolution of 1.0 μm. The alignment accuracy was about 1.0 μm, depending on alignment mark design and the operator's skill. Lithography is used in two different ways - the patterned photoresist is used as either an etch mask or a lift-off layer. In the former case, the photoresist layer was thinner and the exposure dose lower; while in the latter case, the photoresist layer was not only thicker but also treated with a chlorobenzene “soak” in order to increase its surface hardness and improve the lift-off process. In both cases, we used Shipley 1813 positive photoresist. When the wafer surface was covered by silicon dioxide or nitride, a Shipley adhesion promoter was used before the photoresist coating, but otherwise the photoresist was spun directly onto the
wafer. A caution to be noted was that the exposure dose was lower when the photoresist
was on top of silicon dioxide or nitride. Under this circumstance, very careful lamp
calibration was done to assure the correct exposure.

The metalization processes included two major steps - metal evaporation and
metal lift-off. In the metal evaporation step, the wafer mounting was the most critical
operation. Without special adjustment of the two-gun electron beam evaporation system,
the metal flux was not perpendicular to the wafer surface. As shown in Figure 4-1(a), the
metal profile after lift-off had a shape like a chair with a high back. In order to solve this
problem, a collimation device was utilized in combination with alignment of the wafer
normal to the metal flux by tilting the wafer mounting. This adjustment yielded very
good metal profiles, as shown in Figure 4-1(b). In the metal lift-off step, the
chlorobenzene treatment followed by an overnight soak in acetone after metal deposition
were most important. The chlorobenzene soak hardens the photoresist surface and results
in an overhang on the wall of trench patterns after development. This overhang helps
break the metal layer and improves the metal lift-off processes. Figure 4-2 shows a SEM
picture of an overhang processed using the chlorobenzene treatment. The overnight
acetone soak was important because it allowed enough time for the acetone to dissolve
and completely lift the whole photoresist layer. Ultrasound could be used to speed up the
lift-off process but it causes metal adhesion problems, especially for small patterns.
Figure 4-1 The SEM pictures of the metal profiles evaporated without and with the collimatic alignment.
In the standard HBT fabrication process, wet etch techniques were used to etch InGaAs, GaAs and AlGaAs layers. The most frequently used etching solution was the phosphoric acid based recipe - 3 : 1 : 50 (H₃PO₄ : H₂O₂ : H₂O). At 20°C, this recipe had an etch rate of 800 Å/min. for GaAs and about 1000 Å/min. for Al₀.₃Ga₀.₇As. Wet etches are generally anisotropic and crystal orientation dependent. This dependence was extremely important for self-aligned processes in order to control the metal discontinuity. For (100) GaAs wafers, when an emitter finger parallel to the wafer major flat was etched, the cross section of the emitter finger has an upside-down terrace shape suitable for self-aligned HBTs. (In the American standard, the major flat is perpendicular to [0, -1, 1] direction.) When GaAs wafers with other orientations were used, the general rule was that the wet etched facets are always the Ga planes with single bonds pointing out. Another important issue with respect to wet etch is surface wetting. Because a photoresist surface is hydrophobic, a wet etchant will not wet the bottom of the trenches and allow the etching when the trenches are small. We used three methods to solve this problem: 1) oxidize the photoresist surface with an oxygen plasma; 2) replace water by iso-propanol alcohol; 3) use a wetting agent, such as Kodak "dry well" solution. The first method did not offer enough improvement and the second one was not applicable to all etching processes. The third method worked quite well, except that the previously calibrated etch rate decreased slightly with time because the wetting agent on the wafer
surface gradually neutralized the etchant. In order to obtain a better etched profile, dry etch techniques were eventually used.

Three kinds of dielectric were used in the HBT fabrication processes: polyimide, silicon dioxide and silicon nitride. Because polyimide could be directly spun on wafers and it planarizes the wafer surface, it was a good candidate for isolation between the device and the second metal layer, which was fabricated for measurement purposes on discrete HBT wafers. However, the specific type of polyimide available had to be cured for 30 minutes at 400°C, which degrades the metal contacts to the device, thus silicon dioxide was used for the second level metal isolation in our HBTs. Oxide was used to replace the polyimide was because it has a lower dielectric constant than the nitride, thus resulting in smaller parasitic capacitance. Silicon dioxide and nitride were deposited by Plasma Enchanced Chemical Vapor Deposition (PECVD) with silane and oxygen and ammonia as reactants, respectively. The wafer temperature was 300°C during the PECVD deposition. Reactive Ion Etch (RIE) was utilized to pattern the silicon oxide or nitride. Because the stoichiometry of silicon oxide or nitride strongly depends on the deposition conditions and varied from one system to another, the RIE etch rate had to be calibrated for every batch of wafers.

Isolation implant was a process step needed for fabricating RF HBTs. Although the epitaxial structures of RF HBTs were grown on top of semi-insulating substrates, the sub-collector layer was highly conductive and an isolation implant was always used to damage the field region in order to avoid the cross-talk between devices and to reduce the substrate parasitics. For RF HBTs, two proton (H+) implantations and six oxygen (O^{++}) implantations were combined to realize the field isolation. While oxygen ions have a larger carrier killing ratio (~ 50 carriers per ion) than protons (~ 1 carrier per proton), the proton implant has a much deeper projected range (~ 1.51 μm at 200 KeV) than oxygen (~ 0.53 μm at 400 KeV). The implant doses for each species were chosen so that the carrier killing capability was 100 - 200% of the doping concentration in all the epitaxial layers. Because no test wafers were processed to optimize the isolation process, the energies and doses listed in the RF HBT processing flow were just the conditions used, which produced a marginally acceptable isolation - a field sheet resistance of 20 KΩ. The isolation mask consisted of a 0.5 μm thick silicon dioxide plus a 1.5μm thick Al layer. The proton projected ranges in silicon dioxide and Al were similar to those in GaAs, so as long as the total mask thickness is larger than 1.5μm, the surface layer underneath the mask is safe.
4.1.2 Formation of AlGaAs Passivation Ledges

Fully depleted AlGaAs passivation ledges can effectively reduce or eliminate the extrinsic base surface recombination, thus producing HBTs with high current gain, low 1/f noise and a good reliability. In order to fabricate these thin AlGaAs passivation ledges, selective etch processes had to be used.

A wet etchant which selectively etches InGaAs and GaAs over AlGaAs is a 200:1 mixture of hydrogen peroxide (H₂O₂) and ammonia hydroxide (NH₄OH). The etch process stops on top of AlGaAs because the Al₂O₃ formed during the oxidation is dense and not soluble in the etching solution thus producing a protection layer. In order to achieve good selectivity, the 200:1 etch is always used at 10°C where the GaAs etch rate is about 800Å/min. and the GaAs/AlGaAs etch selectivity is about 6. The etch rate and etch selectivity are very sensitive to temperature and freshness of the etch. The temperature variation should be less than 1°C and the solution should be used within 4 hours after mixing.

The AlGaAs passivation ledge processing procedure is described as follows: 1) mix 3:1:50 and 200:1 etching solutions, put them in an isothermal bath at 20°C for the former and 10°C for the latter, and wait for about one hour for solutions to equilibrate; 2) the GaAs etch rates were calibrated in both solutions to insure that the etch rates were the expected values; 3) the wafers were etched in 3:1:50 for 90 seconds and then rinsed in De-Ionized-water (DI-water) for 180 seconds; 4) the wafers were etched in 200:1 for 60 seconds and rinsed in DI-water for 180 seconds. The emitter structure consisted of 200Å-InAs / 300Å-InGaAs / 1000Å-GaAs / 929Å-AlGaAs. The emitter was etched for 90 seconds in 3:1:50 etch, which removed 1300Å. This was followed by 60 seconds in 200:1 etch, which selectively stopped at a depth of 1700Å, which is 200Å into the AlGaAs layer because of the Al composition grading. The finished AlGaAs ledge layer thus had a thickness of about 700Å. This could be reduced to ~600Å if the 200:1 etch time was increased to 120 seconds. The finished AlGaAs surface always exhibited a matte dark gray color due to the Al₂O₃ on the surface.

Although this selective wet etch process was very reliable in producing AlGaAs passivation ledges, it undercut the emitter metal heavily. During the 200:1 selective etch, the vertical etch does not proceed beyond the Al₂O₃ protection layer, however, the
exposed lateral edges continue to etch. When the emitter metal sizes were small, this uncontrollable undercut causes the emitter metal to peel. In order to fabricate small emitter size HBTs, dry etch techniques are needed.

4.1.3 Standard DC HBT fabrication processes

It was relatively easier to fabricate DC HBTs than RF HBTs, so whenever high frequency performance was not of interest for a specific research goal, DC HBTs were designed and fabricated to test the relevant characteristics. For example, in our investigation of minimum passivation ledge lengths, we chose current gain as the measure of passivation effectiveness, thus DC HBTs with different passivation ledge lengths were most suitable for this research task.

Figure 4-3 shows a schematic processing flow for DC HBTs and a detailed narrative version of this processing flow can be found in Appendix A. Each processing step is numbered according to the mask level involved.

4.1.4 Standard self-aligned RF HBT fabrication processes

HBT-ICs must operate at very high frequencies in real applications. In order to study and optimize the high frequency performance of HBT devices, RF HBTs have to be fabricated by the same process as used in HBT-ICs and they must be characterized at high frequencies. In our case, we wanted to compare the high frequency performance of HBTs with and without submicron passivation ledges, so two types of RF HBTs were fabricated - type I fabricated using the standard self-aligned fabrication process without ledges, and type II, using the newly developed processing approach which formed a submicron ledge around the emitter mesa. This section describes the standard approach and while the newly developed approach is discussed in section 4.4.
Figure 4-3 A schematic diagram of the processing flow for fabricating DC HBTs. 1a) emitter metal litho & Ti/Pt/Au evaporation; 1b) emitter metal lift-off & etch down to the passivation layer; 2a) passivation ledge litho; 2b) etch down to the base layer; 3a) base metal litho and Ti/Pt/Au evaporation; 3b) base metal lift-off; 4a) base mesa litho; 4b) wet etch down to collector layer and backside collector metalization; 5) oxide deposition, VIA hole litho and etch; 6) interconnect metal litho, evaporation and lift-off.
Figure 4-4 A schematic diagram of the processing flow for fabricating self-aligned RF HBTs. 1) isolation implant; 2a) emitter metal formation and etch down to the ledge layer; 2b) nitride sidewall and etch down to the base; 3) self-aligned base metal litho and evaporation; 4) base metal lift-off and base mesa litho; 5a) collector metal litho; 5b) etch down to sub-collector and collector metal evaporation; 6) collector metal lift-off and dielectric deposition; VIA litho and VIA etch; 6) interconnect metal litho, evaporation and lift-off.
Figure 4-4 shows a schematic diagram of the processing flow of the self-aligned RF HBT fabrication processes. A detailed narrative version of this processing flow can be found in Appendix B. Each processing step is numbered according to the mask level involved except that the step 2 is skipped here because it had no effect on devices. Basically, mask level 2 opened a window at the alignment mark area so that the 3:1:50 would etch into the wafer and create alignment marks for future processes.

4.1.5 Limitations of standard DC and RF HBT fabrication processes

In order to investigate the minimum passivation ledge lengths, submicron passivation ledges had to be fabricated with very precise dimensions. The standard DC HBT fabrication processes described in section 4.1.3 did not meet this requirement due to the poor critical dimension control in emitter width and insufficient overlay accuracy of the contact mask aligner. If an advanced optical lithography tool, such as an i-line stepper had been available, it might have been possible to fabricate passivation ledges as small as 0.3 μm. In this dissertation research, electron-beam (e-beam) lithography was chosen to overcome this limitation.

The standard self-aligned RF HBT fabrication processes had several disadvantages. First, there were no passivation ledges surrounding the emitter mesa, thus the device performance and reliability suffered from extrinsic base surface recombination. Second, the self-aligned processes were not particularly suitable for manufacturing because the yield was not sufficiently high. Third, the self-aligned lift-off did not allow a thick base metal layer because it used the emitter mesa to lift the base metal. In order to overcome these disadvantages, three processing approaches for fabricating submicron ledges were developed in this dissertation research. Section 4.4 describes these three approaches as well as another that was developed earlier by NTT.

Both the standard DC and RF fabrication processes utilized wet etching to form the emitter mesa, so neither of them was suitable for fabricating submicron emitter size HBTs. Because wet etching always undercuts the mask, precise pattern transfer could not be achieved. When the lateral dimensions are comparable to the vertical dimensions, wet etching causes the metal to peel off. As the leading alternative method, a combination of dry-wet etching offers many advantages over the wet-etch only in terms of emitter size control, passivation ledge formation and GaAs/AlGaAs etch selectivity. Section 4.3 discusses the dry-etch techniques developed for AlGaAs/GaAs HBT fabrication.
4.2 Electron Beam Lithography

4.2.1 Electron beam lithography system

Electron beam lithography (e-beam lithography) offers higher resolution than optical lithography because of the small wavelength of the 10-50 KeV electrons. The system design of e-beam lithography machines also offers much better overlay accuracy than optical lithography systems. A Hitachi HL-700F e-beam lithography system was used in this research. This system uses a field emission type electron gun and could draw ultra-fine patterns with a minimum width of 0.1 μm. Under optimal conditions with alignment mark detection and high accelerating voltage, this system had an overlay accuracy as fine as 0.025 μm (1σ).

The HL-700F uses a vector-scan within each exposure field and exposes in step-repeat fashion from one field to another. In a vector-scan pattern generator, the beam is directed sequentially to the parts of the chip pattern to be exposed. The alignment marks in the field corners are first scanned to set the deflection amplitude, offset, rotation, and x-y scan axis orthogonality. Then the pattern is decomposed into a number of elements and each is filled in by the writing beam. After all elements in the 2 mm x 2 mm field are exposed, the stage is stepped to bring the next field under the beam and the exposure process is repeated.

The resolution of e-beam lithography is not limited by diffraction, but by electron scattering in the resist and by the various aberrations of the electron optics. When the electron beam current is high and the required exposure dose is low, the minimum feature size and the critical dimension (CD) accuracy is also limited by the scan spacing. Once the beam current is set, exposure is controlled by varying the scan spacing, S, and the scan dwell time, T. If the required dose is equal to the resist sensitivity, Q, then the scan spacing can be calculated by

\[ S = \sqrt{\frac{T \cdot T}{Q}} \]  \hspace{1cm} (4.1)

For example, when I = 20 nA and Q = 8 μC/cm², because the minimum scan dwell time is 10 ns, the scan spacing has to be 0.05 μm or larger. When the scan spacing becomes comparable to the feature sizes, the CD accuracy will be poor. In order to reduce the scan
spacing, smaller beam current and higher dose resists, such as PMMA, are used. While the designed resolution of HL-700F is 0.1 μm, features as small as 0.06 μm have been attempted and yielded acceptable results.

The accuracy of the overlay depends on many variables, such as alignment mark condition, mark detection specifications, and stability of field distortion from one level to the next. Marks may be etched trenches in GaAs, or they may be metal layers or other features that give sufficient contrast. A mark position is sensed by the change produced in the number of back-scattered and secondary electrons on scanning over the mark edges. For the HL-700F, due to poor vendor support, design of alignment marks and setup of appropriate mark detection specifications were unknown. We thus designed several alignment marks and tested different mark detection specifications; the test procedure and results are discussed in the next section.
Figure 4-5 The basic configuration of the controlling software and file systems in the HL-700F e-beam lithography system.

