INFORMATION TO USERS

The most advanced technology has been used to photograph and reproduce this manuscript 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.
P-N-P heterojunction bipolar transistors in AlGaAs/InGaAs/GaAs

Hill, Darrell Glenn, Ph.D.
Stanford University, 1990

Copyright ©1990 by Hill, Darrell Glenn. All rights reserved.
P-N-P HETEROJUNCTION BIPOLAR TRANSISTORS IN ALGAAS/INGAAS/GAAS

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

By
Darrell G. Hill
June 1990
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.

James S. Harris, Jr. (Principal Advisor)

Richard Swanson (Electrical engineering)

James D. Plummer (Electrical engineering)

Approved for the University Committee on Graduate Studies:

Dean of Graduate Studies
Abstract

Heterojunction bipolar transistors (HBTs) in GaAs-based materials systems are of interest because of their demonstrated capability for high-speed operation. While almost all work on HBTs has been performed on N-p-n devices, P-n-p HBTs are attractive for several reasons. First, the N-p-n/P-n-p combination is promising for complementary circuits, allowing lower power operation and greater flexibility in circuit design. Also, optimized P-n-p HBTs are expected to have performance comparable to that of the fastest N-p-n devices while offering significant technological advantages in fabrication.

However, the thin base region of high-speed P-n-p structures complicates fabrication. In addition, P-n-p HBT's suffer from the same problems with surface effects which influence the performance of N-p-n devices and limit scaling of lateral dimensions.

In this dissertation the design, epitaxial growth, fabrication, and testing of P-n-p HBT's in AlGaAs/InGaAs/GaAs are described. The use of thin, graded AlGaAs layers for suppression of surface effects is investigated both theoretically and experimentally, leading to the first explanation of the surface passivation mechanism of such layers. This explanation allows design of AlGaAs layers for optimal surface passivation.
Two selective etch solutions were developed for the removal of GaAs and AlGaAs/GaAs layers from underlying InGaAs layers. These solutions are very sensitive to indium content, with as little as 5% indium making a significant difference in the etch rate.

Using optimized AlGaAs layers for suppression of surface effects and selective etching for exposure of the thin base region, highly uniform, high-gain P-n-p HBT's have been demonstrated. Current gains over 400 were obtained with emitter sizes as small as 2.5 x 4 \( \mu m^2 \); standard deviation of current gain was 5%. Despite the very thin 500 \( \AA \) base layer, yields were consistently above 98%. The scaling properties of P-n-p HBT's are shown to be superior to those of N-p-n devices with comparable projected high-speed performance.
Acknowledgments

I am indebted to many people for the support and encouragement given me during my years at Stanford. First and foremost is Professor James Harris. He has always been very supportive of my work, even those ideas which were "far-fetched" (of which I have had more than a few). The breadth of research and the abundance of new ideas coming out of his research group have been a tremendous experience for me.

I would like to thank Professors Richard Swanson and James Plummer for reading my dissertation, and Professor David Bloom for serving on my oral defense committee.

I would also like to thank Won-Seong Lee, Jim Adkisson, and Tony Ma, without whose insight and assistance I would never have learned so much about HBT's or MBE. While their assistance with MBE growth of my wafers was absolutely necessary to my research, I value much more highly the way that they were never too busy to help or chat, in spite of all the work they did to keep the lab functioning (plus their thesis work as well).

I have benefitted from many conversations and collaborations with other members or former members of Professor Harris's research group, most notably Kanji Yoh, William Liu, Damian Costa, and Kevin Lear.

Mak Salimian of Varian was very generous with his time, and I enjoyed very much my collaboration with him.
I would like to thank Gail Chun-Creech and Susie Burns for all of their help (and the occasional sympathetic ear) in addition to keeping the research group running smoothly.

Finally, I would like to thank my wife Cindy. She has been a constant source of encouragement during my time at Stanford. I do not know what I would have done without her support during the two years when just obtaining working transistors seemed an almost impossible goal. She deserves much of the credit for the perseverance which allowed me to continue.
Contents

List of Tables ........................................................................................................ xii
List of Figures ......................................................................................................... xiii
1. Introduction ....................................................................................................... 1

2. Background material ......................................................................................... 6
   2.1 GaAs background ............................................................................................ 6
       2.1.1 Bandgap engineering ............................................................................. 6
       2.1.2 Carrier transport and resistivity ............................................................. 10
       2.1.3 Surface properties and ohmic contacts ............................................... 11
       2.1.4 Thermal conductivity ........................................................................... 12
   2.2 Potential of the HBT for high-speed IC’s ....................................................... 12
       2.2.1 Advantages of the HBT ....................................................................... 13
       2.2.2 Disadvantages of the HBT ................................................................. 15
   2.3 Molecular Beam Epitaxy ............................................................................... 16
       2.3.1 Issues for MBE growth of strained-layer HBT wafers ....................... 18
       2.3.2 Outdiffusion of base dopant and undoped spacer layers ..................... 20
       2.3.3 Grading of heterojunctions in MBE .................................................... 21
       2.3.4 Si-doped AlGaAs and DX centers ....................................................... 22

3. Device physics .................................................................................................. 23
   3.1 Introduction ................................................................................................... 23
   3.2 Operation of conventional bipolar transistors ............................................. 23
   3.3 Operation of heterojunction bipolar transistors .......................................... 25
3.4 Quantitative development of HBT principles..........................27
  3.4.1 Emitter injection efficiency (γ)........................................28
  3.4.2 Base transport factor (α_t).............................................34
3.5 Importance and design of emitter-base grading......................35
  3.5.1 Optimal grading..........................................................36
  3.5.2 The role of abrupt emitter-base heterojunctions..................42
  3.5.3 The Light Hole P-n-p (LHPNP)........................................44
3.6 Device speed -- f_t and f_{max}...........................................45
3.7 Surface effects....................................................................50
  3.7.1 Emitter size effect.........................................................52
  3.7.2 Emitter/collector offset voltage.....................................56
3.8 Design trade-offs between the P-n-p and the N-p-n.................58
3.9 Strained semiconductor systems..........................................60
  3.9.1 Strain and critical thickness..........................................61
  3.9.2 Strain effects on band structure and carriers....................64
4. Initial results for strained-base P-n-p heterojunction bipolar transistors.............................................67
  4.1 Preface to Chapter 4..........................................................67
  4.2 Introduction.........................................................................67
  4.3 Growth and processing......................................................68
  4.4 Results and discussion......................................................70
  4.5 Conclusion..........................................................................72
5. Investigation of the emitter passivation mechanism of thin AlGaAs layers...............................................74
  5.1 Introduction..........................................................................74
  5.2 Analysis...............................................................................76
6. Two selective etching solutions for GaAs on InGaAs
   and GaAs/AlGaAs on InGaAs........................................89
   6.1 Introduction......................................................90
   6.2 Sample preparation.............................................90
   6.3 Experimental results..........................................90
   6.4 Application....................................................97
   6.5 Conclusion.....................................................97

7. Uniform, high-gain AlGaAs/InGaAs/GaAs P-n-p HBT's
   by dual selective etch process.....................................98
   7.1 Introduction....................................................98
   7.2 Device fabrication..............................................98
   7.3 Experimental results..........................................101
     7.3.1 Uniformity of current gain..............................101
     7.3.2 Emitter size effect.......................................102
     7.3.3 Yield........................................................105
     7.3.4 Material quality considerations.........................106
   7.4 Conclusion.....................................................109

8. Conclusions........................................................110
   8.1 Summary........................................................110
   8.2 Suggestions for future work..................................112

Appendices
   A. Grading of the emitter-base heterojunction.................116
B. Dual selective etch HBT fabrication process...............................122
References......................................................................................131
List of Tables

Table 4.1  Epitaxial layer structure for first strained-layer P-n-p HBT........................................................................................................................................69

Table 4.2  Comparison of HBT’s with different emitter and base compositions........................................................................................................................................72

Table 5.1  Epitaxial layer structure for testing effect of surface passivation layer thickness........................................................................................................................................80

Table 6.1  Epitaxial layer structure for testing of H₂O₂/NH₄OH etch solution........................................................................................................................................91

Table 6.2  Epitaxial layer structure for testing of K₃Fe(CN)₆/K₄Fe(CN)₆ etch solution........................................................................................................................................94

Table 7.1  Epitaxial layer structure of dual selective etch P-n-p HBT's........................................................................................................................................99
List of Figures

Figure 2.1 Energy bandgap and lattice constant for various semiconductors..............................................................8

Figure 2.2 Schematic diagram of Stanford's molecular beam epitaxy system.................................................................17

Figure 3.1 Schematic cross-section of a bipolar transistor showing major components of current flow.........................24

Figure 3.2a Band diagram of conventional p-n-p bipolar transistor.................................................................................26

Figure 3.2b Band diagram of P-n-p heterojunction bipolar transistor.............................................................................26

Figure 3.3 Plot of log (β) versus log (Iₒ), showing effect of traps in emitter-base space charge region: current gain improves with increasing current density..............................34

Figure 3.4 Electron and hole concentrations in an HBT under normal operating conditions.........................................35

Figure 3.5 Band diagram for perfect heterojunction between undoped AlAs and undoped GaAs...............................37

Figure 3.6 Band diagram of an abrupt P-n AlAs/GaAs heterojunction (p-n GaAs homojunction case is in dotted lines)........................................................................................................38

Figure 3.7 Band diagram for P-n AlAs/GaAs heterojunction linearly graded over 300 Å, under 1.2 V forward bias...39

Figure 3.8 P-n heterojunction using optimum grading (except for grading, same parameters as in Figure 3.7)..............41