Unlike optical lithography, e-beam lithography does not use a mask. Layout and exposure information stored in data files directly control the e-beam writing. Figure 4-5 shows the basic configuration of the software and files in HL-700F. After the layout
design is finished on DW-2000 or any other CAD tool, the whole library is output to a tape and copied into the HL-700F’s stream data directory, or the whole library is output to a file which is then transferred into the HL-700F by ftp. Inside the stream data directory, software called ELSA converts the layout file into Hitachi language data files, which are stored in a separate directory. Inside the user directory, the user compiles a JTX file which specifies how layout patterns are repeated over the wafer, what type of alignment marks are used and where they are located. A JTX file (such as HBT_LEVEL2.JTX) can include more than one layout file. For example, when the die size is 4 mm x 4 mm but the e-beam field size is only 2 mm x 2 mm, this die can be divided into four squares. The layout of each square is converted into a Hitachi language file independently and the JTX file will then call all four layout files and repeat them alternately so that the pattern on the wafer will have a periodicity of 4 mm x 4 mm. Finally, before running the JTX file for the e-beam writing, the user needs to check the beam current and to specify the dose. These standard procedures can be executed separately by using command lines or together by running a CMD file.

4.2.2 Overlay accuracy

In order to achieve fine alignment, the HL-700F requires two kinds of alignment marks - wafer alignment marks (WMs) and chip alignment marks (CMs). The WMarks are placed on the two sides of a wafer while the CMarks are placed on the four corners of each field (2 mm x 2 mm). Before the wafer is loaded into the HL-700F, the wafer is mounted on a special loading track, which allows the user to rotate the wafer under a microscope so that the WMarks on both sides are at the same y-level. When the HL-700F starts direct writing, the stage will move to the position of the WMarks and the detecting beam will search for the WMarks. Once the WMarks are found, the system will calculate the actual positions of these WMarks based on the reflection signal coming back from the WM edges. Then, the HL-700F uses the actual positions of the WMarks as the reference in following CM search process. During the exposure, every time when the stage moves to bring a new field under the beam, the system will search and scan the four CMarks on the corners of this chip. Then, the deflection amplitude, offset, rotation and x-y scan axis orthogonality will be calculated and used for correcting the writing of this chip.

If the pre-alignment of WMarks under the microscope before loading is not accurate enough or if the WM search specification is set improperly, then the HL-700F may not be able to find the WMarks and the e-beam writing will be aborted. Once WMarks are found, the
e-beam writing will continue, independent of whether or not all of the CMs are found. The overlay accuracy is completely determined by the CM shape and the CM searching and scanning specification.

Figure 4-6 The two-level vernier patterns.

Alignment marks may be etched trenches in GaAs, or they may be metal layers or other features that give sufficient contrast. Our alignment marks were metal crosses made by lift-off processing. The metal lines were 2 μm x 2000 μm for the WMs and 2 μm x 40 μm for the CMs. The advantages of using Ti/Pt/Au metal layers were that they offered high reflection signal contrast and that they were compatible with our processing sequence because our first lithography process was the emitter metal formation. The disadvantage was that the metal line shape could not be precisely controlled by our lift-off process. This was probably the reason that the best overlay accuracy was 0.05 μm
(1σ) instead of the nominal value, 0.025 μm, even after optimizing the mark search and scan specifications.

Figure 4-6 shows the two-level vernier patterns designed for testing the overlay accuracy. Both levels were formed by metal lift-off. The first level patterns were written at the same time when the WM and CMs were defined on the wafer by e-beam lithography. During the writing of the second level patterns, the WM and CMs were used for the e-beam alignment. By measuring the mis-alignments between the 1st and 2nd level patterns across the whole wafer, the overlay accuracy was tested and studied statistically.

<table>
<thead>
<tr>
<th>Overlay Accuracy of HL-700 E-Beam Lithography System</th>
</tr>
</thead>
<tbody>
<tr>
<td>wafer and chip mark search and scan specifications</td>
</tr>
<tr>
<td>condition set #1</td>
</tr>
<tr>
<td>condition set #2</td>
</tr>
<tr>
<td>condition set #3</td>
</tr>
<tr>
<td>condition set #4</td>
</tr>
</tbody>
</table>

Table 4-1 The overlay accuracy of the HL-700F e-beam lithography system.

<table>
<thead>
<tr>
<th>mark edge count</th>
<th>Wafer Mark</th>
<th>Chip Mark</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGC gain factor</td>
<td>Search</td>
<td>Detect</td>
</tr>
<tr>
<td>beam size</td>
<td>1 x 1</td>
<td>1 x 1</td>
</tr>
<tr>
<td>scanning width</td>
<td>μm</td>
<td>500</td>
</tr>
<tr>
<td>scanning starting point</td>
<td>μm</td>
<td>40</td>
</tr>
<tr>
<td>scanning pitch between lines</td>
<td>μm</td>
<td>1.0</td>
</tr>
<tr>
<td>scanning shot pitch</td>
<td>μm</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Table 4-2 The main WM and CM search and scan specifications under the condition set #3.

Table 4-1 summarizes the overlay accuracy of the e-beam alignment done by using four different sets of mark search and scan specifications. For each set of mark search and scan specifications, 100 vernier patterns in 25 dies were measured by using
Scanning Electron Microscope (SEM). A statistics tool called SAS was used to analyze the data. As shown by Table 4-1, the best overlay accuracy was about 0.05 μm in both x and y directions and the average x and y mis-alignments had a value of 0.01 μm due to the asymmetry inside HL-700F. Considering that the smallest ledge dimension we would test for our HBTs was 0.1 μm, this overlay accuracy was adequate. Table 4-2 lists the main WM and CM search and scan specifications in the condition set #3 used to realize the best overlay accuracy.

4.2.3 E-Beam resist and related processing

Two types of e-beam resists had been used in this dissertation research - PMMA and SAL601. PMMA is a positive resist where electron exposure causes bond breaking while SAL601 is a negative resist where electron exposure causes the formation of bonds or cross links between polymer chains. SAL601 is also useful as an image reversal resist because the complete formation of cross links in SAL601 relies on proper post-exposure baking.

PMMA was used as a positive resist for the emitter and base metal lift-off in our HBT processing. The wafer was first baked in an oven at 150°C for 30 minutes, then the 496K-5% PMMA was spun on the wafer at 4 krpm for 30 seconds. After spinning, the wafer was baked at 110°C for 24 hours to remove moisture and solvent inside the PMMA. The thickness of the PMMA was 0.26 - 0.3 μm, depending on the freshness of the PMMA. The e-beam exposure dose was 275 μC/cm², independent of the PMMA thickness. After exposure, the wafer was developed in a mixture of Methyl Iso-Butyl Ketone (MIBK) and iso-propanol. The mixing ratio of this developer was one part MIBK with three parts iso-propanol. Developing was 40 seconds at room temperature followed by a 2 minute rinse in iso-propanol. This 2600Å thick PMMA was used to lift-off a 1600Å thick metal layer and very good metal shape was obtained.

Although PMMA is a resist that offers the highest known resolution, the high exposure dose requires a relatively long e-beam writing time. Therefore, as a positive resist, PMMA can hardly be used for bright field masking, which means that the field area without device structures is not covered by the resist after developing. For example, in the passivation ledge formation process step, only the emitter and its perimeter are protected by the resist so that the etching solution will etch the rest of the wafer down to the base layer. As a result, the whole wafer has to be exposed by e-beam except the
emitter area. If the beam current is 20 nA and the dose is 275 \( \mu \text{C/cm}^2 \), it would take about 76 hours to write a 2-inch wafer (3.8 hr. / cm\(^2\)). For bright field masking, a negative resist is usually used so that only the patterns which stay after developing are exposed.

SAL601 is a negative e-beam resist which requires an exposure dose of 8 \( \mu \text{C/cm}^2 \). The wafer was first baked in an oven at 150°C for 30 minutes, then the SAL601 was spun on the wafer at 4 krpm for 30 seconds. After spinning, the wafer was baked in an oven at 80°C for 30 minutes. Then the wafer was patterned while the HL-700F with an exposure dose of 8 \( \mu \text{C/cm}^2 \). After exposure and before developing, the wafer was baked on a hot-plate at 110°C for 75 seconds. Then the wafer was developed in Shipley Developer Concentrated (no dilution) for 3 minutes followed by DI-water rinse. The hot-plate baking temperature and time have to be well controlled because they affect the critical dimensions after developing.

While most positive resists are very soluble in acetone, negative resists such as SAL601 are not because of the cross links. In order to remove SAL601 after etching down to the base, an oxygen plasma can be used with the following conditions: chamber pressure 1 torr, RF power 250 W, \( \text{O}_2 \) flow rate 20 sccm and etch time 4 minutes. Micropost Stripper 1165 was also tested for removal of SAL601; however the results showed that it is not suitable as it attacks the Titanium metalization as well as etching GaAs. Thus, an oxygen plasma was used to clean the resist in our HBT processing.
4.3 Dry Etch Processing

Wet etch processing has several disadvantages: 1) it always undercuts the mask, thus precise pattern transfer is difficult to achieve; 2) the etch rate of certain material in an etchant solution can be affected by mask pattern sizes. This local variation causes uniformity problems; 3) liquid phase surface tension and mass transport tend to make small trench etching difficult and uneven; 4) etchant solutions lose potency in use, thus causing the etch rate to change over time.

While wet etching involves using etchant solutions, dry-etching is a general term that applies to any gaseous ambient etching technique. The fabrication of AlGaAs/GaAs hetero-junction bipolar transistors and circuits involves etching small features in thin films and epitaxial layers. The critical dimension control, uniformity, and reproducibility are extremely important. As the leading alternative etching method, dry etching offers many advantages over wet etching in terms of emitter size control, passivation ledge formation and GaAs/AlGaAs etch selectivity. These parameters are not just important, but absolutely necessary for fabricating sub-micron HBTs.

The discussion in this section will concentrate on the plasma etching processing developed to fabricate AlGaAs/GaAs HBTs with sub-micron features. Section 4.3.1 introduces plasma etching, including what is a plasma, what basic mechanisms cause etching, and what are the process requirements. Section 4.3.2 discusses GaAs and AlGaAs plasma etching techniques and recipes. Section 4.3.3 describes the difficulties of dry etching InGaAs and the solution developed in a design of experiment (DOE) study on this matter. Section 4.3.4 lists the recipes used for dry-etching silicon dioxide, silicon nitride, polyimide and metal tungsten.

4.3.1 Overview

As shown in Figure 4-7, the plasma reactor consisted of opposed parallel plate electrodes in a chamber that can be maintained at low pressures, typically ranging from 0.01 to 1 Torr. When a high frequency voltage is applied between the electrodes, current flows forming a plasma, which emits a characteristic glow. Reactive radicals are generated by this electrical discharge. Wafers on the electrode surfaces are exposed to reactive neutral and charged species. Some of these species combine with the film or epi-layer and form volatile products that evaporate, thus etching the film or epi-layer.
As shown in Figure 4-7, in our etching chamber, the top electrode and the whole chamber wall are grounded, serving as the anode while the bottom electrode is connected to the RF supply serving as the cathode. When the cathode is at negative potential, a region adjacent to the cathode is called the cathode sheath or the *dark space*. The high mobility of light electrons and the repulsive field cause the electron population in this region to be so depleted that few electronic excitations occur and a glow is therefore not observed.

When a RF signal in the mega-Hertz range is applied to the cathode through an impedance matching network, the responses of the plasma to the positive and negative cycles are different. When the cathode is positive, many highly mobile electrons are accelerated toward the cathode, causing a significant accumulation of negative charge. When the cathode is negative, only heavy, immobile ions are accelerated toward it; however, significantly fewer of these ions strike the cathode than did electrons on the positive cycle. The plasma therefore acts electrically as a diode with a net negative charge building on the cathode. Although this negative charge accumulation could occur...
at both electrodes, only the cathode can keep this negative charge because there is a large blocking capacitor between the cathode and the RF supply while the anode is simply grounded. As a result, a self-established negative DC bias is created on the cathode. Under this negative DC bias, positive ions are accelerated through the sheath toward the wafer and transfer large amounts of energy and momentum to the wafer surface, thus causing the surface material to be ejected, i.e., sputter etching.

Plasma etching can proceed by four basic mechanisms: [4-1] (1) Sputtering: the ion energy mechanically ejects film material, (2) Chemical etching: thermalized neutral radicals chemically combine with film material forming volatile products, (3) Ion-enhanced chemical etching: neutral radicals reaching the surface alone cannot form a volatile product; energetic ions alter the film or product layer so that chemical reactions can occur which gasify the material; since the ion flux mainly bombards surfaces oriented perpendicular to the electric field, lithographically defined feature sidewalls receive minimal ion flux and the reaction is highly directional, (4) Ion-enhanced protective etching: an inhibitor layer coats the surface, forming a protective barrier which excludes the neutral etchant; moderate ion flux disrupts this protective layer and allows chemical etching to proceed anisotropically, in the vertical direction. Although many mechanisms play a role even in the simplest examples of plasma etching, one or two mechanisms usually dominate the process. Etching variables can be tuned to favor some mechanisms so that a desired etching profile can be obtained.

A plasma etching process is usually characterized according to several output quantities: etch rate, etch selectivity, anisotropy, surface roughness and loading effect. Etch selectivity is the ratio of etching rates between two different materials immersed in the same plasma, for example GaAs and AlGaAs. Anisotropy refers to preferential erosion in a direction normal to the surface versus laterally on a wafer. Surface roughness represents the change in surface quality during the etching. Loading effect is the depletion of etchant species from the gas phase, caused by reactant consumption in the etching process.

The instrumental variables, RF power, RF excitation frequency, pressure, temperature, and feed chemistry are normally used to manipulate plasma processing. For our dry etching system, the RF excitation frequency is 13.56 MHz and the chamber is at the room temperature. The input variables used in our design of experiments (DOE) include RF power, pressure, source gas flow rates and additive flow rates.
4.3.2 Dry etching of GaAs and AlGaAs

A key for the dry successful etching of GaAs and AlGaAs is that the etch products be volatile. Table 4-3 lists the volatility of possible etch products when halogen source gases are used to etch GaAs, AlGaAs or InGaAs. Because neither AlF₃, GaF₃ and InF₃ are volatile, chlorine or bromine atom based plasmas must be utilized. These choices include CCl₂F₂ (Freon 12), Cl₂, CCl₄, SiCl₄, BCl₃, HBr, Br₂, etc.. Because freon 12 and CCl₄ contain carbon atoms which tend to form polymers, they are not suitable for GaAs and AlGaAs etching. Freons are normally used together with O₂ so that the polymers formed during the etch are simultaneously removed by the O₂ plasma; however, O₂ is not allowed in GaAs and AlGaAs etch because of its reaction to form surface oxides. In addition, freon 12 has been banned because it generates ozone holes. Because the bromine based source gases are more corrosive and expensive and bromide etch products are less volatile than chloride etch products, the chlorine based source gases offer more advantages. Among the choices left, SiCl₄ was chosen to pursue processing development for fabricating HBTs with sub-micron features.

<table>
<thead>
<tr>
<th>Etch product</th>
<th>Volatile</th>
<th>Boil or subl'm T</th>
<th>ΔH° @25°C (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AlBr₃</td>
<td>yes</td>
<td>263°C</td>
<td></td>
</tr>
<tr>
<td>AlCl₃</td>
<td>yes</td>
<td>178°C</td>
<td>-168.30</td>
</tr>
<tr>
<td>AlF₃</td>
<td>no</td>
<td>1291°C</td>
<td>-359.50</td>
</tr>
<tr>
<td>GaBr₃</td>
<td>yes</td>
<td>279°C</td>
<td></td>
</tr>
<tr>
<td>GaCl₃</td>
<td>yes</td>
<td>201°C</td>
<td>-125.40</td>
</tr>
<tr>
<td>GaF₃</td>
<td>no</td>
<td>800°C</td>
<td>GaF: -60.20</td>
</tr>
<tr>
<td>InBr₃</td>
<td>?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>InCl₃</td>
<td>?</td>
<td>600°C</td>
<td>-128.40</td>
</tr>
<tr>
<td>InF₃</td>
<td>no</td>
<td>&gt;1200°C</td>
<td>InF: -48.61</td>
</tr>
<tr>
<td>SiCl₄</td>
<td></td>
<td></td>
<td>-157.03</td>
</tr>
<tr>
<td>SiF₄</td>
<td></td>
<td></td>
<td>-358.98</td>
</tr>
</tbody>
</table>

Table 4-3 Volatility and forming energy of possible etch products. [4-2]
One advantage of SiCl₄ over chlorinated halo carbon is that it will not produce chlorinated films which can contaminate the sample, the wafer platen and the etching chamber. Furthermore, although SiCl₄ is a liquid at 760 Torr and 300 K, its vapor pressure is high enough that it can be delivered to the etching chamber using a normal Mass Flow Controller (MFC). However, if the liquid SiCl₄ source stays at the same temperature as the source gas feeding lines and MFC, SiCl₄ tends to condense inside the lines and MFC, thus causing corrosion. Instead of heating up the source gas feeding lines and the MFC in our set-up, we cooled the SiCl₄ quartz container to 15°C so that the source gas could be delivered at room temperature without condensation.