Figure 3.9 Comparison of graded and abrupt emitter-base heterojunctions, showing the significant reduction of bandgap (and therefore increased recombination current) for the graded case.........................................................43

Figure 3.10 Pinning of the Fermi level at the surface of GaAs...........51
Figure 3.11  Pinning of the Fermi level at the surface of InGaAs, as a function of In content (after H. H. Wieder, Applied Physics Letters, vol. 38, pp. 170-171, 1980)..................52

Figure 3.12  Degradation in current gain as lateral dimensions of HBT in GaAs are reduced (after Nakajima et. al., Japan. J. Appl. Phys., Vol. 24, pp. L596-L598, 1985)......53

Figure 3.13  The three components which contribute to emitter size effect: 1) Recombination via surface traps in the emitter-base depletion region  2) Recombination at the exposed base surface  3) Diffusion of holes into the bulk extrinsic base.........................................................54

Figure 3.14  Example of the emitter-collector offset voltage for a GaAs HBT: the collector current is negative below an emitter-collector bias of approximately 200 mV........57

Figure 3.15  Exaggerated illustration of strain caused by lattice mismatch between InGaAs and GaAs substrate:  a) stable strained layer  b) unstable strained layer which causes misfit dislocations..................................................61

Figure 3.16  Matthews-Blakeslee limit for thickness of a strained epitaxial layer (after J. W. Matthews and A. E. Blakeslee, Journal of Crystal Growth, vol. 27, pp. 118-125, 1974).................................................................63

Figure 3.17  Schematic representation of the band structure of a) an unstrained tetrahedral semiconductor with a direct bandgap, and b) the same semiconductor under biaxial compression. CB=conduction band, HH=heavy-hole band, LH=light-hole band, and SO=spin-split-off band..........................................................65

Figure 4.1  Schematic cross section of P-n-p HBT.................................70

Figure 4.2  Common emitter I/V characteristics of 24 x 30 μm² graded-base HBT.................................................................71

Figure 4.3  Current gain versus collector current for 24 x 30 μm² graded-base HBT.................................................................71
Figure 5.1  Passivation of the surface surrounding the emitter mesa by means of a thin, wholly-depleted AlGaAs layer (after H. H. Lin and S. C. Lee, Applied Physics Letters, vol. 47, pp. 839-841, 1985) .................................................. 75

Figure 5.2  Band diagram for depleted AlGaAs surface passivation layer ................................................................. 76

Figure 5.3  Band diagram for 500 Å surface passivation layer ......... 81

Figure 5.4  Band diagram for 700 Å surface passivation layer ......... 82

Figure 5.5  Band diagram for 1000 Å surface passivation layer ...... 83

Figure 5.6  Current gain dependence on thickness of AlGaAs surface passivation layer surrounding the emitter perimeter .............................................................................. 84

Figure 5.7  Band diagram for surface passivation layer in H. H. Lin's super-gain HBT (H. H. Lin and S. C. Lee, Applied Physics Letters, vol. 47, pp. 839-841, 1985) .................................................. 86

Figure 6.1  Etching of 1000 Å GaAs layer over InGaAs using H$_2$O$_2$/NH$_4$OH solution of pH 7.3 at 10° C ...................... 92

Figure 6.2  Temperature dependence of GaAs etch rate in H$_2$O$_2$/NH$_4$OH solution of pH 7.3 ................................................. 93

Figure 6.3  Relation of etch depth and etch time for GaAs and Al$_{0.3}$Ga$_{0.7}$As in K$_3$Fe(CN)$_6$/K$_4$Fe(CN)$_6$ solution at 25° C ...................................................................................... 95

Figure 6.4  Etching of 1000 Å GaAs/AlGaAs over InGaAs using K$_3$Fe(CN)$_6$/K$_4$Fe(CN)$_6$ solution at 25° C ...................... 96

Figure 7.1  The two selective etching steps: a) Exposure of AlGaAs using H$_2$O$_2$/NH$_4$OH solution  b) Exposure of InGaAs away from active emitter edge using K$_3$Fe(CN)$_6$/K$_4$Fe(CN)$_6$ solution ............................................. 100

Figure 7.2  Overlaid incremental current gain characteristics for all HBT's on two dies .................................................................................. 102

Figure 7.3  Emitter size effect for P-n-p HBT's with 500 Å base. 103
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.4</td>
<td>Effect of surface passivation on current gain and emitter size effect for P-n-p HBT's</td>
<td>104</td>
</tr>
<tr>
<td>7.5</td>
<td>Comparison of emitter size effect for N-p-n and P-n-p HBT's, with and without surface passivation</td>
<td>105</td>
</tr>
<tr>
<td>7.6</td>
<td>Base and collector currents versus base-emitter bias for P-n-p HBT at a constant emitter-collector bias of -3 V (emitter size is 3 x 10 μm²)</td>
<td>106</td>
</tr>
<tr>
<td>7.7</td>
<td>Improvement of base ideality factor from 3 to 2 by separation of In and Al in the emitter-base space charge region</td>
<td>107</td>
</tr>
<tr>
<td>A.1</td>
<td>Band diagram of p-n junction depletion layer under bias</td>
<td>118</td>
</tr>
<tr>
<td>B.1</td>
<td>Lift-off of emitter contact metal</td>
<td>123</td>
</tr>
<tr>
<td>B.2</td>
<td>Formation of AlGaAs surface passivation layer by selective etching</td>
<td>125</td>
</tr>
<tr>
<td>B.3</td>
<td>Exposure of base (away from emitter edge) by selective etching</td>
<td>127</td>
</tr>
<tr>
<td>B.4</td>
<td>Lift-off of base contact metal</td>
<td>128</td>
</tr>
<tr>
<td>B.5</td>
<td>Formation of base mesa by nonselective etching</td>
<td>128</td>
</tr>
<tr>
<td>B.6</td>
<td>Deposition of silicon nitride by PECVD, then opening of contact holes by RIE</td>
<td>129</td>
</tr>
<tr>
<td>B.7</td>
<td>Patterning of interconnect metal by lift-off</td>
<td>130</td>
</tr>
</tbody>
</table>
Chapter 1
Introduction

The broadest trends characterizing the progress of semiconductor technology are increasing levels of integration, reduction in size of individual devices, and increasing speed. No other industry comes close to the rapid and steady progress exhibited by silicon integrated circuit manufacturing.

While device dimensions and the level of integration continue to advance steadily, the maximum speed of silicon integrated circuits has been improving relatively slowly. In some cases, fundamental limits are being approached. For example, when microwave operation is obtained in silicon (often through the use of sub-micron features defined by the expensive, low-throughput process of electron beam lithography), parasitic capacitance and crosstalk caused by substrate conduction are severe problems.

The fastest silicon circuits are sufficiently technologically complex that gallium arsenide integrated circuits are becoming increasingly attractive. For a given level of speed performance, GaAs devices put less stringent demands on lithographic capability due to superior carrier transport properties. High-quality, semi-insulating GaAs substrates greatly reduce parasitic capacitances. However, GaAs technology is many years behind the silicon industry. The first GaAs VLSI circuit, a 32-bit microprocessor, was announced only in 1988;¹ 16k GaAs static RAM are just becoming available.
Several device approaches have been suggested to take advantage of the speed potential of GaAs. MESFET (metal-semiconductor field effect transistor), HEMT (high electron mobility transistor), and HBT (heterojunction bipolar transistor) devices all have their proponents. For high levels of integration, the HBT seems most promising due to the inherent uniformity of the device turn-on voltage, an absolute requirement for high speed integrated circuits (this topic will be discussed further in section 2.2). The highest level of integration obtained in GaAs to date, the microprocessor previously mentioned, uses the HBT.

The high-speed performance of AlGaAs/GaAs HBT's has been amply demonstrated, with figures of merit ($f_T$ and $f_{max}$) exceeding 100 GHz$^2$ and ring oscillator propagation delays below $6 \times 10^{-12}$ second.$^4$

All of these records were obtained with N-p-n HBT's; to date, most interest in P-n-p devices has been due to desire for a complementary technology in GaAs analogous to CMOS technology in Si (by convention, use of the capital "N" or "P" denotes the wide-bandgap material in an HBT). Nevertheless, theoretical studies have consistently shown that suitably-optimized P-n-p HBT's should have performance comparable to that of the fastest N-p-n devices.$^5$ In addition to their potential for high speed, P-n-p HBT's have significant technological advantages over N-p-n HBT's. Outdiffusion of the base dopant, which is responsible for variations in both turn-on voltage and current gain in N-p-n devices, is much less severe for
P-n-p HBT's; P-n-p HBT's do not suffer from the material quality problems associated with n-type AlGaAs; the availability of fast-diffusing acceptors with high solid solubility makes a planar process more likely for P-n-p devices. One of the key points of this dissertation, the superior lateral scaling properties of P-n-p HBT's, has not been previously recognized.

Despite the interest in complementary circuits and the potential advantages of the P-n-p, only two reports of working P-n-p HBT's in AlGaAs/GaAs had been published when this work was undertaken.\textsuperscript{7,8} The reported devices were rather large (smallest emitter was 10 x 50 μm\textsuperscript{2}), and the maximum current gain of any of the devices was 50.

This dissertation includes the first highly-uniform P-n-p HBT's, as well as the first P-n-p HBT's utilizing strained layers. The superiority of the P-n-p for small device geometries is demonstrated experimentally. Also, the issue of yield for very thin-base devices is addressed with the first reported wet etching solutions for selective removal of GaAs and AlGaAs/GaAs from thin InGaAs layers.

Chapter 2 presents background material on AlGaAs/GaAs heterojunction bipolar transistors, including fundamental differences between Si and GaAs, the potential of the HBT compared to other GaAs device approaches, and description of HBT growth by molecular beam epitaxy.
Chapter 3 begins with a simple, plain-English description of HBT operating principles. A quantitative development of the same material is then presented, along with an analysis of speed-limiting mechanisms and the role of surface effects. Design trade-offs between the N-p-n and P-n-p are discussed in section 3.8, followed by a description of strained semiconductor systems in section 3.9.

Chapter 4 is adapted from a paper describing the first strained-layer P-n-p transistors. Many of the issues discussed in later chapters, such as surface passivation, device yield, and uniformity, arose from the results presented in this chapter.

The first explanation of the surface passivation mechanism of thin AlGaAs layers is developed in Chapter 5. While thin, wholly-depleted AlGaAs layers have been shown to reduce the surface recombination at the edge of the emitter in some cases, design of such layers has been empirical and time-consuming. The analysis in this chapter gives a plausible explanation for the passivation mechanism of such layers and is in very good agreement with experimental results. Using this analysis, such thin AlGaAs layers may be designed for optimal passivation.

Chapter 6 describes the first two wet chemical etching solutions for selective etching of GaAs and AlGaAs/GaAs with respect to InGaAs. Both solutions are very sensitive to In content, with as little as 5% In making a significant difference in the observed etch rate. Such sensitive etches will find application in
fabrication of HBT's, resonant tunneling transistors, and probably other devices as well.

Chapter 7 is a description of the application of the selective etch described in Chapter 6 to the problem of yield for thin-base HBT's. In addition to resolving the issue of yield, the dual selective etch process produced HBT's with outstanding uniformity. The superior potential of the P-n-p HBT for small device geometries is demonstrated by comparison with similar N-p-n HBT's.

Chapter 8 concludes with a summary of results and suggestions for future work.

Many of the things which were learned during the course of this research, while not fundamental in nature, are nevertheless very important from a practical standpoint; the appendices contain this type of information. Appendix A contains a detailed description of the design of optimal emitter-base heterojunction grading for HBT's, as well as execution of grading by molecular beam epitaxy. Appendix B contains an annotated description of the fabrication sequence used for making the HBT's discussed in Chapter 7.
Chapter 2
Background Material

2.1 GaAs background

Si is currently the dominant semiconductor for integrated circuits, not because of fundamental advantages in carrier transport properties, but rather because it offers technological advantages in circuit fabrication. During the infancy of integrated circuitry, the superior electron and hole mobilities of the older, established Ge technology were outweighed largely by the ease with which SiO₂ could be grown on Si wafers for pattern definition, masking, insulation, and surface passivation. The importance of this capability should not be underestimated. Nevertheless, as Si technology has progressed, previously "insurmountable" problems for Ge and GaAs, such as deposition of dielectric layers, were solved for Si. With the appearance of LSI and even VLSI circuits in GaAs, it is appropriate to re-examine the fundamental differences between Si and GaAs.

2.1.1 Bandgap engineering

The single most important reason for considering GaAs for semiconductor devices is the capability for "bandgap engineering." GaAs is lattice-matched to (ie., has the same crystal structure and lattice constant as) AlAs. Consequently, heterojunctions (junctions between two different semiconductors in a single crystal) of GaAs and AlₓGa₁₋ₓAs (0 ≤ x ≤ 1) may be incorporated into devices,
providing the device designer with an extra degree of freedom: he may specify the bandgap of the semiconductor. A change in bandgap acts somewhat like an electrostatic potential difference, but is independent of any purely electrostatic potential variations. Controlling the bandgap profile across a device therefore allows much greater control of the distribution and flow of electrons and holes than is possible in homojunction (single semiconductor) devices.

Few semiconductor pairs are suitable for making device-quality heterojunctions, as many constraints must be satisfied. First, as mentioned previously, the lattice constants of the two semiconductors should be reasonably well-matched. If mismatched materials are used, the structure will be strained, and the thickness of the strained layer must be limited to avoid misfit dislocations at the interface which could seriously degrade device performance. Lattice mismatch strain will be discussed more fully in section 3.9.

Second, the bandgap difference between the two semiconductors should be large, preferably with capability for compositional grading. The constraints of lattice matching and bandgap difference narrow the field significantly. Bandgap energy versus lattice constant is given for a selection of semiconductors in Figure 2.1.

The third requirement for device-quality heterojunctions is demonstrated by the GaAs-Ge system: the two semiconductors should preferably be isoelectronic. While GaAs and Ge are well
lattice-matched and have significantly different bandgaps, GaAs/Ge heterojunction devices are difficult to make, although it has been done.\textsuperscript{12} Because Ga and As act as acceptor and donor impurities in Ge, while Ge can be either a donor or an acceptor in GaAs, even a slight amount of interdiffusion across the heterojunction results in heavily compensated material which is not useful for most device applications. In addition, compositional grading between GaAs and Ge is impossible without very heavy compensation. For these reasons, only isoelectronic pairs of semiconductors (having the same number of valence electrons, such as GaAs/AlAs or Si/Ge) presently seem promising for heterojunction devices.

![Graph showing bandgap vs lattice constant for various semiconductors](image-url)

Figure 2.1 Energy bandgap and lattice constant for various semiconductors

(data taken from reference 15, pp. 848-849)

Heterojunction systems for semiconductor devices must also have lattice-matched substrates available for crystal growth. The substrate need not be either of the materials in the heterojunction system, as shown by the current interest in $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$/
In$_{0.52}$Al$_{0.48}$As. Although GaAs, AlAs, and InAs are not lattice-matched to InP, the proper ternary alloys are. InGaAs/InAlAs transistors have superior properties in many respects due to the lower surface recombination velocity of InGaAs (see section 3.7); however, the physical frailty of the InP substrate and the increased complexity of maintaining lattice matching during growth are practical constraints. As InP growth technology matures, InGaAs/InAlAs/InP will probably play an increasing role in the semiconductor device world.

Other considerations for heterojunction systems for devices include availability of donors and acceptors with a wide range of attainable doping, and compatibility of materials for simultaneous growth. GaAs, AlAs, and their ternary alloys are currently the most commonly used semiconductors for heterojunction devices because GaAs and AlAs meet all of these requirements, and GaAs substrates of high purity are readily available.

The recently-demonstrated ability to obtain Si/SiGe heterojunctions cannot be overlooked.$^{13,14}$ Such a capability is a promising development for Si technology and may eventually make the Si bipolar transistor more competitive for very high speeds. However, Si/SiGe technology is still in its infancy, and problems with outdiffusion of base dopant, as well as issues of compatibility with existing Si processes, must be addressed.
2.1.2 Carrier transport and resistivity

In addition to the benefits of bandgap engineering which are possible in the GaAs/AlAs system, GaAs has the advantage of an electron mobility that is roughly five times larger than the electron mobility in Si (the hole mobilities are comparable). The high mobility reduces both electron transit time of devices and the resistivity of n-type bulk regions, both of which are important for high speed devices.

While the limiting high-field saturation velocities for both Si and GaAs are approximately $10^7$ cm/sec, the velocity in GaAs at fields from 2000 to 5000 V/cm is higher: 2 to $3 \times 10^7$ cm/sec. The electron velocity in Si does not approach $1 \times 10^7$ cm/sec until an electric field of 30,000 V/cm is applied.\textsuperscript{15} In FET devices and in the collectors of bipolar devices, high velocity at this lower electric field results in high speed operation with significantly lower power consumption.

The decrease in electron velocity in GaAs at electric fields above 3000 V/cm is due to the scattering of electrons from the high-mobility ($\Gamma$) valley to the lower-mobility (L) valley in the conduction band. The L valley, about 0.3 eV higher in energy than the $\Gamma$ valley, is normally unoccupied; however, electrons may gain enough energy from the electric field to be transferred to the upper valley, resulting in a negative differential resistance.

If the electron transit time of a high-field region is shorter than the time required for equilibration between the $\Gamma$ and L valleys,
then electron velocities significantly above the $3 \times 10^7$ cm/sec figure may be obtained. Evidence of such "velocity overshoot" has been observed experimentally in both FET and bipolar devices.

While silicon-on-insulator devices have been made using various insulating substrates such as sapphire and silicon dioxide to reduce parasitic capacitances, semi-insulating GaAs substrates are readily available. Substrate capacitance can therefore be eliminated without the materials problems associated with deposition onto substrates with different mechanical properties.

2.1.3 Surface properties and ohmic contacts

As mentioned in the opening of this chapter, the primary reason for the primacy of Si in semiconductor technology is the ease with which silicon dioxide can be grown on the surface for insulation, passivation, and processing. The interface between Si and thermally-grown SiO$_2$ is stable, of very high quality, and is reproducible. Oxidation is an extremely effective method of passivating the Si surface.

In contrast, the surface properties of GaAs are probably the most significant impediment to its use for devices. The GaAs surface is electronically active, with a high concentration of donor and acceptor levels in the bandgap due to Ga and As vacancies and other defects. These levels keep the Fermi level pinned near the center of the GaAs bandgap, regardless of doping density, and act as efficient traps for minority carriers. As a result of this Fermi level pinning, the surface recombination velocity for GaAs is very high (on
the order of $10^6$ cm/sec), and all direct junctions between GaAs and metals act as Schottky diodes. Ohmic characteristics are obtained only by doping very heavily to insure that current flow across the Schottky barrier is dominated by tunneling.\textsuperscript{16} The behavior of the GaAs surface is discussed further in section 3.7.

2.1.4 Thermal conductivity

The thermal conductivity of a semiconductor is important because integrated circuits generate substantial heat. The ultimate limitation on circuit speed is often the amount of heat which can be dissipated before the characteristics of individual devices begin to suffer.