Etching of GaAs and AlGaAs in a moderate SiCl₄ plasma (~50 m-Torr) is generally chemical, and the etching process involves a series of sequential steps: 1) etchant formation - SiCl₄ \( \rightarrow \) SiCl₃⁺ + Cl + e⁻; 2) chemical reaction to form products - GaAs + nCl \( \rightarrow \) GaClₓ + AsClₙ; 3) product removal, either by evaporation or ion-bombardment - GaClₓ(ads) \( \rightarrow \) GaClₓ(gas), where the slowest step, usually removal of the group III chloride products, limits the etching rate. In order to maintain the anisotropy of the etching process and to speed up the removal of the chloride products, relatively lower chamber pressures and He additive were used in our test and final etching recipes.

In order to study the dependence of etch rates on the processing conditions and to optimize the dry etch of GaAs and AlGaAs, a design of experiment (DOE) was performed. In this DOE test, the RF Power, Chamber Pressure and Flow Rates of SiCl₄, SF₆ and He were taken as input variables; the self-established DC Bias and GaAs and AlGaAs etch rates were taken as output variables; and a linear model was used to analyze the dependence of the etch rates on these input variables. Table 4-4 lists the values of input variables used in this DOE test and the results of output variables. Figure 4-8 shows the dependence of the DC bias and GaAs etch rate on input variables.
<table>
<thead>
<tr>
<th>DOE Trial number</th>
<th>RF Power (Watt)</th>
<th>RF Power (m-Torr)</th>
<th>SiCl₄ Flow Rate (sccm)</th>
<th>SF₆ Flow Rate (sccm)</th>
<th>He Flow Rate (sccm)</th>
<th>GaAs Etch Rate (Å/min.)</th>
<th>AlGaAs Etch Rate (Å/min.)</th>
<th>Anode DC Bias (volt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>80</td>
<td>37</td>
<td>22</td>
<td>11</td>
<td>130</td>
<td>2931</td>
<td>0</td>
<td>300</td>
</tr>
<tr>
<td>8</td>
<td>60</td>
<td>27</td>
<td>22</td>
<td>11</td>
<td>90</td>
<td>1815</td>
<td>0</td>
<td>241</td>
</tr>
<tr>
<td>1</td>
<td>60</td>
<td>37</td>
<td>44</td>
<td>11</td>
<td>130</td>
<td>2121</td>
<td>0</td>
<td>211</td>
</tr>
<tr>
<td>5</td>
<td>80</td>
<td>37</td>
<td>22</td>
<td>11</td>
<td>130</td>
<td>2735</td>
<td>0</td>
<td>296</td>
</tr>
<tr>
<td>9</td>
<td>80</td>
<td>37</td>
<td>44</td>
<td>11</td>
<td>90</td>
<td>3419</td>
<td>0</td>
<td>236</td>
</tr>
<tr>
<td>10</td>
<td>60</td>
<td>27</td>
<td>22</td>
<td>0</td>
<td>130</td>
<td>1041</td>
<td>1172</td>
<td>375</td>
</tr>
<tr>
<td>6</td>
<td>60</td>
<td>27</td>
<td>44</td>
<td>0</td>
<td>90</td>
<td>1323</td>
<td>1415</td>
<td>277</td>
</tr>
<tr>
<td>7</td>
<td>80</td>
<td>37</td>
<td>44</td>
<td>0</td>
<td>130</td>
<td>1465</td>
<td>1578</td>
<td>330</td>
</tr>
<tr>
<td>1</td>
<td>60</td>
<td>37</td>
<td>44</td>
<td>11</td>
<td>130</td>
<td>2735</td>
<td>0</td>
<td>211</td>
</tr>
<tr>
<td>4</td>
<td>60</td>
<td>37</td>
<td>22</td>
<td>0</td>
<td>90</td>
<td>1243</td>
<td>1415</td>
<td>299</td>
</tr>
<tr>
<td>3</td>
<td>80</td>
<td>27</td>
<td>44</td>
<td>11</td>
<td>130</td>
<td>1954</td>
<td>0</td>
<td>317</td>
</tr>
<tr>
<td>2</td>
<td>80</td>
<td>27</td>
<td>22</td>
<td>0</td>
<td>90</td>
<td>1548</td>
<td>1709</td>
<td>420</td>
</tr>
<tr>
<td>2</td>
<td>80</td>
<td>27</td>
<td>22</td>
<td>0</td>
<td>90</td>
<td>2051</td>
<td>2051</td>
<td>422</td>
</tr>
<tr>
<td>4</td>
<td>60</td>
<td>37</td>
<td>22</td>
<td>0</td>
<td>90</td>
<td>1641</td>
<td>1578</td>
<td>311</td>
</tr>
<tr>
<td>3</td>
<td>80</td>
<td>27</td>
<td>44</td>
<td>11</td>
<td>130</td>
<td>2564</td>
<td>0</td>
<td>324</td>
</tr>
</tbody>
</table>

Table 4-4 The values of the input and output variables in the DOE test performed for optimizing SiCl₄-based GaAs and AlGaAs dry etch.

First, examining the DC Bias, because it is established by the asymmetry between the cathode and anode, it becomes an output variable instead of a controllable input parameter. (With proper feedback design, the DC Bias may become controllable provided that the RF Power is not fixed.) As shown in Figure 4-8, the DC Bias increases when the RF Power increases or the Chamber Pressure decreases. When the Flow Rate of SiCl₄ or SF₆ was increased, the DC Bias decreased. However, increasing the He Flow Rate had an opposite effect - the DC Bias was higher with a larger He flow; this is the reason that a large He flow rate was used in order to maintain the etch anisotropy.
Figure 4-8  The dependence of the DC Bias and GaAs etch rate on input variables.  a) the effects of RF Power and Pressure (33 sccm SiCl₄, 5.5 sccm SF₆ and 110 sccm He); b) the effects of SiCl₄ and SF₆ (70 watts RF Power, 32 m-torr Chamber Pressure, 110 sccm He); the effects of He and SF₆ (70 watts RF Power, 32 m-torr Chamber Pressure, 33 sccm SiCl₄).
The GaAs etch rate had a positive dependence on the RF Power because more reactant was generated during the plasma discharge under a larger RF Power. When the chamber pressure was lower, the GaAs etch rate decreased although the DC Bias was higher. This dependence suggests that the GaAs etching mechanism is basically chemical etching - the etch rate strongly depends on the partial pressures of the etching species. Additional evidence of this etching mechanism is that an increase in He Flow Rate, which reduced the partial pressure of SiCl₄, resulted in a drop in the GaAs etch rate in spite of a higher DC Bias. While a small GaAs etch rate increase with larger SiCl₄ Flow Rate was expected, the observed dramatic increase in the GaAs etch rate with SF₆ Flow Rate was not expected because GaF₃ is not highly volatile. This is probably due to the low forming energy of GaF and the high forming energy of SiF₄. The presence of SF₆ tended to free more Cl by the reaction SiCl₄ + F → SiCl₃F + Cl, and as a result, increase the available reactant.

The presence of SF₆ can not only increase the GaAs etch rate, but also enhance the etch selectivity of GaAs over AlGaAs. Without SF₆, AlGaAs was etched by the SiCl₄ plasma at a rate similar to the GaAs etch rate. Because AlCl₃ is slightly more volatile than GaCl₃, the AlGaAs etch rate was about 10% higher than the GaAs etch rate. However, when SF₆ was added into the chamber, the AlGaAs etch was completely stopped while the GaAs etch rate increased dramatically. The opposite effects of SF₆ on the GaAs and AlGaAs etch resulted in excellent etch selectivity. This etch selectivity can be used for fabricating sub-micron HBTs with AlGaAs passivation ledges.

This etch selectivity was due to the formation of AlF₃. Because the formation energy of AlF₃ is much larger than that of AlCl₃, when both SiCl₄ and SF₆ are present in the chamber, a layer of non-volatile AlF₃ is formed first, thus stopping the AlGaAs etch completely. Because the formation energy of AlF₃ is large, this etch stop layer can sustain the bombardment of moderate energy ions. Although AlF₃ is not volatile, it is very soluble in water. Therefore, this etch stop layer can be easily removed by a DI-water rinse and does not affect the next processing step.

It is worthy to point out that in order to etch AlGaAs successfully, the chamber base pressure had to be low (~ 2 x 10⁻⁶ torr) because any residual SF₆ or O₂ may form non-volatile products, such as AlF₃ or Al₂O₃ with Al so that the AlGaAs etch can not proceed.
4.3.3 Dry etching of InGaAs

For Npn AlGaAs/GaAs HBTs, a thin layer of InGaAs or InAs is usually grown on top of the emitter cap layer to lower the contact resistance of the emitter metal contacts. As shown in Table 4-3, none of the Indium products is volatile, so this InGaAs or InAs contact layer makes the dry etch process of HBT devices very difficult. One possible solution is to utilize a wet etch process to first remove the thin InGaAs or InAs cap layer before the dry etch process. Unfortunately, this solution has two disadvantages. First, our experiments have shown that the wet etch process is very non-uniform and tends to undercut the emitter metal so that the HBTs would suffer from high emitter series resistance problems. Second, separate etch steps in the emitter mesa formation process requires a much longer turn around time, which hinders productivity.

In order to develop dry etch process conditions suitable for removing the thin InGaAs or InAs layer, a DOE test run was performed together with the DOE test described in the last section. Table 4-5 lists the values of input variables used in the DOE test run for InGaAs etch and the results of output variables. For InGaAs etch, the most important output parameter is the etch rate. Because neither InF$_3$ nor InCl$_3$ is volatile, obtaining an adequate InGaAs etch rate is the major issue. Another important output parameter is the GaAs/InGaAs etch ratio. In the HBT epitaxial structures, there is a 1000 – 3000 Å GaAs cap layer between the InGaAs contact layer and the AlGaAs emitter layer. Because the GaAs etch rate is usually larger than that for InGaAs, during the InGaAs etch process, any over-etch would quickly etch the GaAs cap layer away and affect the formation of passivation ledges. In order to control the effect of the over-etch, which is necessary because the dry etch rate is not sufficiently uniform across the whole wafer, a relatively low GaAs/InGaAs etch ratio is desirable. For example, if the total InAs/InGaAs layer thickness is 500 Å and the etch ratio is 4, then a 50% InGaAs over-etch will remove $500 \times 50\% \times 4 = 1000$ Å GaAs underneath InGaAs at most. Therefore, this dry etch condition can be used as long as the GaAs cap layer is thicker than 1000 Å. Figure 4-9 shows the dependence of the InGaAs etch rate and GaAs/InGaAs etch ratio on input variables analyzed by using a linear model.

Looking first at the InGaAs etch rate, Figure 4-8 and Figure 4-9 clearly show that the InGaAs etch rate and the DC Bias have the same dependence on input variables, which suggests that the InGaAs etch process is basically energy-driven. As shown in Figure 4-9, the InGaAs etch rate was higher when the RF Power was larger or the
Chamber Pressure was lower. When the Flow Rate of SiCl$_4$ or SF$_6$ was reduced, the InGaAs etch rate increased because of the higher self-established DC Bias under these conditions. This is completely different than the GaAs etch rate which was dominated by a chemical etching mechanism.

The InGaAs/GaAs etch ratio is defined as the ratio of the two etch rates under the same etch conditions. It is very important to lower this etch ratio during the InGaAs layer etch. Because the etch rate around the wafer edge is usually larger than at the center of the wafer, in order to remove the InGaAs completely, a 50 to 100 percent over etch is required. During this over etch, the GaAs cap layer around the wafer edge where the InGaAs is first cleared will be etched away at the GaAs etch rate. Therefore, in order to prevent the passivation layer from being affected by this over etch, the InGaAs/GaAs etch ratio must be smaller than the thickness ratio of the cap layer over the contact layer. In our epitaxial design, this thickness ratio is between 2 and 6, so the etch ratio has to be controlled under 6. As indicated by Table 4-5 and Figure 4-9, increasing RF Power, decreasing Chamber Pressure and not using SF$_6$ all lower the etch ratio. Although a lower etch ratio always means a high DC Bias, thus enhancing the plasma damage to the etched surface, the InGaAs etch is not the final step, so it is not very critical to maintain good surface quality. In the case of etching down to the base, the base surface is always finished by a wet polishing etch.
<table>
<thead>
<tr>
<th>Input Variables</th>
<th>Output Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>DOE</td>
<td>RF Power (Watt)</td>
</tr>
<tr>
<td>Trial number</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>80</td>
</tr>
<tr>
<td>8</td>
<td>60</td>
</tr>
<tr>
<td>1</td>
<td>60</td>
</tr>
<tr>
<td>5</td>
<td>80</td>
</tr>
<tr>
<td>9</td>
<td>80</td>
</tr>
<tr>
<td>10</td>
<td>60</td>
</tr>
<tr>
<td>6</td>
<td>60</td>
</tr>
<tr>
<td>7</td>
<td>80</td>
</tr>
<tr>
<td>1</td>
<td>60</td>
</tr>
<tr>
<td>4</td>
<td>60</td>
</tr>
<tr>
<td>3</td>
<td>80</td>
</tr>
<tr>
<td>2</td>
<td>80</td>
</tr>
<tr>
<td>2</td>
<td>80</td>
</tr>
<tr>
<td>4</td>
<td>60</td>
</tr>
<tr>
<td>3</td>
<td>80</td>
</tr>
</tbody>
</table>

Table 4-5 The values of the input and output variables in the DOE test performed for developing SiCl₄-based InGaAs dry etch.
Figure 4-9 The dependence of the InGaAs etch rate and GaAs/InGaAs etch ratio (GaAs_InGaAs_Sel in the plots) on input variables. a) the effects of RF Power and Pressure (33 sccm SiCl₄, 5.5 sccm SF₆ and 110 sccm He); b) the effects of SiCl₄ and SF₆ (70 watts RF Power, 32 m-torr Chamber Pressure, 110 sccm He); the effects of He and SF₆ (70 watts RF Power, 32 m-torr Chamber Pressure, 33 sccm SiCl₄).
4.3.4 Dry etching of dielectric and metal

In our HBT fabrication processes, dry etching techniques were also used to etch several other materials, which include 1) tungsten (W) metal during the emitter metalization process, 2) silicon nitride during the emitter mesa formation process, 3) silicon dioxide during the interconnect metal via hole process, and 4) polyimide during the interconnect metal via hole process. Although these dry etch processes were relatively easy when compared with the III-V compound semiconductor dry etch, some cautions are still required in order to obtain optimal results.

Table 4-6 lists the recipes that were developed for dry etching sputtered W, PECVD silicon nitride and oxide, and polyimide. As shown by the table, the chamber pressure for all these recipes is 100 m-torr, which is relatively high, so that some isotropic etch will undercut the mask. As a result, the critical dimension control is limited if the total etch time is long. O\textsubscript{2} is included as a reactant because some polymer will form during the freon etching processes and the O\textsubscript{2} plasma can remove the polymer. Sometimes, an extra 30 seconds pure O\textsubscript{2} plasma is necessary to completely clear the wafer after freon etch.