Si is the clear winner in this area, with a room-temperature thermal conductivity of 1.5 W/cm·°C, compared to GaAs's 0.46 W/cm·°C.\textsuperscript{15} This is somewhat compensated by the larger bandgap of GaAs, allowing GaAs devices to run at much higher temperatures. Nevertheless, the higher temperatures mean greater device-to-device temperature differentials and therefore greater variation in device parameters on the same chip. If heat dissipation limits the performance of a circuit, then the GaAs IC must have substantially lower power consumption at a given speed (which is indeed often the case) to be competitive with the Si circuit.

2.2 Potential of the HBT for high-speed IC's

Three devices in GaAs (or AlGaAs/GaAs) have demonstrated potential for high-speed integrated circuits: the metal-
semiconductor field-effect transistor (MESFET), the heterojunction bipolar transistor (HBT), and the modulation-doped field-effect transistor (MODFET; also known as HEMT, for high electron mobility transistor).

The MESFET was the first commercially viable three-terminal device technology in GaAs. With a relatively simple process and speeds considerably beyond that of commercially available Si, MESFET's are still in widespread use in fast gate arrays and microwave amplifiers.

The MODFET employs a heterojunction between a doped AlGaAs layer and an undoped GaAs channel layer. Because the electrons are confined to the undoped GaAs channel, the mobility is very high, and the saturated velocity is reached at much lower electric field (and therefore lower applied bias) than for the MESFET. This performance gain is offset by increased complexity in fabrication.

The HBT has many fundamental advantages over the MESFET and the MODFET for IC applications; once again, however, these advantages come at the cost of increased complexity. If yield and device scaling continue to improve for HBT's in GaAs, then high-speed IC's using these devices will be commercially viable.

2.2.1 Advantages of the HBT

The fundamental requirements for high-speed VLSI have been examined by Eden.\textsuperscript{17} Since the minimum dynamic switching energy (gate power dissipation times gate propagation delay) is proportional to the square of the logic voltage swing, successful
circuits must use devices which are not only fast, but fast with small logic swings.

Small logic swings impose stiff constraints on devices. For noise considerations, the spread in device threshold voltage in a circuit must be a small fraction of the logic swing; therefore, strict control of threshold voltage is an absolute necessity for fast VLSI. Also, very high transconductance is required so that these small voltage swings can be turned into usable signals.

For bipolar transistors (homojunction as well as heterojunction), the "threshold voltage" -- the value of $V_{be}$ for some specific $I_c$ -- is largely determined by the bandgap of the base layer. Consequently, uniformity of threshold voltage to within a few mV is easily obtainable. In contrast, the threshold voltage of FET devices is very sensitive to variations in thickness and doping density of the channel region.

The exponential current-voltage characteristic of bipolar transistors guarantees a very high transconductance above the turn-on voltage, so that low voltage swings are practical. In addition, the high current drive capability makes bipolar transistors less sensitive to the loads imposed by fan-out or long (or off-chip) signal lines.

The incorporation of a heterojunction into the bipolar transistor allows additional freedom in choosing the doping densities of the emitter and base layers. Compared to conventional bipolar transistors, the HBT has lower base resistance and emitter-
base junction capacitance, giving it a significant speed advantage if parasitic resistances and capacitances are properly controlled.

2.2.2 Disadvantages of the HBT

Fabrication of bipolar transistors is more complex than that of FET devices. Three different epitaxial layers must be contacted rather than two, and ohmic contacts to both n- and p-type material are required.

From the standpoint of device physics, the most important shortcomings of HBT's in GaAs can be traced to the electronic behavior of the surface (this behavior and its effect on devices is described in section 3.7). Because high current density is required for best speed performance (section 3.6), power consumption is excessive unless the device dimensions are sufficiently small; however, surface effects cause the current gain to degrade as lateral dimensions are reduced.

Surface effects are also responsible for the complexity of GaAs ohmic contact technology. Due to the high current densities involved, ohmic contacts must have very low resistance if they are not to limit device performance.

Although it does not address ohmic contacts, this dissertation addresses the issue of device scaling by investigating passivation of the surface by means of thin AlGaAs capping layers, and also by demonstrating the P-n-p HBT's relative insensitivity to surface effects because of its thinner base layer.
2.3 Molecular Beam Epitaxy

Molecular beam epitaxy (MBE) is a technique for growing crystals which allows very precise control of crystal composition. The crystal growth takes place in an ultra-high vacuum chamber which contains heated effusion cells or furnaces. Each furnace contains one of the constituent materials of the crystal (or an impurity for doping the crystal) and is equipped with a shutter. During growth, each furnace is heated to the temperature at which the material it contains begins to evaporate. Because of the low pressure, the atoms or molecules which are boiled off form atomic or molecular beams which impinge on the substrate. The substrate is heated so that the atoms or molecules which strike the surface have some lateral mobility and arrange themselves onto proper crystal lattice sites, extending the crystal in an orderly fashion.

The flux of the beams, and therefore the growth rate (or impurity density), is controlled by varying the temperature of the furnaces. A typical growth rate for GaAs and related compounds is one atomic layer per second. Since the shutters on the furnaces can be opened and closed in a small fraction of a second, true atomic-layer control of the composition of epitaxial layers is routinely achieved. Figure 2.2 shows a schematic diagram of the growth chamber of an MBE system.

Using sophisticated pumping systems and liquid-nitrogen cooled cryo-shrouds to keep background pressures below $10^{-10}$ Torr, the ultra-high vacuum chamber prevents contamination during
growth. Also, the purest available material is used in the source furnaces. High quality semiconductor crystals suitable for devices are therefore obtained. Quality of the surface can be monitored during the growth by RHEED (reflection high energy electron diffraction), which is sensitive to surface morphology.

Figure 2.2 Schematic diagram of Stanford's molecular beam epitaxy system (figure courtesy of Won-Seong Lee)

When growing GaAs, AlAs, InAs, or other III-V compounds, the flux of the group V elements (As, for these materials) is usually several times higher than the group III element flux. The group V
atoms tend to stick only to group III atoms on the surface with unsatisfied bonds, so that the proper stoichiometry is preserved. Because As tends to desorb from GaAs at temperatures above 500° C, the shutter on the As furnace is left open to maintain significant As overpressure on the GaAs substrate whenever the substrate is at elevated temperatures.

The group III atoms, however, have sticking coefficients close to unity; i.e., very few of the group III atoms will desorb from the surface once they land there. If a ternary alloy, such as Al₀.₃Ga₀.₇As, is desired, then the furnace temperatures are set so that

$$\frac{Al\text{ flux}}{Al\text{ flux} + Ga\text{ flux}} = 0.3.$$ 

Transients in flux have been noted following opening of furnace shutters during MBE growth. These transients are caused by a slight cooling which occurs when the shutter, which has been radiatively heated by the furnace, moves away from the furnace opening to expose the relatively cool interior of the MBE growth chamber. The furnace then cools slightly due to increased blackbody radiation loss.

In Stanford's MBE system, the furnaces have been modified somewhat so that the shutter is further from the furnace opening, reducing these transients to less than 5% of the steady-state flux.

2.3.1 Issues for MBE growth of strained-layer HBT wafers

While the temperature/flux characteristic is calibrated regularly for each furnace, the strained InGaAs layers in this
dissertation required especially strict control of both composition and thickness, since a few percent variation in either parameter could cause the layer to exceed the critical thickness and form large numbers of misfit dislocations (discussed in section 3.9.1). Therefore, for growth of wafers containing InGaAs layers, the fluxes of the Ga, Al, and In furnaces were checked very carefully before initiating growth. The different fluxes are measured by rotating an ionization gauge into the position normally occupied by the substrate, and then opening one shutter at a time. The flux corresponding to a given growth rate is known for each furnace, and a linear relationship between measured flux and growth rate has been established.

Material quality is dependent on substrate temperature during growth, with different materials requiring different growth temperatures. The GaAs and AlGaAs layers in this work were grown at either 600° or 630° C, while the InGaAs layers were grown at 500° C due to the desorption of In which occurs at higher temperatures. Between InGaAs and GaAs layers, the growth was interrupted by closing all shutters except the one on the As furnace while the substrate temperature was reset. If GaAs was to be grown on an InGaAs layer, then a 20 Å layer of GaAs was grown before the temperature was raised in order to prevent In desorption at the higher temperature.
2.3.2 Outdiffusion of base dopant and undoped spacer layers

In growth of N-p-n HBT's in AlGaAs/GaAs by MBE and MOCVD (metalorganic chemical vapor deposition\textsuperscript{20}), outdiffusion of the base dopant must be taken into consideration. At the concentrations used for high-performace N-p-n HBT's (1 \times 10^{19}/\text{cm}^3 or higher), the acceptor dopants used in these technologies tend to stay at the surface and ride along the growth front. Because the base is more heavily doped than the emitter, a "tail" of the base doping profile extending into the emitter converts part of the emitter from n to p-type, moving the p-n junction into the wider bandgap material and drastically affecting the device current gain and turn-on voltage.

The most common way of dealing with this problem is to grow an undoped spacer layer on top of the base layer to "absorb" this doping tail before growth of the emitter. However, such spacer layers decrease the maximum current gain and make current gain more strongly dependent on current density.\textsuperscript{21} Because the effective width of this spacer layer depends on the severity of acceptor outdiffusion, the device characteristics will almost certainly vary from wafer to wafer, and possibly even across a wafer, due to nonuniform outdiffusion. This is a serious problem at high doping densities, since strict control of device turn-on voltage is an absolute requirement for high-speed IC's, as described in section 2.2.1.
Because the base sheet resistivity is still a limiting factor in the speed of N-p-n HBT's, there is interest in raising the base doping beyond the low $10^{19}$ level. In order to keep the acceptors from diffusing hundreds of Angstroms into the emitter during growth, the substrate temperature is often lowered, and the As flux is increased by a factor of three to five or more. Unfortunately, HBT's grown in this fashion have low minority carrier lifetime in the base layer -- even lower than expected from the inverse relationship of lifetime to doping density -- and rarely have maximum current gains above 30. The lowered substrate temperature and increased As flux result in a higher concentration of As antisite defects (As atoms occupying Ga lattice sites) which may be responsible for the lower base lifetime. However, the lower lifetime could also be due to Auger recombination, which is nonlinear in doping density.

Because the resistivity of n-type GaAs is much lower than that of p-type GaAs at the same doping density (roughly 1/15th), lower base doping is acceptable for P-n-p HBT's. In addition, the diffusivity of Si donors is much lower than that of Be or Zn acceptors used for MBE and MOCVD, respectively. Consequently, undoped spacer layers have been shown to be unnecessary in P-n-p HBT's. P-n-p devices therefore retain the excellent threshold voltage uniformity required for high-speed integrated circuits.

2.3.3 Grading of heterojunctions in MBE

For most applications, the emitter-base heterojunction of an HBT should be compositionally graded to remove discontinuities in
the conduction and valence bands at the interface. Because the grading occurs over a short distance (typically 200 - 300 Å), the MBE furnaces do not have time to respond to changing temperatures. Instead, the fluxes are modulated mechanically by opening and closing the shutter(s) with a varying duty cycle. Motivation for heterojunction grading is given in section 3.5, while grading design and execution of grading by MBE is described in detail in Appendix A.

2.3.4 Si-doped AlGaAs and DX centers

N-type AlGaAs is known to have a high concentration of deep donors, particularly for Al content of 30% or more.\textsuperscript{24,25} Such deep donors act as generation-recombination centers, causing current gain to degrade at low current densities, and increasing the noise of the device. The DX center in particular has been associated with Al-Si interactions, and increases in concentration as either the Al content or Si doping concentration is increased.

In N-p-n HBT's, the n-type AlGaAs emitter is necessarily doped with Si, so that DX centers account for recombination in the emitter-base space charge region, and also contribute low frequency (1/f) noise which is undesirable for some applications. P-n-p HBT's, however, have Be rather than Si in the AlGaAs. There are no known deep levels associated with Be in AlGaAs, so that P-n-p HBT's may offer a significant advantage for low noise operation.
Chapter 3
Device Physics

3.1 Introduction

This chapter discusses the operating principles of heterojunction bipolar transistors. Sections 3.2 and 3.3 provide a simple, plain-English description of how heterojunction bipolar transistors work, while a more quantitative treatment is given in the following sections. The fundamental relationships between design parameters (doping density, layer thickness, alloy composition) and performance parameters (current gain, unity current gain cutoff frequency, etc.) are presented, along with a discussion of surface effects.

The design trade-offs between high-speed P-n-p and N-p-n devices are discussed in section 3.8, followed by a description of the fundamentals of strained semiconductor systems.

3.2 Operation of conventional bipolar transistors

Bipolar transistors consist of two n-type semiconductor layers separated by a thin, p-type layer ("n-p-n" transistor), or two p-type layers separated by a thin, n-type layer ("p-n-p" transistor). As this work deals primarily with the latter, the p-n-p transistor will be used for all examples; operation of the n-p-n transistor is analogous.

During normal operation of a p-n-p bipolar transistor, one of the p-n junctions is forward biased, while the other junction is
reverse-biased. Current flows across the forward-biased junction. In a well-designed p-n-p transistor, almost all of the current across the forward-biased junction is carried by holes injected from the p-type layer (the "emitter") into the thin n-type layer (the "base"). These holes diffuse across the base layer until they encounter the reverse-biased p-n junction. The electric field associated with the reverse-biased junction then accelerates the holes into the p-type "collector" layer (Figure 3.1).

![Schematic cross-section of a bipolar transistor showing major components of current flow](image)

While most of the holes injected into the base make it to the collector, some of them recombine with electrons in the n-type base layer. In steady-state, the electrons which recombine must be replaced by electrons supplied externally through the base contact. In addition, a small fraction of the current across the forward-biased p-n junction (the emitter-base junction) is due to electrons being injected from the base into the emitter. These electrons lost from the base must also be replaced by electrons from the base contact.
If most of the current across the emitter-base junction is due to holes injected from the emitter into the base, and if most of these holes diffuse across the base to the collector without recombining in the base, then the number of holes reaching the collector is much greater than the number of electrons which have to be supplied from the base contact. The ratio of the hole current into the collector to the electron current into the base is known as the "common-emitter current gain," or \( \beta \) of the transistor. By controlling the electron current into the base, the relatively large collector current can be modulated.

### 3.3 Operation of heterojunction bipolar transistors

A p-n junction in which the p- and n-type layers are composed of the same semiconductor is known as a homojunction; the junction between dissimilar semiconductors is called a heterojunction. The diagram in Figure 3.2a shows a homojunction bipolar transistor in which the emitter, base, and collector layers are made of the same material.

In heterojunction bipolar transistors (HBT's), the emitter layer is composed of a different semiconductor from the base. The semiconductor for the emitter layer is chosen so that the energy bandgap is greater than that of the base layer semiconductor. If the heterojunction is properly graded (compositional grading of heterojunctions will be discussed in section 3.5), the band diagram will be as shown in Figure 3.2b.
As shown in Figure 3.2, the barriers for hole and electron injection, $\Delta E_p$ and $\Delta E_n$, are equal for the homojunction transistor (Figure 3.2a), while in the HBT they differ by $\Delta E_g$, the bandgap difference between the emitter and the base. The holes injected from the emitter into the base therefore see a lower barrier than do the electrons injected from the base into the emitter. Because current density depends exponentially on the height of the barrier to injection, fewer electrons actually leave the base, resulting in lower base current and higher current gain.
3.4 Quantitative development of HBT principles

Emitter injection efficiency, or $\gamma$, is defined as the ratio of hole current to total current carried by electrons and holes across the emitter base junction (for p-n-p bipolar transistors). Therefore, for an emitter current $I_E$, the hole current into the base is $I_E \gamma$, and the electron current from the base into the emitter is $I_E (1 - \gamma)$.

Base transport factor ("$\alpha_t$") is defined as the ratio of hole current across the emitter-base junction to hole current across the base-collector junction; i.e., the fraction of minority-carrier holes which diffuse across the base without recombining.

Common emitter current gain ("$\beta$") is the ratio of collector current to base current, and may be expressed in terms of $\alpha_t$ and $\gamma$. As discussed before, for an emitter current $I_E$, the undesirable current of electrons injected into the emitter is $I_E (1 - \gamma)$. The hole current injected into the base is $I_E \gamma$. The fraction of these holes which traverse the base without recombining is $\alpha_t$, so the hole current across the base-collector junction is given by $I_E \alpha_t \gamma$. The fraction which recombines is $1 - \alpha_t$, giving a recombination current of $I_E (1 - \alpha_t) \gamma$.

In steady state, the electrons which are lost from the base due to injection into the emitter, or due to recombination with holes diffusing across the base, must be replaced by electrons supplied externally as base current. The base current $I_B$ is just the sum of these two loss currents:
\[ I_B = I_E(1 - \gamma) + I_E(1 - \alpha_t) \gamma = I_E(1 - \gamma \alpha_t) \] (3-1)

and the current gain is

\[ \beta = \frac{I_C}{I_B} = \frac{\gamma \alpha_t}{1 - \gamma \alpha_t} \] (3-2).

To understand the physical mechanisms underlying \( \alpha_t \) and \( \gamma \) one must look at the carrier transport in greater detail.

For the sake of simplicity, the derivations in the rest of this chapter will implicitly assume the low injection condition (minority carrier concentration much smaller than majority carrier concentration) and one-dimensional transport, unless otherwise stated.

### 3.4.1 Emitter injection efficiency (\( \gamma \))

The classic equation describing current flow in a p-n junction diode was first derived by Shockley:\(^{26}\)

\[ J = \left( \frac{qD_p n_{0p}}{L_p} + \frac{qD_n n_{0p}}{L_n} \right) \times \left( e^{qV/kT} - 1 \right) \] (3-3)

where \( J \) is current density, \( q \) is the hole charge (electron charge is \(-q\)), \( V \) is applied bias, \( D_p \) and \( D_n \) are the hole and electron diffusivities, \( L_p \) and \( L_n \) are the hole and electron diffusion lengths, \( p_{n0} \) is the equilibrium minority carrier (hole) density in the base, and \( n_{p0} \) is the equilibrium minority carrier (electron) density in the emitter. The two terms in the equation represent hole and electron diffusion currents, respectively, across the junction.
In the case of a p-n-p bipolar transistor, the injected holes diffuse across the base and are accelerated into the collector over a distance much shorter than \( L_p \), the hole diffusion length. Consequently, it is appropriate to replace \( L_p \) with \( W_B \), the base width, to obtain an expression for emitter current of a homojunction bipolar transistor under normal operating conditions (emitter-base junction forward biased, base-collector junction reverse biased):

\[
I_e = \left( \frac{qD_{np}n_0}{W_B} + \frac{qD_{rp}p_0}{L_n} \right) \times (e^{\frac{-qV}{kT}} - 1)
\]  
(3-4).