In the case of etching PECVD SiN or SiO\textsubscript{2}, the etch rate varies a lot depending the film synthesis conditions. This is because the density and stoichiometry of the PECVD film vary from machine to machine or when the different synthesis conditions, such as substrate temperature and chamber pressure, are used. Therefore, in order to control the etch process, a dummy wafer covered by an identical film should be used to calibrate the etch rate before etching the real wafers. In addition, some over etch is always needed because of the loading effect in the dry etch processes.

In the case of etching polyimide, patterned photoresist is usually the masking layer. Because an O\textsubscript{2} plasma etches photoresist as fast as it etches polyimide, the photoresist has to be thicker than the polyimide. Because the O\textsubscript{2} plasma etches the photoresist laterally as well, the opening of the polyimide pattern after etching is wider than the designed size by the amount of the polyimide thickness.
<table>
<thead>
<tr>
<th>Material To Be Etched</th>
<th>RF Power (Watt)</th>
<th>Chamber Pressure (m-Torr)</th>
<th>Freon Flow Rate (sccm)</th>
<th>Additive Flow Rate (sccm)</th>
<th>Etch Rate (Å/min.)</th>
<th>Cautions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sputtered W</td>
<td>300</td>
<td>100</td>
<td>C₂F₆: 100</td>
<td>O₂: 10</td>
<td>~207</td>
<td>50% over etch to clear</td>
</tr>
<tr>
<td>PECVD SiN</td>
<td>100</td>
<td>100</td>
<td>CHF₃: 100</td>
<td>O₂: 10</td>
<td>~620</td>
<td>etch rate varies</td>
</tr>
<tr>
<td>PECVD SiO₂</td>
<td>100</td>
<td>100</td>
<td>CHF₃: 100</td>
<td>O₂: 10</td>
<td>~270</td>
<td>etch rate varies</td>
</tr>
<tr>
<td>Polyimide</td>
<td>100</td>
<td>100</td>
<td>-</td>
<td>O₂: 10</td>
<td>~3826</td>
<td>partially cured</td>
</tr>
</tbody>
</table>

Table 4-6 The conditions developed for dry etching sputtered W, PECVD silicon nitride and oxide, and polyimide.
4.4 Four Approaches to Form Sub-micron Passivation Ledges

Depleted AlGaAs passivation ledges have been used to reduce surface recombination around the emitter perimeter of AlGaAs/GaAs hetero-junction bipolar transistors since 1985. \textsuperscript{[4-4]} Since then, research has shown that this passivation ledge not only increases the current gain, \textsuperscript{[4-5]} but also reduces the low frequency noise \textsuperscript{[4-6]} and improves the reliability of HBTs. \textsuperscript{[4-7]} However, this passivation ledge increases the base-collector junction capacitance and the base resistance, thus decreasing the cut-off frequency, \( f_t \), and the maximum oscillation frequency, \( f_{\text{max}} \), of HBTs. The research results presented in Chapter 5 show that 0.3 \( \mu \text{m} \) is the minimum ledge length necessary to fully passivate the extrinsic base of 0.1 \( \mu \text{m} \) base width HBTs. The ledge length is defined as the distance from emitter mesa to the end of AlGaAs ledge. This observation indicates that RF device performance can be optimized by utilizing passivation ledges as short as 0.3 \( \mu \text{m} \). The challenge then is to incorporate sub-micron passivation ledges into HBT fabrication. This section will describe four different processing approaches and compare the relative advantages and disadvantages of each. These approaches include: 1) Electron-Beam Lithography, 2) Selective Undercut Etching, 3) Self-Aligned Undercut Etch and Base Lift-off, and 4) Silicon Dioxide Sidewalls.

4.4.1 The Electron-Beam Lithography Approach

The process flow of the Electron-Beam Lithography approach is outlined in Figure 4-10. The emitter metal and alignment mark patterns are defined by e-beam lithography and formed by PMMA lift-off; the emitter mesa is formed by selectively wet-etching down to the AlGaAs layer. (Figure 4-10a) The passivation ledges are defined by e-beam direct writing using image reversal e-beam resist SAL601, and non-selectively etching down to the base. (Figure 4-10b) The residual SAL601 is removed using a high pressure \( \text{O}_2 \) plasma. (Figure 4-10c) A SEM picture of the finished emitter and passivation ledges is shown in Figure 4-10d.
Figure 4-10 The process flow of the Electron-Beam Lithography Approach.

The Electron-Beam Lithography approach is straightforward and has precise control of ledge lengths, but it requires an electron beam direct writing system, which is complex and intrinsically limits process throughput for even low integration levels. The passivation ledge length is defined by e-beam lithography. To minimize the e-beam writing time, an image reversal e-beam resist, Shipley SAL601 in this case, has to be used. This approach is a useful research tool because HBTs with different passivation ledge lengths can be processed on the same wafer, and because the dimensions can be easily changed, since there is no mask.

4.4.2 The Selective Undercut Etching Approach

The process flow of the Selective Undercut Etching approach is outlined in Figure 4-11. The emitter mesa is formed by anisotropically etching away the InGaAs contact layer and GaAs cap layer. During the etching process, the emitter metal serves as the etching mask and the AlGaAs passivation ledge layer as the etch stop. (Figure 4-11a) The selective undercut is realized using a special high chamber pressure and low DC-bias etching recipe. Due to the formation of AlF₃ that is non-volatile on the top of the AlGaAs passivation ledge layer, this recipe gives high etch selectivity. This undercut etching technique can also be used to form the sub-micron emitter mesa. (Figure 4-11b) After the undercut etch, the wafers are coated with positive resist and exposed without a mask. Because the resist underneath the undercuts stays after developing, passivation ledges with a length roughly
equal to the undercut size will form after the finish of the next etch step. (Figure 4-11c) The residual resist is removed and base metalization is done by normal processing techniques described in section 4.1. (Figure 4-11d) A SEM picture of the undercut structure processed using this approach is shown in Figure 4-11e.

![Diagrams](image)

**Figure 4-11** The process flow of the Selective Undercut Etching Approach.

Table 4-7 summarizes the etching recipes used in the Selective Undercut Etching approach. For InGaAs etch, since InCl₃ is not very volatile, the etching mechanism is basically physical sputtering. However, by using a relatively high DC-bias, an etch rate of
about 200 Å/min., which is adequate for HBT fabrication, can be achieved. For the GaAs emitter cap etch, SF\textsubscript{6} is used together with SiCl\textsubscript{4} to provide the etch selectivity. With a moderate DC-bias, both anisotropy and selectivity can be obtained. For the GaAs undercut etch, a high chamber pressure and large reactive gas flow rates are maintained to lower the DC-bias and enhance the chemical etching mechanism. Under this condition, the experimental data show that the vertical and the lateral etch rates are roughly the same so that the undercutting amount can be controlled by timing the etch. Finally, for the AlGaAs emitter layer etch, only SiCl\textsubscript{4} is used and in order to minimize the plasma damage to the base surface, the RIE process is controlled to stop about 300 Å above the base layer and followed by a wet etch step to realize a smooth base surface.

The Selective Undercut Etching approach is a process that can form sub-micron passivation ledges and sub-micron emitters, and thus is suitable for fabricating high performance sub-micron HBTs. The advantages of this approach are three-fold: first, a sub-micron passivation ledge is formed; second, with undercut, the emitter mesa width can be made less than one micron - smaller than the resolution of optical lithography; finally, because the selective etch does not affect the emitter contact area, sub-micron emitter width HBTs processed with this approach will not suffer from high emitter contact resistance. The disadvantages of this approach are mainly two: first, this approach requires a highly selective GaAs/AlGaAs RIE system with very good GaAs etch rate uniformity because the AlGaAs ledge is formed by intentionally undercutting the GaAs with respect to the emitter mask in a selective RIE of the emitter mesa; second, the ledge length may not be very accurate because the undercut distance is controlled by timing the overetch.

<table>
<thead>
<tr>
<th></th>
<th>RF Power Watts</th>
<th>Chamber Pressure mTorr</th>
<th>SiCl\textsubscript{4} Flow Rate sccm</th>
<th>SF\textsubscript{6} Flow Rate sccm</th>
<th>He Flow Rate sccm</th>
<th>Measured DC Bias Volts</th>
</tr>
</thead>
<tbody>
<tr>
<td>InGaAs Contact Layer Etch</td>
<td>60</td>
<td>27</td>
<td>22</td>
<td>0</td>
<td>130</td>
<td>375</td>
</tr>
<tr>
<td>GaAs Emitter Cap Etch</td>
<td>40</td>
<td>27</td>
<td>22</td>
<td>11</td>
<td>130</td>
<td>200</td>
</tr>
<tr>
<td>GaAs Undercut Etch</td>
<td>40</td>
<td>60</td>
<td>50</td>
<td>15</td>
<td>90</td>
<td>70</td>
</tr>
<tr>
<td>AlGaAs Emitter Layer Etch</td>
<td>40</td>
<td>27</td>
<td>22</td>
<td>0</td>
<td>130</td>
<td>260</td>
</tr>
</tbody>
</table>

Table 4-7 A summary of RIE recipes used in the Selective Undercut Etching approach.
4.4.3 The Self-Aligned Undercut-Etch and Base Lift-off Approach

The process flow of the Self-Aligned Undercut-Etch and Base Lift-off Approach is outlined in Figure 4-12. Standard wet-etching techniques are used to selectively etch down to the top of AlGaAs passivation ledge layer. (Figure 4-12a) The wafer is coated first by PECVD silicon dioxide and then by PMMA. The base metal position is defined by e-beam lithography. After developing the PMMA, the silicon dioxide underneath the PMMA is undercut using a buffer HF etch. Then a non-selective wet etchant is used to etch down to the base, followed by base metal deposition. (Figure 4-12b) PMMA is used to lift-off the base metal. (Figure 4-12c) Figure 4-12d shows a SEM picture of the oxide undercut and the passivation ledge. Figure 4-12e shows a SEM picture of the base metal and the sub-micron spacing formed by this approach.

Figure 4-12 The process flow of the Self-Aligned Undercut-Etch and Base Lift-off approach.

In this approach, the sum of the ledge length and the ledge-to-base-metal spacing is defined by e-beam lithography, i.e., the layout design. Because the spacing itself is determined by the amount of the oxide undercut, the passivation ledge length can be
controlled by controlling the oxide undercut etching time. Figure 4-13 shows the dimension of the spacing as a function of the BOE undercut etching time. The sensitivity of the critical dimension control is about 500 Å per 15 seconds, which is good enough for fabricating HBTs with 0.1 μm ledges. It is important to point out that this approach does not require e-beam lithography. As long as a very high overlay accuracy can be reached, an i-line stepper can replace the e-beam lithography system, thus making this approach more manufacturable and less costly.

![Graph showing spacing vs PMMA line width](image)

**Figure 4-13** The Dimension of the ledge-to-base-metal spacing as a function of the BOE undercut etching time.

The Self-Aligned Undercut Etch and Base Lift-off approach is a process which simultaneously forms a sub-micron passivation ledge and a sub-micron ledge-to-base-metal spacing. This processing approach requires neither e-beam lithography nor selective GaAs/AlGaAs RIE. Instead, the dimensions of the passivation ledges can be controlled by the precise inter-level alignment of a stepper and by the amount of oxide undercut. In addition to its simplicity, this approach allows a thick base metal layer because it uses resist instead of the emitter mesa to lift-off the base metal. The disadvantage of this approach is
that it cannot be used for submicron emitter size HBTs because wet-etch techniques are used to form the emitter mesa.

4.4.4 The Silicon Dioxide Sidewall Approach

The previous three processing approaches were developed during this dissertation research. The fourth one - the Silicon Dioxide Sidewalls approach - was developed and published by NTT. In this processing approach, an emitter mesa is formed by an anisotropic etch using photoresist as a mask. After reaching the AlGaAs passivation ledge layer, a thin silicon dioxide layer ( < 0.2 μm) is deposited and then anisotropically etched away. As a result, a thin silicon dioxide sidewall forms. The following etch utilizes this sidewall as a mask to etch down to the base layer, thus forming a short ledge around the emitter mesa. The passivation ledges formed this way have a dimension roughly equal to the thickness of the sidewall. Figure 4-14 shows a schematic drawing of the cross-section of a HBT device processed by this approach.

![Figure 4-14 The Silicon Dioxide Sidewalls approach.][4-8]

The main advantage of this approach is the control of the ledge dimensions. However, the ledge length in this approach is limited by the thickness of the sidewall dielectric. This thickness is typically no larger than 0.2μm to allow a well controlled etch back of the dielectric. To fabricate passivation ledges as large as 0.3μm, some modifications might be necessary, such as a small undercut formed before the sidewall fabrication.
In summary, four different processing approaches for fabricating sub-micron ledges have been described. The Selective Undercut Etching approach is probably the best way to realize sub-micron HBTs with low emitter resistance, high current gain and good reliability. The Self-Aligned Undercut Etch and Base Lift-off approach and Silicon Dioxide Sidewall approach are generally suitable for manufacturing HBT-ICs, but the emitter sizes are larger. Finally, if very precise dimension control is required, e-beam lithography can be used. This approach is a valuable research tool, but is unlikely to be used in HBT manufacturing because of its low process throughput.
Chapter 5
DC HBT Performance with and without Sub-Micron Passivation Ledges

5.1 Device Simulation Results

In order to study the effects of different passivation ledge lengths and find the minimum ledge length for full passivation, both analytical calculations and numerical simulations were done to evaluate the current gain as a function of the passivation ledge lengths. Section 5.1.1 describes the analytical formulation and results. Section 5.1.2 presents the Semi-Cad based 2-dimensional numerical simulation results.

5.1.1 Analytical Simulation

As discussed in Chapter 3, the current gain of HBTs is mainly limited by the base transport factor because the emitter injection efficiency is very close to unity. In general, the base current of a passivated Npn HBT consists of five components, each indicated in Figure 5-1: (1) $I_1$, interface recombination current occurring at the interface between the passivation ledge and the extrinsic base region covered by the ledge; (2) $I_2$, interface recombination current at the base contact; (3) $I_3$, surface recombination current at the exposed extrinsic base surface between the ledge tip to the contact; (4) $I_4$, bulk recombination current occurring in the base region and (5) $I_5$, trap recombination current in the emitter-base space-charge region. Figure 5-1 shows a schematic drawing of the base region and parameters used in the analytical model and calculations. Only half of the device is shown and required for calculations because the other half is identical and the boundary condition along the bisected emitter plane is reflective. In Figure 5-1, the numbers in parentheses represent the location of each of the base recombination currents described above.

In order to analytically describe the device in Figure 5-1, the fifth component of the base current will not be considered; otherwise, a two-dimensional numerical simulation has to be used in order to solve Poisson's equation and continuity equation simultaneously. (Section 5.1.2 presents the simulation results obtained using a two-dimensional simulator called Semi-Cad.) Although this component is generally not negligible, the analytical calculation is greatly simplified and yields reasonable accuracy when the surface and interface recombination components dominate the base current, which they do unless the base passivation is very good.
\[ d_1 = \frac{W_e}{2} \]
\[ d_2 = \frac{W_e}{2} + L \]
\[ X_0 = \frac{W_e}{2} + L + S + W_{bm} \]

Figure 5-1 The schematic drawing of the base region where the analytical simulation is performed to investigate the different components of the base current. The width of the structure along the dimension perpendicular to the cross section is \( W \).

The equation to be solved is the 2-dimensional current continuity equation in the base region:

\[ \nabla^2 n(x, y) + f \frac{\partial n(x, y)}{\partial y} - \frac{n(x, y)}{\tau_n D_n} = 0. \]  \hspace{1cm} (5.1)

where \( n(x, y) \) is the electron concentration distribution in the base region, \( \tau_n \) and \( D_n \) are the lifetime of the minority carrier electrons and the diffusion constant for electrons in the base region, respectively. \( D_n \) is taken to be \( q\mu_n kT \) and \( \mu_n \) is the minority electron mobility. \( f \), the normalized base electric field, is defined as

\[ f = \frac{q}{kT} E_{\text{base}}. \]  \hspace{1cm} (5.2)
where \( E_{\text{base}} \) is the magnitude of the uniform base quasi-electric field in the \( y \)-direction. If the sign of this quasi-electric field is negative, then it increases the base transport factor and reduces the surface recombination effect, thus improving device performance. In AlGaAs/GaAs HBTs, such an electric field can be established by grading the Al content in the base region from the emitter to the collector.