As described previously, the current gain of the transistor depends strongly on having most of the current across the junction result from holes injected from the emitter into the base. Since the two terms in the preceding equation represent hole and electron currents, the ratio of hole current to electron current is given by

\[
\frac{I_p}{I_n} = \frac{\frac{qD_{np}n_0}{W_B}}{\frac{qD_{rp}p_0}{L_n}} = \frac{D_{np}n_0}{D_{rp}p_0W_B} \quad (3-5).
\]

Note that \( \gamma \) is related to this current ratio by the expression

\[
\gamma = \frac{\frac{I_p}{I_n}}{1 + \frac{I_p}{I_n}}
\]  
(3-6)

so that, as \( I_p/I_n \) approaches infinity, \( \gamma \) approaches unity. Equation (3-5) may be taken one step further by making use of the fact that,
in equilibrium, the product of the electron and hole concentrations is a constant, $n_i^2$, which is characteristic of the semiconductor. The equilibrium minority carrier concentration is then given by $n_i^2 / N$, where $N$ is the doping concentration. In the general case where the emitter and base may be made of different semiconductors,

$$\frac{I_p}{I_n} = \frac{D_p n_{IB}^2 N_AL_n}{D_n n_{IE}^2 N_D W_B} \left(\frac{n_{IB}^2}{n_{IE}^2}\right)$$

where $N_A$ and $N_D$ are emitter and base doping, and $n_{IE}^2$ and $n_{IB}^2$ are the electron-hole products for the emitter and base, respectively.

The right-hand side of the preceding equation has been rearranged so that the first factor represents the hole/electron current ratio for a homojunction bipolar transistor, since $n_{IE}^2$ and $n_{IB}^2$ are the same in this case (neglecting bandgap shrinkage of the emitter due to heavy doping, which can significantly reduce the gain of homojunction bipolar transistors\textsuperscript{27}). Only the emitter and base doping densities and the base width may be used as design parameters for the homojunction bipolar; once these are specified, the carrier diffusivities and diffusion length are largely determined.

For the heterojunction bipolar transistor, $n_{IE}^2$ and $n_{IB}^2$ are no longer the same. The general expression for $n_i^2$ is
\[ n_i^2 = N_c N_v e^{-\frac{E_g}{kT}} \]  

(3-8)

where \( N_c \) and \( N_v \) are the effective densities of states in the conduction and valence bands, respectively, and \( E_g \) is the bandgap. Therefore,

\[ \frac{n_{IB}^2}{n_{IE}^2} = \frac{N_{CB} N_{VB} e^{-\frac{E_{ge}}{kT}}}{N_{BE} N_{VE} e^{-\frac{E_{ve}}{kT}}} \sim e^{\frac{\Delta E_g}{kT}} \]  

(3-9)

where \( \Delta E_g \) is the difference in energy bandgap between the emitter and base. Equation (3-7) may now be re-expressed to emphasize the role of the heterojunction in bipolar transistors:

\[ \frac{I_p}{I_n} \sim \frac{D_p N_A L_n \Delta E_g}{D_n N_D W_B e^{\frac{\Delta E_g}{kT}}} \]  

(3-10)

For sufficiently large \( \Delta E_g \), \( I_n \) (electron current from the base into the emitter) will be completely suppressed, virtually independent of either the emitter and base doping or base width. For example, AlGaAs/GaAs HBT's typically use an Al mole fraction of 30%. At this mole fraction, \( \Delta E_g = 0.37 \) eV, so that \( e^{\frac{\Delta E_g}{kT}} = e^{0.026 \times 10^6} \) (at room temperature). With this additional factor of \( 10^6 \), the device designer may choose doping levels and base width freely in order to optimize other aspects of performance, such as speed.

Unfortunately, the Shockley equation for diode current (3-3) does not tell the whole story, since it does not include the effect of
traps in the emitter-base junction. In regions relatively depleted of carriers (such as p-n junctions), where screening by large numbers of majority carriers no longer occurs, such traps are effective recombination sites. This effect is well known in silicon bipolar transistors, causing $\beta$ to degrade at low current levels. At higher current levels, the trap recombination current becomes less important, and equation (3-3) accurately describes the current flow.

Because of the higher concentration of mid-bandgap traps in GaAs and AlGaAs compared to Si, the base current of GaAs bipolar transistors almost always has a significant component due to base carriers recombining with emitter carriers at such traps, causing $\beta$ to depend on current density at all current levels. In the past, the trap levels which cause this recombination were thought to be entirely due to the exposed surface of the emitter-base junction. While surface recombination certainly contributes to this effect, even devices which have very effective surface passivation still show some dependence of current gain on current density due to intrinsic defects in the AlGaAs-GaAs junction (probably in the AlGaAs). The surface passivation technique described in Chapter 5 allows one to distinguish between intrinsic and surface recombination.

The dependence of $\beta$ on current density may be used to deduce the severity of trap-assisted recombination. Because the traps are near mid-bandgap, the current associated with trap-assisted recombination has a dependence of $I \sim e^{2kT}$ rather than the diffusion
current dependence of \( I \sim e^{\frac{qV}{kT}} \) given by the Shockley equation (3-3). If trap-assisted recombination in the junction is dominating the base current, then the base current will be proportional to \( e^{\frac{qV}{2kT}} \); if such recombination is significant but not dominant, then the base current may be described as being proportional to \( e^{\frac{qV}{nkT}} \), where \( n \) (known as the "ideality factor") is between 1 and 2.

The collector current, however, is not influenced by the trap-assisted recombination current in the emitter-base junction space charge region, but rather is proportional to the diffusion current. The collector current therefore goes as \( e^{\frac{qV}{kT}} \). This leads immediately to the dependence of current gain on current density:

\[
\beta = \frac{I_C}{I_B} \sim \frac{qV}{e^{\frac{qV}{e^{kT}}} n} \sim e^{\frac{qV}{nkT}} \left( 1 - \frac{1}{n} \right) \tag{3-11}
\]

A plot of \( \log(\beta) \) versus \( \log(I_C) \) therefore has a slope equal to \((1 - 1/n)\); the lower the slope, the less the influence of emitter-base junction traps on device characteristics. Figure 3.3 shows this effect for a typical AlGaAs/GaAs HBT with surface passivation; the slope of \( \log(\beta) \) versus \( \log(I_C) \) is approximately 1/2, showing that recombination in the emitter-base space charge region is significant.
Figure 3.3 Plot of log ($\beta$) versus log ($I_C$), showing effect of traps in emitter-base space charge region: current gain improves with increasing current density

3.4.2 Base transport factor ($\alpha_t$)

Once holes are injected past the junction space charge region into the neutral base of a bipolar transistor, they move primarily by diffusion toward the collector side of the base, where the reverse-biased base-collector junction keeps the hole concentration very low. The electron and hole concentrations in a P-n-p HBT under normal operating conditions (base-emitter forward biased, base-collector reverse biased) are given schematically in Figure 3.4.

The fundamentals of base transport for HBT's with uniform base layers are virtually identical to those for homojunction bipolar transistors, which may be found in texts on the subject.\textsuperscript{29} Base transport and device speed can be improved for HBT's by the addition of compositionally graded base layers, since minority carriers then
drift as well as diffuse and minority carrier storage in the base is decreased at a given current level. However, for state-of-the-art HBT's, the effect of base grading on $\alpha_t$ is much less significant than a reduction of surface effects which results from base grading. Consequently, $\alpha_t$ will not be discussed further. Section 3.7 contains a discussion of base grading and surface effects, while device speed issues will be discussed in section 3.6.

![Figure 3.4 Electron and hole concentrations in an HBT under normal operating conditions](image)

3.5 Importance and design of emitter-base grading

The derivation in the preceding section assumed that the emitter-base heterojunction was compositionally graded so that the conduction and valence bands varied smoothly and monotonically from emitter to base. While HBT's with an abrupt heterojunction between the emitter and base may be favored for certain applications (see section 3.5.2), digital circuits will almost always use graded-heterojunction HBT's for two distinct reasons.

First, one of the principle advantages of bipolar transistors for integration is the guaranteed uniformity of threshold voltage.
Such uniformity is very important in obtaining circuits with high levels of integration. While the turn-on voltage of field-effect transistors (FET's) is very sensitive to variations in channel doping and thickness of insulating layers, the threshold voltage of bipolar transistors is controlled by the bandgap of the base layer, and depends only weakly on doping and processing parameters.

If the emitter-base heterojunction is abrupt, then current will be carried predominantly by tunneling through the barrier formed at the discontinuity. Tunneling current is very sensitive to slight variations in the thickness and composition of such barriers. Additionally, even small redistributions of dopants near the heterojunction may move the p-n metallurgical junction into either the wide- or narrow-gap semiconductor, causing substantial changes in the height of the potential barrier (and the threshold voltage).

The second reason that abrupt heterojunctions are not desirable for HBT's for integration is that the resulting discontinuity in the band structure increases the operating voltage. To obtain a given current density, an abrupt heterojunction requires a higher bias voltage than a graded junction. This higher bias voltage means higher power consumption, which is always undesirable, since power consumption is a serious limitation for high speed bipolar devices.

3.5.1 Optimal grading

The grading of the emitter-base heterojunction must be done with care to retain the advantages described in the previous section.
Improper grading may increase recombination in the junction, or cause higher operating voltages.

Consider an undoped sample of GaAs attached perfectly to an undoped sample of AlAs such that the crystal lattice continues undisturbed from one to the other. The band diagram of the interface would look something like Figure 3.5.

![Band diagram for perfect heterojunction between undoped AlAs and undoped GaAs](image-url)

However, if the two semiconductors, rather than being undoped, are doped n- and p-type, then the discontinuity of the interface is superimposed on the electrostatic potential created by a p-n junction. By convention, an uppercase letter is used to denote the wider-bandgap material, while a lowercase letter is used for the narrower-bandgap material. Figure 3.6 shows an abrupt P-n heterojunction (AlAs is doped at $1 \times 10^{18}$ p-type, GaAs is doped $5 \times 10^{18}$ n-type). Rather than the smooth junction depicted in Figure 3.2b, this heterojunction contains discontinuities in both the conduction and valence bands. The discontinuity in the valence band
is of particular concern for a P-n-p HBT, since the "spike" will hinder the injection of holes from the emitter into the base. Note that the bandgap difference between the AlAs and the GaAs is distributed between the conduction and valence bands (for Al content of 45% or less, $\Delta E_C = 0.6 \Delta E_g$ and $\Delta E_V = 0.4 \Delta E_g$; beyond 45%, the bandgap is indirect, and such simple ratios do not apply).

Figure 3.6 Band diagram of an abrupt P-n AlAs/GaAs heterojunction (p-n GaAs homojunction case is in dotted lines)
CHAPTER 3. DEVICE PHYSICS

Compositional grading of the interface -- from AlAs to 
Al\textsubscript{x}Ga\textsubscript{1-x}As to GaAs, with x varying smoothly from 1 to 0 over some 
nonzero distance -- reduces or even eliminates the spike. Linear 
grading is sometimes employed for simplicity's sake. Figure 3.7 
shows the actual calculated band diagram for a graded AlAs/GaAs

![Graph](image)

Figure 3.7 Band diagram for P-n AlAs/GaAs heterojunction linearly graded 
over 300 Å, under 1.2 V forward bias

P-n heterojunction (same doping as for Figure 3.6, with linear 
grading over 300 Å; the junction is shown under typical forward bias 
of 1.2 V). Because the grading width of 300 Å was carefully chosen 
to match the depletion width under forward bias, the spike has been 
removed; however, there is now a slight hump in the valence band
which will tend to confine holes in the space charge region. This is actually the best that can be done with linear grading; a poor choice of grading width will make this situation even worse, causing both excessive operating voltage and carrier trapping. Linear grading cannot provide a monotonic transition of the bands between the emitter and base under forward bias.

The solution for optimal grading was proposed by J. R. Hayes and coworkers at Bell Laboratories. To remove the barrier completely, one wants to create a parabolic potential change through compositional grading which exactly matches the parabolic electrostatic potential change due to the fixed charge of depleted acceptors in the emitter. This compensation of the potentials must occur over the depletion layer width at a forward bias equivalent to the bandgap of the narrow-bandgap material. An analogous method, employed in this dissertation (see Appendix A for details), avoids some ambiguities associated with that of Hayes et al. In this method, the alloy composition at depth \( t \) is given by

\[
\frac{\text{Al content at } t}{\text{Maximum Al content}} = \frac{V(t)}{V_{\text{max}}} \quad \text{where } V(t) = \text{electrostatic potential at } t, \text{ and } V_{\text{max}} = \text{total electrostatic potential across the junction (built-in voltage minus applied voltage) when the junction is forward-biased by an amount equal to the bandgap of the base. Note that the Al grading extends a short distance into the n-type (GaAs) layer in this grading scheme, since the n-type layer is partly depleted under normal operating conditions.}\
\]
Figure 3.8 shows the calculated band diagram for an optimally-graded heterojunction under 1.2 V forward bias. The material parameters are identical to those of Figure 3.7 except for the parabolic grading.

![Graph showing band diagram](image)

Figure 3.8 P-n heterojunction using optimum grading under 1.2 V forward bias (except for grading, same parameters as in Figure 3.7)

As required, the bands are smooth and monotonic between the emitter and base. Note that the entire bandgap difference, $\Delta E_g$, now appears in the conduction band, rather than being distributed between the valence and conduction bands as in Figure 3.6 (the slight bump in the conduction band is where the Al content is around 45%; beyond this point, the semiconductor is indirect bandgap).
Grading of the emitter-base heterojunction, as well as practical execution of grading during molecular beam epitaxy (MBE), is discussed in detail in Appendix A.

3.5.2 The role of abrupt emitter-base heterojunctions

While grading of the emitter-base heterojunction is desirable for digital integrated circuits, such grading does not necessarily yield the highest performance (in terms of either gain or speed) from individual devices. One of the original motivations for making HBT's with graded junctions was the expectation that current gain would be increased because of an increase in $\gamma$ (emitter injection efficiency). As was shown in the preceding section, proper grading increases the effective barrier height for electron injection from the base into the emitter (for P-n-p HBT's) to the full bandgap difference between the base and emitter.

In practice, however, the increase in $\gamma$ is offset by an increase in recombination current in the emitter-base space charge region. For a given trap density, the Shockley-Read-Hall recombination rate is proportional to the intrinsic carrier density $n_i$.

The value of $n_i$ varies exponentially with bandgap (equation 3-8). Since grading lowers the bandgap over a large portion of the space charge region compared to the abrupt heterojunction case (Figure 3.9), the recombination current is significantly increased. Reduced current gain for graded-heterojunction HBT's has been reported.
Figure 3.9 Comparison of graded and abrupt emitter-base heterojunctions, showing the significant reduction of bandgap (and therefore increased recombination current) for the graded case.

A more intriguing (but less likely) explanation for higher current gain observed in abrupt heterojunction HBT's is hot-carrier transport. The discontinuity in the bands may serve as a filter for energetic carriers, since an excess of energy is required for a carrier to surmount the barrier. Once past the barrier, these energetic carriers might maintain higher than average energy (corresponding to higher than average velocity) as they cross the base. Because transit of the base requires less time, fewer of the carriers recombine, resulting in higher current gain.

While the hot-carrier hypothesis is often cited for improved current gain of abrupt heterojunction HBT's, the previously-described increase in recombination in the space charge region is
frequently overlooked. One method of distinguishing between the two mechanisms is a comparison of the cutoff frequencies \( f_T \) of devices which are identical except for the emitter-base heterojunction. Apparently, no such comparison has been performed as of this writing. If hot carriers do maintain a higher-than-average velocity across the base, then even HBT's which employ heterojunction grading may benefit from the inclusion of a slight discontinuity at the edge of the base depletion region.

3.5.3 The Light Hole P-n-p (LHPNP)

An effect reminiscent of the hot-carrier hypothesis of the preceding section, but quite distinct and peculiar to P-n-p HBT's, may occur in devices with abrupt emitter-base heterojunctions. In GaAs and related alloys, the top of the valence band is actually a superposition of two bands with different effective masses. The macroscopic hole mobility is actually a weighted average of the mobilities of the holes with low effective mass ("light holes") and those with high effective mass ("heavy holes"). Because the probability of quantum mechanical tunneling has an exponential dependence on effective mass, thin barriers act as mass filters, allowing the higher mobility light holes to penetrate. If the time required for light holes to redistribute themselves into the heavy hole band is longer than the base transit time, then a P-n-p HBT with a thin barrier at the emitter edge of the base may be an extremely fast device -- significantly faster than N-p-n HBT's. With a short base transit time similar to that of the N-p-n due to the low
effective mass of the light hole (only 20% greater than that of electrons) and the low base resistance of the P-n-p, the Light Hole P-n-p (LHPNP) transistor would have high speed performance far outweighing the penalty of increased operating voltage due to the barrier. To the best of my knowledge, the LHPNP device has not been described prior to this dissertation.

There is evidence to support the claim of light hole transport across the base of P-n-p HBT's. The one report of high-speed performance of a P-n-p$^3$ used a device with an abrupt heterojunction and a 1000 Å base doped at 3 \times 10^{18}/cm^3. At this doping level, the hole mobility is approximately 120; by the Einstein relation, the hole diffusivity is $D_p = \frac{\mu_p k T}{q} \approx 3$. The base transit time should therefore be 

$$\tau_B = \frac{W_B^2}{2 D_p} = \frac{\left(1000 \times 10^{-8}\right)^2}{2 (3)} = 17 \text{ psec},$$

putting an upper limit on $f_T$ of $f_T \leq \frac{1}{2\pi \tau_B} = 10 \text{ GHz}$ (see section 3.6 for discussion of base transit time and $f_T$). However, the reported $f_T$ for this device was 19 GHz, indicating a base transit time much below 8 psec. This lends strong credibility to the idea of LHPNP devices.

### 3.6 Device speed -- $f_T$ and $f_{\text{max}}$

Two figures of merit are commonly used in quantifying the speed performance of heterojunction bipolar transistors: $f_T$ and $f_{\text{max}}$. 
The term $f_T$ refers to the frequency at which a transistor's current gain drops to unity, and is often discussed in the context of transistor switching speed, as in digital integrated circuit applications. The maximum frequency of oscillation, or $f_{\text{max}}$, represents the frequency at which the power gain drops to unity; it is more relevant for microwave applications.

The cutoff frequency $f_T$ is simply related to $f_{\text{max}}$ by the expression

$$f_{\text{max}} = \sqrt{\frac{f_T}{8 \pi r_B C_{\text{BC}}}} \tag{3-12}$$

where $r_B$ is the base resistance and $C_{\text{BC}}$ is the base-collector junction capacitance. Because the two terms are simply related, and $f_T$ has an intuitively satisfying description based on delay time arguments, only $f_T$ will be discussed at length in this dissertation.