The boundary condition for solving the current continuity equation is:

\[
\vec{N} \cdot \nabla n + n \vec{N} \cdot \vec{E} = -\frac{s}{D_n} n(x,y),
\]  

(5.3)

where \( \vec{N} \) is the unit outward normal vector to the surface under consideration, \( \vec{E} \) is the unit vector of electric field at the boundary, and \( s \) is the surface recombination velocity. Referring to Figure 5-1, we can express the boundary conditions for different interfaces and surfaces as follow:

\[
\frac{\partial n}{\partial y} + f n = -\frac{J_n}{q D_n}, \quad 0 \leq x < d_1 \quad \text{and} \quad y = 0, \quad \text{underneath emitter} \quad (5.4a)
\]

\[
\frac{\partial n}{\partial y} + f n = \frac{s_{d}}{D_n} n, \quad d_1 \leq x < d_2 \quad \text{and} \quad y = 0, \quad \text{passivated interface} \quad (5.4b)
\]

\[
\frac{\partial n}{\partial y} + f n = \frac{s_{\text{max}}}{D_n} n, \quad d_2 \leq x \leq X_0 \quad \text{and} \quad y = 0, \quad \text{unpassivated surfaces} \quad (5.4c)
\]

\[
\frac{\partial n}{\partial x} = 0, \quad x = 0 \quad \text{and} \quad 0 \leq y \leq Y_0, \quad \text{reflective boundary} \quad (5.4d)
\]

\[
-\frac{\partial n}{\partial x} = \frac{s_1}{D_n} n, \quad x = X_0 \quad \text{and} \quad 0 \leq y \leq Y_0, \quad \text{etched mesa surface} \quad (5.4e)
\]

\[
n(x,Y_0) = 0, \quad 0 \leq x \leq X_0 \quad \text{and} \quad y = Y_0, \quad \text{interface to collector} \quad (5.4f)
\]

For the base-collector interface, the electron concentration is taken to be zero because the base-collector junction is reverse biased in the normally active bias regime in which HBTs are commonly operated.
A solution to Equation (5.1) with boundary conditions given by Equations (5.4) is obtained by separation of variables. A general solution of \( n(x,y) \) that matches the boundary conditions (5.4d), (5.4e) and (5.4f) can be expressed as follow:

\[
n(x,y) = e^{-f\gamma/2} \sum_m A_m \sinh \left[ \gamma_m (y - Y_0) \right] \cos \left( \lambda_m x \right), \quad (5.5)
\]

where \( \lambda_m \) is the \( m \)th root of the transcendental equation:

\[
\lambda_m X_0 \tan \left( \lambda_m X_0 \right) = \frac{s_1}{D_n} X_0. \quad (5.6)
\]

and \( \gamma_m \) satisfies the following equation:

\[
\gamma_m^2 = \lambda_m^2 + \left( \frac{f}{2} \right)^2 + \frac{1}{\tau_n D_n}. \quad (5.7)
\]

\( A_m \) is an unknown multiplying constant determined from the remaining boundary conditions (5.4a), (5.4b) and (5.4c). In order to determine \( A_m \), one can express \( A_m \) as \( \psi_m K_m \), where:

\[
K_m = \left[ \gamma_m \cosh \left( \gamma_m Y_0 \right) - \frac{f}{2} \sinh \left( \gamma_m Y_0 \right) \right]^{-1}. \quad (5.8)
\]

With substitution of the general solution, (5.5), into the remaining boundary conditions, \( \psi_m \) can be determined from the following reduced boundary conditions:

\[
\frac{\partial n}{\partial y} + fn = \sum_m \psi_m \cos \left( \lambda_m x \right) = -\frac{J_n}{qD_n}, \quad 0 \leq x < d_1 \text{ and } y = 0, \quad (5.9a)
\]
\[
\frac{\partial n}{\partial y} + fn = \sum_m \psi_m \cos \left( \lambda_m x \right) = \frac{s_a}{D_n} n, \quad d_1 \leq x < d_2 \text{ and } y = 0, \quad (5.9b)
\]
\[
\frac{\partial n}{\partial y} + fn = \sum_m \psi_m \cos \left( \lambda_m x \right) = \frac{s_{\text{max}}}{D_n} n, \quad d_2 \leq x \leq X_0 \text{ and } y = 0. \quad (5.9c)
\]
The solution of $\psi_m$ is obtained using the technique described by Kennedy in references \(^5\). After determination of $\psi_m$, the solution of $n(x,y)$ is known for the entire base region and the different base current components are obtained as follows:

$$I_{\text{intfp}} = q s_a \int_A n(x,y) \, dA = -qs_a W \sum_m \psi_m K_m \sinh \left( \gamma_m Y_0 \right) \frac{\sin \left( \lambda_m d_2 \right) - \sin \left( \lambda_m d_1 \right)}{\lambda_m},$$

passivated interface \hspace{1cm} (5.10a)

$$I_{\text{surf}} = q s_{\text{max}} \int_A n(x,y) \, dA = -qs_{\text{max}} W \sum_m \psi_m K_m \sinh \left( \gamma_m Y_0 \right) \frac{\sin \left( \lambda_m X_0 \right) - \sin \left( \lambda_m d_2 \right)}{\lambda_m},$$

unpassivated surface \hspace{1cm} (5.10b)

$$I_{\text{edge-surf}} = qs_1 \int_A n(x,y) \, dA = qs_1 W \sum_m \psi_m K_m B_m \cos \left( \lambda_m X_0 \right),$$

etched mesa surface \hspace{1cm} (5.10c)

$$I_{\text{bulk}} = q \frac{1}{\tau_n} \int_V n(x,y) \, dV = q \frac{1}{\tau_n} W \sum_m \psi_m K_m B_m \frac{\sin \left( \lambda_m X_0 \right)}{\lambda_m},$$

bulk base region \hspace{1cm} (5.10d)

where

$$B_m = \frac{\exp \left( \frac{\left( \gamma_m + \frac{f}{2} \right) Y_0}{2} \right)}{\gamma_m - \frac{f}{2}} \sinh \left( \frac{\left( \gamma_m - \frac{f}{2} \right) Y_0}{2} \right) - \frac{\exp \left( \frac{\left( \gamma_m - \frac{f}{2} \right) Y_0}{2} \right)}{\gamma_m + \frac{f}{2}} \sinh \left( \frac{\left( \gamma_m + \frac{f}{2} \right) Y_0}{2} \right)$$

and $W$ is the device width defined in Figure 5-1. $I_{\text{intfp}}$ is the passivated extrinsic base interface recombination current, $I_{\text{surf}}$ is the unpassivated extrinsic base surface recombination current, $I_{\text{edge-surf}}$ is the base mesa edge surface recombination current, and $I_{\text{bulk}}$ is the bulk recombination current inside the base region.
The number of terms m to be used in the summation of equation (5.9) depends on the accuracy desired and computation time. An estimate of the accuracy of the solution can be obtained by calculating the collector current across the surface y = Y₀:

\[ I_c = -qD_n W e^{-\frac{F_0}{2}} \sum_m \psi_m K_m \gamma_m \frac{\sin(\lambda_m X_0)}{\lambda_m} \]  

(5.12)

and comparing this to the relative difference between \((I_c - I_b)\) as the number of terms m increases. In this research, m = 200 was used and leads to excellent accuracy (error < 0.1%).

Figure 5-2  The calculated current gain of HBTs as a function of passivation ledge length with different base widths. The emitter size is 2 x 10 μm², the recombination velocity is 1 x 10⁶ cm/sec. for an unpassivated surface and 1 x 10³ cm/sec. for passivated interfaces. There is no base electric field.

One of the most important results obtained from this analytical calculation is the saturation in HBT current gain when the passivation ledge length becomes larger than a critical value. Figure 5-2 shows the calculated current gain of HBTs as a function of the passivation ledge length with different base widths. When passivation ledge length increases, current gain increases because fewer of the electrons injected from the emitter are able to diffuse laterally and reach an unpassivated surface where they recombine with holes, thus creating recombination current. The point at which the current gain reaches its
95% of its maximum is shown in Figure 5-2 and the current gain curve saturates indicating that most of the surface recombination has been eliminated and thus become a small component among the remaining base current constituents.

The analytical calculation also shows that the minimum ledge length for full passivation is proportional to the base width. Using the 95% definition shown in Figure 5-2, the ledge length required to produce full passivation is about 3x the base width when there is no built-in base field. Experimental verification of this result would be very useful for optimizing HBT design. Sections 5.2 and 5.3 describe the design of experiments and compare the measured results against this analytic simulation.

![Graph showing current gain vs passivation ledge length](image)

**Figure 5-3** The calculated current gain of HBTs as a function of passivation ledge length with different base grading. The emitter size is $2 \times 10 \, \mu m^2$, the recombination velocity is $1 \times 10^6 \, cm/sec.$ for unpassivated surface and $1 \times 10^3 \, cm/sec.$ for passivated interface. The base width is 1000Å.

Figure 5-3 shows the calculated current gain of HBTs as a function of passivation ledge length for different base grading. When the Al composition is linearly graded from $x = 0.15$ (or 0.10, and 0.05) to 0.00 through the 1000Å thick base, a quasi-electric field of -18000 (or 12000, and 6000) V/cm exists. This electric field reduces the base transit time, which reduces both the bulk base and surface recombination currents, thus significantly increasing the current gain. This electric field also makes the minimum ledge length for full passivation smaller than its zero field value. Using the same 95% current gain definition,
the minimum ledge length for full passivation decreases from 0.3 μm for zero field base structure to about 0.16 μm for 0.05 to 0.0 graded base, 0.08 μm for 0.10 to 0.0 graded base, and 0.06 μm for 0.15 to 0.0 graded base. This decrease in passivation length results from electrons being accelerated by the strong vertical field toward the base-collector junction and their lateral diffusion towards the unpassivated base surface is greatly reduced so that the extrinsic base surface recombination is eliminated.

Figure 5-4 shows the current gain dependence on the value of surface recombination velocity underneath the passivation ledge. This surface recombination velocity, \( s_a \), is used in equations (5.4b) and (5.9b), i.e., the boundary condition for region (1) in Figure 5-1. Similar to the trend that has been observed in Figure 5-2, the current gain dependence on the surface recombination velocity has a saturation plateau of \( 1 \times 10^4 \) cm/sec. below which the surface recombination velocity underneath the passivation ledge does not dominate the current gain. This result demonstrates that the AlGaAs/GaAs interface underneath the passivation ledge does not have to be perfect or yield zero recombination velocity because the current gain is insensitive to \( s_a \) as long as the interface is “reasonably” passivated (i.e. \( s_a \leq 1 \times 10^4 \) cm/sec.).

![Current Gain Dependence on Surface Recombination Velocity](image)

Figure 5-4: The current gain dependence on surface recombination velocity, \( s_a \), underneath the passivation ledge.

Figure 5-5 shows the effect of the unpassivated base surface. \( s_{\text{max}} \) is the surface recombination velocity used in equations (5.4c) and (5.9c), i.e., the boundary condition for
region (2) and (3) in Figure 5-1. While for an unpassivated GaAs surface, \( s_{\text{max}} \) is about 1 \( \times \) \( 10^6 \) cm/sec, it can become as large as the thermal velocity, \( 2 \times 10^7 \) cm/sec., underneath the metal contact. Figure 5-5 shows that when an HBT is not fully passivated, the larger the value of \( s_{\text{max}} \), the smaller the current gain; however, if the passivation ledge is longer than the minimum passivation ledge length, the value of \( s_{\text{max}} \) makes no difference. For a fully passivated HBT, the surface recombination current is a very small component of the total base current and the value of \( s_{\text{max}} \) is unimportant.

![Graph showing current gain dependence on surface recombination velocity](image)

**Figure 5-5** The current gain dependence on the surface recombination velocity \( s_{\text{max}} \) at the unpassivated base surface.

5.1.2 Semi-Cad 2-dimensional Device Simulation

Semi-Cad is a general purpose two-dimensional device simulator, capable of simulating a wide range of devices, such as diodes, FETs, BJTs, thyristors, and CCDs. We chose Semi-Cad to simulate the ledge passivation effects on AlGaAs/GaAs HBTs for three reasons. First, Semi-Cad allows the user to create and simulate devices with heterostructures, such as AlGaAs/GaAs HBTs. Second, Semi-Cad allows the user to model Fermi level pinning by simulating dynamic traps at surfaces. It also allows the user to supply values of srm and srvp, parameters which describe carrier surface recombination velocities for Shockley-Read-Hall traps. Finally, Semi-Cad allows the user to specify any profile combination of composition and doping by creating a user-defined analytically
function. This flexibility is extremely useful in describing the parabolic grading at the emitter-base junction and the composition-dependent carrier lifetimes in the graded emitter.

Comparison of the analytical calculation described in the previous section with the Semi-Cad simulation exhibits several differences. First, because Semi-Cad solves both carrier continuity equations and Poisson's equation simultaneously, it is easy to include the recombination current component in the emitter-base space-charge region. As a result, the simulated current gain is expected to be lower when comparing to those obtained previously. Second, the Semi-Cad default GaAs mobility models are used and they are different than the assumed constant value used in the analytical calculation. Finally, in the Semi-Cad simulation, the effects of both surface recombination velocity and the built-in electric field from surface Fermi level pinning are considered.

![Graph showing current gain vs. emitter-base bias for different passivation ledge lengths.](image)

Figure 5-6 The Semi-Cad simulation results for current gain dependence on emitter-base bias for HBTs with different passivation ledge lengths.

The Figure 5-6 shows the Semi-Cad simulation results for current gain as a function of emitter-base bias. When the passivation ledge length increases, the current gain increases up to a passivation ledge length of 0.4 μm, after which the current gain curves saturate. Figure 5-7 compares the Semi-Cad simulation with the analytical calculation. Both approaches predict the same trend in terms of current gain saturation and the existence of a minimum ledge length for full passivation. However, because the Semi-Cad simulation uses different mobility models and includes the emitter-base space-charge region recombination, the predicted current gain is lower than the analytical approach. In addition,
the passivation ledge length required to reach gain saturation is about 0.4 \( \mu m \) from Semi-Cad simulation compared of 0.3 \( \mu m \) for the analytical solution. In Section 5.3, the experimental data are to be compared for both simulation approaches.

![Graph showing current gain vs passivation ledge length](image)

**Figure 5-7** Comparison between Semi-Cad simulation and analytical calculation for current gain versus passivation ledge length.

In order to understand why a relatively short ledge can fully passivate the extrinsic base of a HBT, it is instructive to examine the carrier distributions in the device. Figure 5-8 shows a contour map of the electron concentration of a passivated HBT device under high current injection. As shown by the contour map, the electron concentration drops by several orders of magnitude with lateral distance away from the emitter edge of a few tenths of a micron. The surface recombination current is proportional to the surface electron concentration, so if the unpassivated base surface is 0.3 \( \mu m \) away from the emitter edge, the surface recombination current is expected to decrease by at least one order of magnitude, thus making the surface effect insignificant. Figure 5-9 shows a zoom-in of Figure 5-8 near the emitter edge and the electron concentration in the extrinsic base. The plot at the bottom left of Figure 5-9 shows that the electron concentration right underneath the extrinsic base surface drops by about a factor of 20 within 0.2 \( \mu m \) distance and then starts to increase and form a peak at \( x = 0.3 \mu m \). The formation of the peak is because the passivation ledge length is only 0.3 \( \mu m \), beyond which the unpassivated base surface has a very high recombination velocity so that a lot of electrons pile up there. Figure 5-10 shows a similar contour map for an unpassivated HBT device. When there is no passivation ledge
at all, the lateral spreading of electrons is much wider inside the extrinsic base, thus making the surface recombination current the dominant base current component and reducing the gain.

Figure 5-8  Contour map of electron concentration for a passivated HBT device under normal operation. The emitter size is 2 x 10 μm² and only half of the device is shown. The base width is 1000Å and the passivation ledge length is 0.3 μm. The HBT device is biased at $V_{be} = 1.4$ volts and $V_{ce} = 1.4$ volts.
Figure 5-9 A zoom view of the contour map in Figure 5-8 and the electron concentration plots in the extrinsic base. Top: zoom view of the contour map in Figure 5-8. Bottom Left: electron concentration right underneath the extrinsic base surface ($0 < x < 0.3$ passivated, and $0.3 < x$ unpassivated). Bottom Right: electron concentration 100Å beneath the extrinsic base surface.
Figure 5-10  Contour map of electron concentration for an unpassivated HBT device under normal operation. The emitter size is 2 x 10 μm² and only half of the device is shown. The base width is 1000Å and there is no passivation ledge. The HBT device is biased at $V_{be} = 1.4$ volts and $V_{ce} = 1.4$ volts.
5.2 Design of Experiments

In order to study the effects of different passivation ledge lengths experimentally and verify the design rules for the minimum ledge length for full passivation predicted by the simulations, we designed and fabricated a group of HBT devices. The epitaxial structure of these HBTs has been shown in Table 3-2 and discussed in Section 3.6.2. Section 5.2.1 describes the overall process flow that integrates all the processing modules and Section 5.2.2 presents the layout used for the fabrication.

5.2.1 Process Integration

Because current gain is the main parameter used to delineate the effects of different passivation ledge lengths, DC HBTs were chosen for this study. The main steps for this DC process includes the following: (1) emitter metal formation, (2) passivation ledge formation, (3) base metal formation, (4) base mesa formation, (5) dielectric deposition and 2nd-metal VIA definition, and (6) 2nd-metal formation.

The whole process flow is identical to the standard DC HBT fabrication processes shown in Figure 4-3 except steps (1) and (2). In order to fabricate the sub-micron passivation ledges and accurately control the ledge dimension, e-beam lithography was used to form the emitter metal and passivation ledges as described in Section 4.4.1. When e-beam and optical lithography technologies are combined into one process flow, two important issues have to be considered. The first is the size of the layout cell. Because the field size in e-beam lithography is always small -- 2 mm x 2 mm for the Hitachi HL-700F, one optical lithography layout cell is usually a composite of several e-beam fields which have to be taped out separately and added together during the exposure process. The second is the parity of the layout patterns. In optical lithography, a mask is written with the metal side facing up and used with the metal side facing down, so the parity of actual patterns printed on a wafer through optical lithography is opposite to the design. As a result, when the layout design is taped out for e-beam lithography, its parity has to be changed so that at the later processes, optical lithography patterns can be aligned to the e-beam patterns.
5.2.2 Layout Design

In our layout, the device matrix has four different emitter sizes and for each emitter size, there are ten different passivation ledge lengths. Figure 5-11 shows the actual layout of the device matrix. In order to verify the alignment and monitor the process, we also put several test structures in our layout. The cross section SEM picture of the finished emitter and passivation ledges is shown in Figure 4-10d.

The layout for a typical DC HBT device is shown in Figure 5-12. The device shown has an emitter of $2 \times 10 \, \mu m^2$ and a passivation ledge of $0.3 \, \mu m$. The spacing between the passivation ledge and the base metal contact is always $3.0 \, \mu m$. It is worth pointing out that the passivation ledge length of $0.3 \, \mu m$ is measured along the direction perpendicular to the emitter finger. Along the other direction, i.e., parallel to the emitter finger, the passivation ledge is always $2.0 \, \mu m$ long. This becomes especially important when we are calculating the actual peripheral length of the emitter.
Figure 5-11 The device matrix in the layout. Each row contains ten HBTs with ten different passivation ledge lengths.
Figure 5-12 The layout for a single DC HBT device.
5.3 Experiment Results

DC current gain, defined as the ratio of the emitter to the base current, is the main parameter measured from the finished HBT devices. The measurement is done with a HP4145B parameter analyzer to forward bias the emitter-base junction from 0.0 to 2.0 volts and monitor the base and collector currents. From this measurement, not only the current gain, but also a Gummel Plot can be obtained for each measured device. Measurements were performed at both room temperature and elevated temperatures. The main experimental results and comparison with the simulations are presented in the following sections.

5.3.1 Current Gain

Figure 5-13 shows the current gain results. It shows current gain measured or simulated as a function of the passivation ledge length at a collector current density of $1 \times 10^4$ A/cm$^2$. The experimental data, represented by open symbols, include four emitter sizes - 2 x 10, 4 x 10, 8 x 10 and 12 x 10 $\mu$m$^2$. The simulated data, represented by solid symbols, are for a single emitter size of 2 x 10 $\mu$m$^2$.

In the figure, the solid line represents measured data and the two dotted lines are from simulations. The experimental results show the same trend as predicted by both simulation approaches and the experimental minimum ledge length for full passivation is about 0.3 $\mu$m. When there is no passivation ledge or the passivation ledge is shorter than 0.3 $\mu$m, the extrinsic base surface recombination current dominates the base current so that the agreement between the measured data and the analytical simulation is very good. When the extrinsic base has been completely passivated by 0.3 $\mu$m or longer ledges, the measured data are smaller than the prediction of the analytical simulation. This is because the analytical simulation does not include emitter-base space-charge region recombination, thus over-estimating the current gain. Comparing the measured data to the Semi-Cad simulation, there is some discrepancy from using a model with too pessimistic mobility, thus the predicted current gains are relatively low.
Figure 5-13 Measured and simulated current gain as a function of passivation ledge length. The open symbols are for experimental results and the solid ones for simulation results. The lines are for an emitter size of 2 x 10 μm².

Consider the measured data for different emitter sizes. For devices which are not fully passivated, i.e., those devices in which the passivation ledge is shorter than 0.3 μm, HBTs with a smaller emitter size suffer more from extrinsic base surface recombination and have a lower current gain. When the passivation ledge becomes equal to or longer than 0.3 μm, the difference in current gain for all the emitter sizes vanishes. This observation verifies the concept as well as the value of the minimum ledge length for full passivation.

5.3.2 Emitter Size Effect

Another way to examine the difference in current gain for different emitter sizes is to look at the emitter size effect. For a bipolar transistor, the DC current gain can be expressed as [5-4]

\[
\frac{I_B}{I_C} = \frac{I_{D}^{surface} + I_{D}^{bulk}}{I_C} = \frac{C_1 \cdot P + C_2 \cdot A}{C_3 \cdot A} = \left(\frac{C_1}{C_3}\right) \frac{P}{A} + \left(\frac{C_2}{C_3}\right).\tag{5.13}
\]
where P and A are the perimeter and the area of the emitter, respectively. As shown by the formula, the base current contains two portions, one of which is proportional to the perimeter and the other to the area. The collector current is always proportional to the emitter area, provided that emitter current crowding is not a concern. Therefore, if we plot 1/B as a function of the perimeter-area ratio, the slope represents the strength of the surface recombination current. If the slope becomes zero, then the device is fully passivated.

Figure 5-14 shows the emitter size effects for HBTs with different passivation ledge lengths. The HBTs without passivation ledges have a large slope in the 1/B versus P/A plot, which means the current gain of these devices are dominated by the surface recombination current. For the HBTs with passivation ledges, as shown in Figure 5-14a, even with passivation ledges as short as 0.1 μm, the current gain is greatly improved and the slope becomes almost flat. Figure 5-14b, which is an enlargement of the lower portion in plot, shows that HBTs with passivation ledges smaller than 0.3 μm still have a detectable emitter size effect. Only after the passivation ledge length is larger than the minimum ledge length for full passivation does the emitter size effect completely disappear.
5.3.3 Gummel Plots and Ideality Factors

Figure 5-15 shows the measured Gummel plots of the HBTs with zero or 0.3 µm passivation ledges. For both devices, the ideality factors of the collector current $I_c$ are close to unity. However, the ideality factors of the base current, $I_b$, are quite different for these two devices. The unpassivated device has an ideality factor of 1.26, which is closer to unity than the 1.60 - value of the ideality factor from the passivated device. This indicates that the ideality factor for surface recombination current component is close to 1 instead of 2. [5-5] Based on the discussion in Section 3.1.5, this observation means that the non-equilibrium assumption should be used in modeling the surface recombination process.

![Gummel plots](image)

Figure 5-15 Measured Gummel plots of HBTs with zero or 0.3 µm passivation ledges.
Figure 5-16 Measured current gain as a function of the collector current density for HBTs with zero or 0.3 μm passivation ledges.

5.3.4 Base Pushout Effect

Figure 5-16 shows the measured current gain as a function of the collector current density for the HBTs with zero, 0.1 and 0.3 μm passivation ledges. The interesting feature in this plot is the rolling-down of the current gain as the collector current density becomes higher than 9 × 10^4 A/cm². This is probably due to the base pushout effect or Kirk effect. [5-6] When the device is operating under high current injection, the concentration of the injected electrons in the base-collector space charge region can be expressed as

\[ n_{injected} = \frac{J_c}{q \nu_{sat}} \]  \hspace{1cm} (5.14)

where \( J_c \) is the collector current density and \( \nu_{sat} \) is the electron saturation velocity. If \( n_{injected} \) is higher than the collector dopant concentration, it neutralizes the base-collector space-charge region so that the effective neutral base becomes wider. With a wider base, the base transit time increases and current gain drops. In our case, the nominal collector dopant concentration is 2 × 10^{16} cm⁻³, which allows a maximum collector current density of 6.4 × 10^4 A/cm² if the saturation velocity is assumed to be 2 × 10^7 cm/sec. Because the
accuracy of the doping concentration is relatively poor, the actual collector dopant concentration may be 40% higher, which yields a value where the current gain starts to drop of \( J_c = 9 \times 10^4 \text{ A/cm}^2 \).

5.3.5 Temperature Dependence

![Graph showing current gain vs. temperature for different ledge lengths](image)

Figure 5-17 Measured current gain at room and elevated temperatures for HBTs with different passivation ledge lengths. These HBTs have an emitter size of 2 x 10 \( \mu \text{m}^2 \).

Figure 5-17 shows the measured current gain at room and elevated temperatures. In general, the current gain drops as temperature increases. This can be explained by the following expression: [5-7]

\[
\beta = \beta_0 \exp \left( \frac{1}{n_c} - \frac{1}{n_b} \right) \frac{qV}{kT}. \tag{5.15}
\]

where \( n_c \) and \( n_b \) are the ideality factors of the collector current and base current, respectively. Because \( n_c \) is very close to 1 and \( n_b \) is a number between 1 and 2, the positive exponential term makes the current gain decrease as temperature increases. For HBTs without passivation ledges, the difference between \( n_c \) and \( n_b \) is smaller, so the temperature dependence is weaker for these devices. For HBTs that have been fully
passivated, the temperature dependence is stronger. Although weak temperature dependence is usually desired for circuit design, passivation is essential for achieving high current gain and good reliability. One solution is to improve the quality of emitter-base space-charge region so that $n_b$ will become close to 1 even when HBTs are fully passivated.

![Graph](image)

**Figure 5-18** Current gain as a function of the passivation ledge length measured at room and elevated temperatures. These HBTs have an emitter size of $2 \times 10^6 \text{ m}^2$.

Figure 5-18 shows the current gain as a function of passivation ledge length measured at different temperatures. Although the current gain is lower at higher temperatures, the minimum ledge length for full passivation is about 0.3 m for all temperatures. This observation seems counter intuitive - the minimum ledge length for full passivation should become longer at higher temperature due to the faster diffusion of electrons. However, in the $5 \times 10^{18} \text{ cm}^{-3}$ doped base region, the diffusion length of electrons is about 2.1 m, and the Semi-Cad simulation shows that the decrease in the electron concentration away from the emitter edge is much faster than $\exp\left(-\frac{x}{L_D}\right)$, where $L_D$ is the diffusion length. Therefore, at elevated temperatures, the larger diffusion length does not spread the electrons noticeably wider. This observation makes the concept and
value of the minimum ledge length for full passivation an important and valuable design criteria for HBTs.
Chapter 6
RF HBT Performance with and without Sub-Micron Passivation Ledges

6.1 Design of Experiments

In chapter 5, both simulation and experimental results of current gain of DC HBTs demonstrated that a sub-micron ledge could fully passivate a HBT. This conclusion indicates that RF HBT performance could similarly be optimized by utilizing passivation ledges as short as 0.3 μm. In order to verify that there is no difference in RF performance for HBTs with and without sub-micron passivation ledges if a proper layout design is used, a group of RF HBTs were designed and fabricated. The epitaxial structure of these RF HBTs were described Table 3-3 and discussed in Section 3.6.2. Section 6.1.1 describes the overall process flow used to fabricate RF HBTs with and without sub-micron passivation ledges. Section 6.1.2 presents the layout used for fabrication as well as high frequency measurements.

6.1.1 Process Integration

Among the four processing approaches that can be used to form sub-micron passivation ledges, the Self-Aligned Undercut-Etch and Base Lift-off (SAUBL) approach was used to fabricate these RF HBTs. Figure 6-1 shows the process flow with the split for devices with and without the passivation ledge. At the beginning of the processing, the MBE wafer was cleaved into two half-wafers. These two half-wafers went through identical processes before the split, then one half-wafer was processed by optical lithography and the self-aligned lift-off process to form HBTs without passivation, while the other half-wafer was processed using e-beam lithography and the SAUBL process to form HBTs with the passivation ledge. After the base metalization step, the two half wafers were brought back to identical processes to finish the rest of the fabrication as shown in Figure 6-1.
Figure 6-1 The process flow with split used to fabricate RF HBTs with and without sub-micron passivation ledges. The two schematic drawings of HBT cross sections show the difference in device structures processed in the two processing sequences.
6.1.2 Layout Design

Figure 6-2 shows the layout of RF HBTs with and without passivation ledges. The layout is identical for these two types of HBTs except at the base metal level. For a passivated HBT, shown in Figure 6-2(a), the base fingers do not overlap with the emitter finger and the gap defines the sum of the passivation ledge length and the ledge-to-base-metal spacing. For a HBT without ledges, shown in Figure 6-2(b) the base finger covers the emitter finger so that self-aligned base metal lift-off can be realized during the process.

Figure 6-3 shows the layout of RF HBTs with Ground-Signal-Ground transmission lines designed for RF probing. Figure 6-3 also shows the open circuit and short circuit structures placed on the die in order to measure and be able to strip the pad parasitics. By the side of each structure, an equivalent circuit has been drawn to show the connections that can not be seen directly from the layout. Section 6.2.2 discusses how to use the open and short circuit structures to extract the intrinsic HBT performance from the measurements which include parasitic elements.
Figure 6-2 Layout of RF HBTs with and without passivation ledges.
Figure 6-3 Layout and equivalent circuits of the transistor and open and short circuit structures connected to the Ground-Signal-Ground transmission lines for RF probing.
6.2 High Frequency Measurement

6.2.1 Measurement Set-up and Procedure

In order to characterize the high frequency performance of HBTs, a network analyzer was used to measure the S-parameters of the devices under common emitter configuration. Figure 6-4 shows the measurement set-up that includes an HP8720B network analyzer to sweep the signal frequency and measure S-parameters, an HP4145B parameter analyzer to bias the devices and the measurements were controlled and data acquired and stored with an HP9000 computer. The devices were mounted in a high precision probe station using Cascade Microtech probes.

![Diagram](image.png)

Figure 6-4 The S-parameter measurement set-up.
Figure 6-5 Calibration procedure for S-parameter measurements.

At the beginning of the measurement, the set-up must be calibrated using a calibration standard kit. Figure 6-5 shows the calibration procedure called SLOT (Short-Load-Open-Through). Figure 6-5 also shows the shape of the calibration structures on the standard substrate and the nominal impedance values of short, load and open, i.e., 0, 50 and $\infty$ ohms, on Smith Chart. After this calibration, all the parasitics up to the tips of the Cascade probes have been automatically stripped by the network analyzer.

During the S-parameter measurement, the emitter-base bias is automatically set by the HP4145B to yield a designated collector current level. The network analyzer then sweeps the signal frequency from 500 MHz to 20 GHz and measures the four S-parameter components at each frequency point. Each HBT and the open and short circuit structures are measured under the same connection so that the parasitics of the Ground-Signal-Ground (GSG) transmission lines can be stripped after data acquisition.
6.2.2 Data Processing and Performance Extraction

\[ S_{ij} \text{ transistor + pads + leads} \]
\[ \downarrow \text{convert} \]
\[ Y_{ij} \text{ transistor + pads + leads} \]
\[ \downarrow \text{strip} \]
\[ Y_{ij} \text{ pads} \]

\[ Y_{ij} \text{ transistor + leads} \]
\[ \downarrow \text{convert} \]
\[ Z_{ij} \text{ transistor + leads} \]
\[ \downarrow \text{strip} \]
\[ Z_{ij} \text{ leads} \]

\[ Z_{ij} \text{ transistor} \]
\[ \downarrow \text{convert} \]
\[ S_{ij} \text{ transistor} \]
\[ \rightarrow h_{ij} \text{ transistor} \]
\[ \rightarrow f_t \]

\[ \text{Power Gain} \quad \rightarrow f_{\text{max}} \]

Figure 6-6 The data processing and parameter extraction procedure.

The SLOT calibration kit provides parasitic extraction only to the tips of the Cascade probes. Thus parasitic capacitances and resistances of the GSG transmission lines on the HBT substrates are still included in the measured data. Figure 6-6 shows the procedure used to process the data and extract the high frequency parameters of HBTs.

The S-parameters of the open and short circuit structures are measured first. Then values for \( Y_{ij}^{\text{pads}} \) are obtained by converting the open circuit structure S-parameters to Y-parameters, and values for \( Z_{ij}^{\text{leads}} \) are obtained by converting the short circuit structure S-parameters to Z-parameters. For transistors, under each bias condition and signal frequency, the S-parameters are converted to Y-parameters. The pad capacitances are then stripped away by subtracting the values for \( Y_{ij}^{\text{pads}} \) from the total Y-parameters. After stripping the pad capacitances, the Y-parameters are converted to Z-parameters, and the lead resistances are stripped away by subtracting the values of \( Z_{ij}^{\text{leads}} \) from the total Z-parameters. After stripping both the pad capacitances and lead resistances, the pure
transistor parameters and high frequency performance can be represented by any of these two-port parameters—Z, Y, S, or H. [6-2] In order to determine the cut-off frequency, the measured values of $h_{21}$ are plotted as a function of frequency and extrapolated to unity current gain to obtain a value of $f_t$. In order to determine the maximum oscillation frequency, the measured values of unilateral power gain are plotted as a function of frequency to extrapolate the value of $f_{\text{max}}$.

6.3 Experimental Results

6.3.1 $f_t$ and $f_{\text{max}}$

Figure 6-7 shows the measured current gain and unilateral power gain for the highest values of $f_t$ and $f_{\text{max}}$ for HBTs without and with passivation ledges. The measurements were done from 0.5 GHz to 20 GHz. The extrapolation of $f_t$ and $f_{\text{max}}$ is done by using a 20 dB per decade slope for $h_{21}$ and $U$ from values measured at 10 GHz. This extrapolation is conservative because the extrapolated values of $f_t$ and $f_{\text{max}}$ would be larger if the extrapolation were taken from the measured values at 20 GHz.
Figure 6-7 The highest measured values of $f_t$ and $f_{\text{max}}$ for HBTs without and with passivation ledges. (a) without passivation ledges. (b) with 0.3 μm passivation ledges.
For HBTs without passivation ledges, the highest measured values of $f_t$ is 59.4 GHz and $f_{\text{max}}$ is 68.1 GHz at a collector current of 7 mA. For the HBTs with 0.3 µm passivation ledges, the highest measured values of $f_t$ is 62.9 GHz and $f_{\text{max}}$ is 63.5 GHz at a collector current of 16 mA. In terms of the highest $f_t$ and $f_{\text{max}}$, the HBTs without and with passivation ledges have nearly identical performance, as predicted by the design. This observation indicates that the use of sub-micron passivation ledges can be optimized so that the high frequency performances will not be sacrificed. The reason that the $f_{\text{max}}$ of the HBT with 0.3 µm ledges is slightly lower than that of the HBT without passivation ledges is probably because the base resistance is higher for the passivated HBTs, which have a narrower base contact metal and a longer base contact to emitter distance.

In terms of the collector current where the HBTs without and with passivation ledges achieve their highest values of $f_t$ and $f_{\text{max}}$, there is a significant difference. Figure 6-8 shows the measured dependence of $f_t$ and $f_{\text{max}}$ on collector current for HBTs without and with passivation ledges. The difference is probably due to the emitter mesa size. Although for both types of HBTs, the emitter mesa sizes are designed to be 2 x 10 µm$^2$, the actual emitter mesa size of the unpassivated HBTs is smaller than that of the passivated HBTs because the undercut etch is a far more serious problem when the transistors are not passivated. If we assume the undercut is about 0.5 µm on each side, then the two types of HBTs achieve their maximum values of $f_t$ and $f_{\text{max}}$ at about the same collector current density - 8 x 10$^4$ A/cm$^2$. Comparing Figure 5-16 and Figure 6-8, we find that the roll over of the current gain and $f_t$ starts at similar values of collector current density. This is because the base pushout effect decreases not only the current gain but also $f_t$ by increasing the base transit time. The value of $f_{\text{max}}$ is proportional to the square root of $f_t$, thus it also starts to decline at the same point that the injected electrons trigger the base pushout effect.
Figure 6-8 Measured dependence of $f_t$ and $f_{\text{max}}$ on collector current for HBTs without and with passivation ledges. (a) without passivation ledges. (b) with 0.3 $\mu$m passivation ledges.
6.3.2 Discussion

As shown by our experimental results, HBTs with sub-micron passivation ledges can achieve comparable high frequency performance to that of traditional RF HBTs fabricated by a self-aligned processes. These sub-micron passivation ledges will solve some of the reliability problems caused by extrinsic base surface recombination. However, the collector current to achieve the maximum values of $f_t$ and $f_{max}$ is still quite high. A collector current of 15 mA means 75 mW dissipation for each transistor and 75 W for a circuit with 1000 transistors when biased at 5V. This high power consumption will not only limit the integration density but also cause reliability problems. Submicron emitter size HBTs will be absolutely essential to overcome this power limitation and they also help reduce the self-heating related reliability problems.
Chapter 7
Conclusions and Future Work

7.1 Conclusions

The research described in this dissertation has concentrated on the scaling issues for AlGaAs/GaAs Hetero-junction Bipolar Transistors (HBTs) and the minimum passivation ledge lengths required to fully passivate the exposed surface regions. As HBTs are scaled down, the processing approaches to fabricate sub-micron passivation ledges become quite different and require considerable engineering to produce reliable and reproducible devices. The optimization in HBT device epitaxial layer design, layout design, fabrication, and scaling rules for HBT miniaturization were developed in this thesis research.

Both simulation and experimental results show that the minimum passivation ledge lengths required to fully passivate AlGaAs/GaAs HBTs are about three times the base thickness when the base layer is not graded. For example, for HBTs with a base thickness of 1000Å, the minimum ledge length to produce full passivation is 0.3 μm. The measurement results also show that this critical ledge length is not sensitive to temperature because the lateral spreading of electrons in base region is limited by the base thickness rather than the electron diffusion length. This observation makes the concept and value of the minimum ledge length for full passivation an important and valuable design criteria for HBTs. When the base is graded, the minimum ledge length for full passivation decreases from 0.3 μm to less than 0.1 μm depending on the electric field created by grading the base layer.

Three processing approaches were developed to fabricate these sub-micron passivation ledges: (1) Electron-Beam Lithography approach, (2) Self-Aligned Undercut Etch/Base Lift-off approach, and (3) Selective Undercut Etching approach. The Selective Undercut Etching approach proved to be the best way to realize sub-micron HBTs with low emitter resistance, high current gain and good reliability. The Self-Aligned Undercut Etch/Base Lift-off approach is suitable for manufacturing HBTs with larger emitter sizes. The Electron-Beam Lithography approach is a valuable research tool and is the best approach to test the fabrication limits because of its very precise dimension control; however, it is unlikely to be used in HBT manufacturing because of its low process throughput and high cost.
The Self-Aligned Undercut Etch/Base Lift-off approach was used to fabricate specially designed RF HBTs with sub-micron passivation ledges. For comparison, RF HBTs without passivation ledges were also fabricated using the standard Self-Aligned processes. The experimental results show that HBTs with sub-micron passivation ledges can achieve comparable high frequency performance to that of traditional RF HBTs fabricated by the self-aligned processes, but with higher current gain and greater reliability. For the HBTs with 0.3 µm ledges, the highest measured value of \( f_t \) is 62.9 GHz and \( f_{\text{max}} \) is 63.5 GHz. For HBTs without passivation ledges, the highest measured value of \( f_t \) is 59.4 GHz and \( f_{\text{max}} \) is 68.1 GHz.

The turn-on voltage of HBTs is determined by the band-gap of the base material, which does not scale with device size, but only if a different material system is used. Thus the miniaturization of HBTs is mainly in lateral dimensions. Because the emitter contact resistance increases faster than the base-emitter junction capacitance decreases during the scaling and some of the other parasitic components, such as the base contact resistance, do not scale at all, the high frequency performance of HBTs decreases as HBTs are scaled down to sub-micron dimensions. In order to maximize \( f_t \) and \( f_{\text{max}} \), the emitter contact resistance must be reduced. The Selective Undercut Etching approach developed in this dissertation research provides an approach which minimizes this problem.

7.2 Future Work

Based upon the results of this dissertation research, several interesting areas are worthy of further investigation.

- The emitter contact resistance is identified to be an important parasitic in limiting the high frequency performance and scalability of HBTs. This dissertation research also shows that the specific contact resistances are very sensitive to surface conditions. A fabrication procedure in which the emitter epitaxial surface is protected, thus reducing the emitter contact resistance, will be very useful. For example, if the emitter metal can be placed on top of the emitter epitaxial surface before the isolation implant, the surface degradation caused by the oxide deposition and removal will be eliminated.
Further efforts to directly measure the mean time to failure (MTTF) of HBTs with sub-micron passivation ledges are worthwhile. For this investigation, Carbon doped HBTs should be used because Beryllium diffusion may dominate the measured MTTF if the base layer is Beryllium doped. Second, the fabrication should not use H+ implant for isolation because hydrogen redistribution and neutralization of p-type doping may affect the reliability of these HBTs.

Minimum passivation ledge lengths of HBTs using other material systems, such as InAlGa/InGaAs HBTs grown on InP substrates, should be investigated. Because the lateral spreading of electrons and surface recombination velocity in InGaAs are different, the design rules for minimum passivation ledge lengths may change.

Pnp HBTs are important for complementary circuits. The concept of minimum ledge length for full passivation is certainly applicable to these devices. However, because holes, the minority carriers inside the base region of Pnp HBTs, have different transport properties, the values of the minimum ledge length for full passivation must be determined by experimental investigations.

Improving the material quality around the emitter-base junction is very important, especially for HBTs that are fully passivated by sub-micron passivation ledges. The improvement in carrier lifetime within the emitter-base junction will not only increase current gain but also make current gain less collector current or temperature dependent.

These advances would increase the performance, reliability and applications of HBTs.
Chapter 8

References

Chapter 1 References


[1-2] Chenming Hu, “MOSFET Scaling in the Next Decade and Beyond”,


[1-4] William Liu, Damian Costa and James Harris, Jr., “Comparison of the Effects of
      Surface Passivation and Base Quasi-Electric Field on the Current Gain of
      AlGaAs/GaAs Heterojunction Bipolar Transistors Grown on GaAs and Si

Damian Costa and James Harris, Jr., “Low-Frequency Noise Properties of N-p-n
AlGaAs/GaAs Heterojunction Bipolar Transistors”, IEEE Trans. Electr. Dev.,

Lovell H. Cmnitz, “Improved Current Gain Stability in C-doped HBTs with
AlGaAs Surface Passivation Layer”, GaAs REL Workshop, 1991.

Chapter 2 References

      During Molecular-Beam Epitaxy, Ph.D Dissertation, Stanford University, Dec.
      1994, pp. 17.

      During Molecular-Beam Epitaxy, Ph.D Dissertation, Stanford University, Dec.
      1994, pp. 23.


[2-22] Same as the 2nd reference in [1-4].


Chapter 3 References


[3-8] Discussion with Dr. Thomas Low of Hewlett-Packard Company.


[3-10] Same as the reference [3-3].


[3-14] Same as the 2nd and 3rd references in [1-4].


[3-17] Same as the reference [2-8].


[3-28] Same as the reference [3-15].


[3-32] Same as the 3rd reference in [1-4].


[3-35] Same as the reference [3-30].

[3-37] Same as the 3rd reference in [1-4].

[3-38] Same as the reference [3-15].


[3-40] Same as the reference [3-15].


Chapter 4 References


[4-4] Same as the reference [1-3].

[4-5] Same as the 1st reference in [1-4].

[4-6] Same as the 2nd reference in [1-4].

[4-7] Same as the 3rd reference in [1-4].


Chapter 5 Reference


[5-5] Same as the reference [3-13].

[5-6] Same as the reference [3-21].


Appendix A

The Processing Flow for DC HBTs with Passivation Ledges

Note: 3:1:5 means H₃PO₄ : H₂O₂ : H₂O = 3 : 1 : 50, @20°C.
1:200 means NH₄OH : H₂O₂ = 1 : 200, @10°C.
S1813 means Shipley photoresist 1813.
'. ' and "." mean minute(s) and second(s), respectively.
* means alternative approaches.

normal_ledge_split produces passivation ledges larger than 2 μm.
submicron_ledge_split produces passivation ledges of submicron sizes.
except the marked ones, processing steps apply to both splits.

1A. Wafer degrease and clean-up
   a. TCA 5' / aceton 5' / methanol 5' / IPA 5' / DI-water 5'
   b. BOE 6:1 dip 1' / DI-water 3' / blow dry

1B. Emitter metal photolithography (EMIT Mask) normal_ledge_split
   a. singe 20' @150°C / cool 10' @27°C
   b. spin S1813 at 3500 rpm for 30"
   c. bake 10' @80°C / clean the extra resist at corners / bake 10' @80°C
   d. soak in C₆H₅Cl for 10' / bake 5' @80°C
   e. expose 165 mJ/cm² / 1:1 developer for 1' / DI-water 3' / blow dry

1B* Emitter metal e-beam lithography (LEVEL-1 File) submicron_ledge_split
   a. singe 30' @150°C / cool 10' @27°C
   b. spin PMMA (5%, 496K) at 4000 rpm for 30"
   c. bake 24 hours @110°C
   d. expose to e-beam 275 μC/cm²
   e. 1:3 (MIBK : IPA) developer for 40" / IPA 2' / blow dry

1C. Emitter metal evaporation and lift-off
   a. evaporate emitter metal Ti/Pt/Au = 400/200/1000Å aligned and no rotation
   b. acetone 5' / acetone 5' / soak in acetone 24 hours
   c. acetone 5' / methonal 5' / blow dry

1D. Emitter mesa etch
a. calibrate the etch rates of 3:1:50 and 1:200  
b.  3:1:50 etch 1'30" / DI-water 3'  
c.  1:200 etch 1' / DI-water 3'  
d.  dektek the total etched thickness

2A.  Passivation ledge photolithography (PASS Mask)  normal_ledge_split
  a.  singe 20' @150°C / cool 10' @27°C  
b.  spin S1813 at 5000 rpm for 30"  
c.  bake 10' @80°C / clean the extra resist at corners / bake 10' @80°C  
d.  expose 137 mJ/cm² / 1:1 developer for 1' / DI-water 3' / blow dry  
e.  bake 10' @80°C

2A*  Passivation ledge e-beam lithography (LEVEL-2 File)  submicron_ledge_split
  a.  singe 20' @150°C / cool 10' @27°C  
b.  spin SAL 601 at 4000 rpm for 30"  
c.  bake 30' @80°C  
d.  expose to e-beam 9 µC/cm²  
e.  hot-plate-baking 75" @110°C  
f.  concentrated Shipley developer for 3' / DI-water 3' / blow dry

2B.  Passivation ledge etch
  a.  calibrate the etch rates of 3:1:50  
b.  3:1:50 etch 55" / DI-water 3' / blow dry

2C.  Photoresist removal  normal_ledge_split
  a.  acetone 5' / acetone 5' / methanal 5' / blow dry  
b.  dektek the passivation ledge thickness

2C*  SAL 601 e-beam resist removal  submicron_ledge_split
  a.  O₂ plasma etch for 4' (O₂ 20 sccm, base pressure 100 mT, 250 Watts)  
b.  dektek the passivation ledge thickness

3A.  Base metal photolithography (BASE Mask)
  a.  singe 20' @150°C / cool 10' @27°C  
b.  spin S1813 at 3500 rpm for 30"  
c.  bake 10' @80°C / clean the extra resist at corners / bake 10' @80°C
d. soak in C₆H₅Cl for 10' / bake 5' @80°C

3B. Base metal evaporation and lift-off
   a. evaporate emitter metal Ti/Pt/Au = 400/200/2000Å rotation
   b. acetone 5' / acetone 5' / soak in acetone 24 hours
   c. acetone 5' / methanol 5' / blow dry

4A. Base mesa photolithography (MESA Mask)
   a. singe 20' @150°C / cool 10' @27°C
   b. spin S1813 at 5000 rpm for 30"
   c. bake 10' @80°C / clean the extra resist at corners / bake 10' @80°C
   d. expose 137 mJ/cm² / 1:1 developer for 1' / DI-water 3' / blow dry
   e. bake 10' @80°C

4B. Base mesa etch
   a. calibrate the etch rates of 3:1:50
   b. 3:1:50 etch 4 / DI-water 3' / blow dry
   c. acetone 5' / acetone 5' / methanal 5' / blow dry
   d. dektek the base mesa thickness

4C. Backside collector contact
   a. TCA 5' / aceton 5' / methanol 5' / IPA 5' / DI-water 5'
   b. BOE 6:1 dip 1' / DI-water 3' / blow dry
   c. evaporate metal Au/Ge/Ni/Ti/Au = 680/170/150/1000/2000 Å rotation
   d. rapid thermal anneal 30" @430°C

5A. Dielectric deposition
   a. deposit 5000Å SiO₂

5B. VIA hole photolithography (VIAS Mask)
   a. singe 20' @150°C / cool 10' @27°C
   b. spin S1813 at 5000 rpm for 30"
   c. bake 10' @80°C / clean the extra resist at corners / bake 10' @80°C
   d. expose 85.75 mJ/cm² / 1:1 developer for 1' / DI-water 3' / blow dry
   e. bake 10' @80°C
5C. VIA hole etch
   a. plasma etch for 25' (CHF₃ 100 sccm, O₂ 10 sccm, 100 mT, 100Watts)
   b. O₂ plasma ash for 1' (O₂ 20 sccm, base pressure 200 mT, 50 Watts)

6A. Interconnect metal photolithography (INTE Mask)
   a. singe 20' @150°C / cool 10' @27°C
   b. spin S1813 at 3500 rpm for 30''
   c. bake 10' @80°C / clean the extra resist at corners / bake 10' @80°C
   d. soak in C₆H₅Cl for 10' / bake 5' @80°C
   e. expose 87.5 mJ/cm² / 1:1 developer for 1' / DI-water 3' / blow dry

6B. Interconnect metal evaporation and lift-off
   a. evaporate emitter metal Ti/Au = 1000/9000Å rotation
   b. acetone 5' / acetone 5' / soak in acetone 24 hours
   c. acetone 5' / methonal 5' / blow dry
Appendix B

The Processing Flow for RF HBTs without and with Passivation Ledges

Note: 3:1:5 means H₃PO₄ : H₂O₂ : H₂O = 3 : 1 : 50, @20°C.
1:200 means NH₄OH : H₂O₂ = 1 : 200, @10°C.
S1813 means Shipley photoresist 1813.
' and " mean minute(s) and second(s), respectively.
* means alternative approaches

no_ledge_split produces RF HBTs without passivation ledges.
submicron_ledge_split produces RF HBTs with submicron passivation ledges.
except the marked ones, processing steps apply to both splits.

1A. Wafer degrease and clean-up
   a. TCA 5' / aceton 5' / methanol 5' / IPA 5' / DI-water 5'
   b. BOE 6:1 dip 1' / DI-water 3' / blow dry

1B. Dummy oxide deposition
   a. deposit 5000Å oxide

1C. Isolation photolithography (ISOL AC-Mask)
   a. singe 20' @150°C / cool 10' @27°C
   b. spin S1822 at 3500 rpm for 30'
   c. bake 10' @80°C / clean the extra resist at corners / bake 10' @80°C
   d. soak in C₆H₅Cl for 10' / bake 5' @80°C
   e. expose 165 mJ/cm² / 1:1 developer for 1' / DI-water 3' / blow dry

1D. Isolation metal evaporation and lift-off
   a. evaporation Al = 1.5 μm
   b. acetone 5' / acetone 5' / soak in acetone 24 hours
   c. acetone 5' / methonal 5' / blow dry

1E. Dummy oxide etch
   a. plasma etch for 30' (CHF₃ 100 sccm, O₂ 10 sccm, 100 mT, 100Watts)

1F. Isolation implant
   a. H+ 150 Kev dose 7 x 10¹⁴ cm⁻² beam 100mA tilt 7°C
b. H+ 100 Kev dose 5.9 x 10^{14} cm^{-2} beam 100mA \hspace{1cm} \text{tilt 7°C}  \\
c. O++ 400 Kev dose 8 x 10^{12} cm^{-2} beam 100mA \hspace{1cm} \text{tilt 7°C}  \\
d. O++ 300 Kev dose 8 x 10^{12} cm^{-2} beam 100mA \hspace{1cm} \text{tilt 7°C}  \\
e. O++ 240 Kev dose 8 x 10^{12} cm^{-2} beam 100mA \hspace{1cm} \text{tilt 7°C}  \\
f. O++ 200 Kev dose 1.5 x 10^{13} cm^{-2} beam 100mA \hspace{1cm} \text{tilt 7°C}  \\
g. O++ 100 Kev dose 8 x 10^{12} cm^{-2} beam 100mA \hspace{1cm} \text{tilt 7°C}  \\
h. O++ 40 Kev dose 8 x 10^{12} cm^{-2} beam 100mA \hspace{1cm} \text{tilt 7°C}  \\

2A. Post-implant clean-up  
   a. TCA 5' / acetone 5' / methanol 5' / IPA 5' / DI-water 5'

2B. Etch window photolithography (WINE AC-Mask)  
   a. singe 20' @150°C / cool 10' @27°C  
   b. spin S1813 at 5000 rpm for 30''  
   c. bake 10' @80°C / clean the extra resist at corners / bake 10' @80°C  
   d. expose 137 mJ/cm² / 1:1 developer for 1' / DI-water 3' / blow dry  
   e. bake 10' @80°C

2C. Alignment mark window etch  
   a. calibrate the etch rates of 3:1:50  
   b. 3:1:50 etch 5' / DI-water 3' / blow dry

2D. Photoresist removal  
   a. acetone 5' / acetone 5' / methanol 5' / blow dry

2E. Implant mask removal  
   a. 1:1 (H₃PO₄ : H₂O) etch 10' @70°C / DI-water 3' / blow dry  
   b. BOE 6:1 etch 4' / DI-water 3' / blow dry

3A. Emitter metal W deposition  
   a. sputter metal W = 2500Å (1500 watts, 360 V, Ar pressure 5 mbar, RTA 2000)

3B. Emitter metal Ti/Al photolithography (EMIT AC-Mask)  
   a. singe 20' @150°C / cool 10' @27°C  
   b. spin S1813 at 3500 rpm for 30''  
   c. bake 10' @80°C / clean the extra resist at corners / bake 10' @80°C
d. soak in C₆H₅Cl for 10' / bake 5' @80°C  
e. expose 165 mJ/cm² / 1:1 developer for 1' / DI-water 3' / blow dry

3C. Pre-deposition DESCUM and acid dip  
   a. O₂ plasma DESCUM for 20" (O₂ 20 sccm, base pressure 100 mT, 50 Watts)  
   b. 1:1 (H₃PO₄ : H₂O) dip 15" / DI-water 3' / blow dry

3D. Emitter metal Ti/Al evaporation and lift-off  
   a. evaporate emitter metal Ti/Al = 700/1000Å aligned and no rotation  
   b. acetone 5' / acetone 5' / soak in acetone 24 hours  
   c. acetone 5' / methonal 5' / blow dry

3E. W etch  
   a. dry etch W 30' (C₂F₆ 100 sccm, O₂ 10 sccm, 100 mT, 300 Watts)  
   b. O₂ plasma ash for 10' (O₂ 20 sccm, base pressure 200 mT, 50 Watts)

3F. Emitter contact resistance measurement  
   a. measure emitter contact resistance using Transmission Lines  
   b. TCA 5' / acetone 5' / methanol 5' / IPA 5' / DI-water 5'

3G. Emitter mesa etch  
   a. calibrate the etch rates of 3:1:50 and 1:200  
   b. 3:1:50 etch 1'30" / DI-water 3'  
   c. 1:200 etch 1' / DI-water 3'  
   d. dektek the total etched thickness

4A. Nitride deposition  no_ledge_split  
   a. deposit 1200Å Si₃N₄

4A* Oxide deposition  submicron_ledge_split  
   a. deposit 2400Å SiO₂

4B. RIE nitride  no_ledge_split  
   a. RIE nitride 2'15" (C₂F₆ 20 sccm, 44 mT, 200Watts)

4B* Base metal electron-beam lithography (BASE AC-File)  submicron_ledge_split
a. singe 30' @150°C / cool 10' @27°C  
b. spin PMMA (5%, 496K) at 4000 rpm for 30''  
c. bake 24 hours @110°C  
d. expose to e-beam 275 µC/cm²  
e. 1:3 (MIBK : IPA) developer for 40'' / IPA 2' / blow dry  

4C. Etch down to base *no_ledge_split*  
a. calibrate the etch rates of 3:1:50  
b. 3:1:50 etch 55'' / DI-water 3' / blow dry  

4C* Oxide undercut etch *submicron_ledge_split*  
a. wet the wafer surface in the diluted "Dry Well" (20 drops in 1000 ml) for 10''  
b. 6:1 BOE etch 1'30'' / DI-water 3' / blow dry  

4D. Base metal photolithography (BASE AC-Mask) *no_ledge_split*  
a. singe 20' @150°C / cool 10' @27°C  
b. spin S1813 at 3500 rpm for 30''  
c. bake 10' @80°C / clean the extra resist at corners / bake 10' @80°C  
d. soak in C₆H₅Cl for 10' / bake 5' @80°C  
e. expose 165 mJ/cm² / 1:1 developer for 1' / DI-water 3' / blow dry  

4D* Etch down to base *submicron_ledge_split*  
a. calibrate the etch rates of 3:1:50  
b. wet the wafer surface in the diluted "Dry Well" (20 drops in 1000 ml) for 10''  
c. 3:1:50 etch 55'' / DI-water 3' / blow dry  

4E. Pre-deposition DESCUM and acid dip *no_ledge_split*  
a. O₂ plasma DESCUM for 20'' (O₂ 20 sccm, base pressure 100 mT, 50 Watts)  
b. 1:1 (H₃PO₄ : H₂O) dip 15'' / DI-water 3' / blow dry  

4F. Base metal evaporation and lift-off  
a. evaporate emitter metal Ti/Pt/Au = 280/200/620Å aligned and no rotation  
b. acetone 5' / acetone 5' / soak in acetone 24 hours  
c. acetone 5' / methonal 5' / blow dry  

5A. Base mesa photolithography (MESA AC-Mask)
a. singe 20' @150°C / cool 10' @27°C
b. spin S1813 at 5000 rpm for 30"
c. bake 10' @80°C / clean the extra resist at corners / bake 10' @80°C
d. expose 137 mJ/cm² / 1:1 developer for 1' / DI-water 3' / blow dry
e. bake 10' @80°C

5B. Base mesa etch
   a. calibrate the etch rates of 3:1:50
   b. 3:1:50 etch 1'40" / DI-water 3' / blow dry
   c. acetone 5' / acetone 5' / methanol 5' / blow dry
d. dektelk the base mesa thickness

6A. Collector metal photolithography (COLL AC-Mask)
a. singe 20' @150°C / cool 10' @27°C
b. spin S1813 at 3500 rpm for 30"
c. bake 10' @80°C / clean the extra resist at corners / bake 10' @80°C
d. soak in C₆H₅Cl for 10' / bake 5' @80°C
e. expose 165 mJ/cm² / 1:1 developer for 1' / DI-water 3' / blow dry

6B. DESCUM and etch into subcollector
   a. O₂ plasma DESCUM for 20" (O₂ 20 sccm, base pressure 100 mT, 50 Watts)
b. calibrate the etch rates of 3:1:50
c. wet the wafer surface in the diluted "Dry Well" (20 drops in 1000 ml) for 10"
d. 3:1:50 etch 4'30" / DI-water 3' / blow dry

6C. Collector metal evaporation and lift-off
   a. evaporate metal AuGe/Ni/Au = 272/68/1000 Å aligned and no rotation
   b. acetone 5' / acetone 5' / soak in acetone 24 hours
c. acetone 5' / methonal 5' / blow dry
d. O₂ plasma ash for 10' (O₂ 20 sccm, base pressure 200 mT, 50 Watts)
e. rapid thermal anneal 5" @420°C

7A. Dielectric deposition
   a. deposit 5000Å SiO₂

7B. VIA hole photolithography (VIAS AC-Mask)

173
7C. VIA hole etch
   a. plasma etch for 25' (CHF \textsubscript{3} 100 sccm, O\textsubscript{2} 10 sccm, 100 mT, 100 Watts)
   b. O\textsubscript{2} plasma ash for 1' (O\textsubscript{2} 20 sccm, base pressure 200 mT, 50 Watts)

8A. Interconnect metal photolithography (INTE AC-Mask)
   a. singe 20' @150°C / cool 10' @27°C
   b. spin S1813 at 3500 rpm for 30''
   c. bake 10' @80°C / clean the extra resist at corners / bake 10' @80°C
   d. soak in C\textsubscript{6}H\textsubscript{5}Cl for 10' / bake 5' @80°C
   e. expose 87.5 mJ/cm\textsuperscript{2} / 1:1 developer for 1' / DI-water 3' / blow dry

8B. Pre-deposition DESCUM and acid dip
   a. O\textsubscript{2} plasma DESCUM for 20'' (O\textsubscript{2} 20 sccm, base pressure 100 mT, 50 Watts)
   b. 1:1 (H\textsubscript{3}PO\textsubscript{4} : H\textsubscript{2}O) dip 15'' / DI-water 3' / blow dry

8C. Interconnect metal evaporation and lift-off
   a. evaporate emitter metal Ti/Au = 1000/9000Å rotation
   b. acetone 5'/ acetone 5'/ soak in acetone 24 hours
   c. acetone 5'/ methonal 5' / blow dry