The delay time description of $f_T$ is based on summing of various delay components during the switching of a bipolar transistor. The total delay, $\tau_{\text{EC}}$ (for emitter-collector delay time), is related to $f_T$ by

$$f_T = \frac{1}{2 \pi \tau_{\text{EC}}} .$$

The delay components comprising $\tau_{\text{EC}}$ arise from actual transit times of carriers (holes, for p-n-p devices) through the device, plus delays caused by the finite charging time of junction and parasitic capacitances:

$$\tau_{\text{EC}} = \tau_{E^+} + \tau_{B^+} + \tau_{C^+} + \tau'_C \tag{3-13}.$$
The term $\tau_E$ represents the time required to charge the capacitances associated with the emitter-base and base-collector junctions, as well as any parasitic capacitance connected to the base:

$$\tau_E = R_E(C_{EB} + C_{BC} + C_P) \quad (3-14).$$

In this expression, $R_E$ represents the total emitter resistance, which is composed of the series resistance $r_E$ and an effective resistance of $\frac{kT}{q I_E}$ arising from the current-voltage characteristic of p-n junctions,\(^\text{35}\) where $k$ is Boltzmann's constant, $T$ is temperature, $q$ is hole charge, and $I_E$ is the emitter current. Because $R_E$ decreases as $I_E$ increases, $\tau_E$ is reduced at high current levels.

$\tau_B$ is the time required for the carriers to cross the neutral base region. Assuming that the carriers are not velocity-saturated, the base transit time is given by

$$\tau_B = \frac{W_B^2}{\eta D_B} \quad (3-15)$$

with $W_B$ the width of the neutral base region and $D_B$ the minority-carrier diffusivity in the base. The parameter $\eta$ is determined by the composition of the base layer; for base layers with uniform doping and composition, $\eta = 2$. If the doping density or the bandgap of the base is graded to produce a built-in field of magnitude $E_{bi}$, then $\eta$ is given by
\[
\eta = 2 \left[ 1 + \left( \frac{E_{bi} \mu_B W_B}{2 D_B} \right)^2 \right]^{3/2}
\] (3-16)

with \( \mu_B \) the minority carrier mobility in the base.

If the base layer is sufficiently thin, or \( E_{bi} \) is large enough, then equation (3-15) no longer holds due to velocity saturation of the carriers. In these cases, the base transit time is given by

\[
\tau_B = \frac{W_B}{2 v_{sat}}
\] (3-17)

where \( v_{sat} \) is the saturated velocity of the carrier. The factor of 2 in the denominator arises because carriers crossing the high-field base region induce a displacement current before the carriers actually arrive. The larger of the values of equations (3-15) and (3-17) should be used to calculate the base transit time \( \tau_B \).

The third term in the delay time expression in equation (3-13) is the time required for the carriers to cross the base-collector depletion region. Since the electric field in the depletion region is high, the carriers are assumed to travel at the saturated velocity, so

\[
\tau_C = \frac{W_{dep}}{2 v_{sat}}
\] (3-18)

with \( W_{dep} \) = width of base-collector depletion region.

The final term in equation (3-13) represents the R-C time constant of the collector series resistance \( r_C \) and the collector
capacitance $C_{BC}: \tau'_{C} = r_{C} C_{BC}$. While this term is often neglected because $r_{C}$ can be made small in epitaxially-fabricated transistors, it should be included in calculations for P-n-p HBT's because of the higher resistivity of the p-type collector.

The HBT's fundamental advantage over conventional bipolar transistors comes from the ability to dope the base more heavily than the emitter while maintaining good current gain. Because the emitter need no longer be heavily doped (typical emitter doping in HBT's is $5 \times 10^{17}$/cm$^3$), the emitter-base capacitance ($C_{EB}$) may be reduced, so that $\tau_{E}$ is reduced according to equation (3-14). More importantly, the base resistance $r_{B}$ is lower, so $f_{\text{max}}$ is further reduced (equation 3-12). Because the base doping density is very high, high-level injection effects, such as conductivity modulation or the Webster effect,$^{36}$ occur at much higher current levels, and in practice are never observed for AlGaAs/GaAs HBT's.

However, the preceding analysis of $\tau_{EC}$ breaks down at sufficiently high current densities for another reason. When the density of minority carriers leaving the base and entering the collector approaches the collector doping density, the electric field in the collector is decreased. At this point, the $\tau_{C} = \frac{W_{\text{dep}}}{2 v_{\text{sat}}}$ term of equation (3-18) is replaced by a diffusion-like term; in effect, the base has become wider (the Kirk effect$^{37}$).

The current density at which base widening begins is determined by the collector doping density and the carrier saturated
velocity: \( J_{bw} \approx q v_{sat} N_{coll} \). Base widening therefore occurs at the same current density for homojunction and heterojunction bipolar transistors which have the same saturated velocity and collector doping. At the onset of base widening, however, the HBT will still be capable of higher speed operation than the homojunction device.

3.7 Surface effects

The most important limitations on the performance of GaAs HBT's are related to the electronic behavior of the GaAs surface. Large numbers of Ga and As vacancies near the surface act as donor and acceptor levels in the bandgap, so that the Fermi level at the surface is pinned in the bandgap, regardless of the doping density. The Ga vacancy donor level is approximately 0.5 eV above the valence band maximum, while the As vacancy acceptor level is approximately 0.75 eV above the valence band maximum.\(^{38}\) The band diagram of the GaAs surface with p- and n-type doping is given in Figure 3.10.

Note that in both cases, the electric field is in a direction which accelerates minority carriers toward the surface. Because of this accelerating field, plus the high density of mid-bandgap states which act as recombination sites, the surface recombination velocity of minority carriers in GaAs is very high, on the order of \(10^6\) cm/sec. One method of reducing this electric field is to put a thin, wholly-depleted layer of oppositely-doped material on the surface (i.e., a thin n-type AlGaAs layer on the p-type base). This technique will be examined in Chapter 5.
Interestingly, the potential of the Fermi level at the surface for p- and n-type material is observed to be independent of Al or In content. While the bandgap increases with increasing Al and decreases with increasing In, the Fermi level stays at 0.5 eV above the valence band maximum for p-type material, or 0.75 eV above the valence band for n-type material. Figure 3.11 shows the surface pinning potential for InGaAs as a function of In content. As In content increases, the pinning potential is fixed with respect to the valence band. However, the bandgap decreases so that the conduction band approaches the pinning level, reducing the effectiveness of the vacancies as recombination sites. The InGaAs
surface therefore has a much lower surface recombination velocity for significant In content.

![Graph](image)

*Figure 3.11 Pinning of the Fermi level at the surface of InGaAs, as a function of In content (after H. H. Wieder, *Applied Physics Letters*, vol. 38, pp. 170-171, 1980).*

### 3.7.1 Emitter size effect

As discussed in section 3.6, bipolar transistors yield maximum speed performance at high current densities. Since low power consumption is required for these transistors to be used in integrated circuits, ideally one would like to make the transistors as small as possible. Unfortunately, as lateral device dimensions are scaled down to sizes of a few microns, current gain falls off dramatically in GaAs-based bipolar transistors unless special care
is taken to prevent this. This degradation in gain is known as the emitter size effect.\textsuperscript{40}

Figure 3.12 shows typical results for HBT’s with various emitter sizes.\textsuperscript{40} As the square emitter is scaled down from $48 \times 48$ $\mu$m$^2$ down to $6 \times 6$ $\mu$m$^2$, the current gain drops from 130 down to approximately 40. Note that a $6 \times 6$ $\mu$m$^2$ emitter is still relatively large compared to the dimensions one would want for low power and integration.

Figure 3.12 Degradation in current gain as lateral dimensions of HBT in GaAs are reduced (after Nakajima et. al., \textit{Japan. J. Appl. Phys.}, Vol. 24, pp. L596-L598, 1985).
The data for these and other HBT's indicate that while the collector current scales with emitter area as dimensions are reduced, the base current contains a component which scales with the length of the emitter perimeter. Since the area decreases more quickly than the perimeter as dimensions are reduced, the gain falls off at small dimensions.

Figure 3.13 shows schematically the three base current components which scale with emitter perimeter (while they are depicted as hole currents, the subsequent recombination with electrons shows up as base current). All of these components are lumped together under the heading "emitter size effect."

![Diagram of HBT components]

Figure 3.13 The three components which contribute to emitter size effect:

1) Recombination via surface traps in the emitter-base depletion region
2) Recombination at the exposed base surface
3) Diffusion of holes into the bulk extrinsic base
First, at the edge of the emitter, where there is an exposed vertical surface in the emitter-base junction depletion region, large numbers of traps exist in the bandgap. These traps act as generation-recombination centers, causing a base current component which has an ideality factor of two \( \frac{qV}{e^2 kT} \). An AlGaAs ring surrounding the emitter periphery greatly reduces this surface component.\(^{41}\)

The second current component is due to holes which recombine at the exposed surface of the base. Since the surface of the base is outside of the emitter-base depletion region, this component of the base current has an ideality factor of one \( I \sim e^{1.7 kT} \). This component appears as a dependence of current gain on the width of the passivation layer surrounding the emitter.\(^{42}\)

The final component of base current which scales with emitter perimeter is due to holes which diffuse laterally into the extrinsic base region, where they eventually recombine in the bulk. This component also has an ideality factor of one \( I \sim e^{1.7 kT} \), and is independent of the surface properties of the base. Ion implantation, alloyed contacts which produce metal spiking, or other treatments which reduce the minority carrier lifetime in the extrinsic base region will increase the magnitude of this component of emitter size effect by steepening the minority carrier gradient in the lateral direction. On the other hand, inclusion of a built-in electric field in the base layer or the use of a very thin base will reduce this
component by insuring that most of the holes will eventually be swept into the collector.\textsuperscript{43} To a lesser degree, a built-in field and a thinner base also reduce the severity of the second component (recombination at the horizontal base surface), since the minority carrier concentration near the surface is reduced by the suppressed lateral diffusion.

### 3.7.2 Emitter/collector offset voltage

The common emitter characteristics of GaAs HBT's sometimes exhibit an offset voltage in device turn-on: below a certain value of emitter-collector bias, the collector current is negative. This is due to the collector-base junction having a lower turn-on voltage than the emitter-base junction, so that the collector-base junction begins to have forward conduction first. Figure 3.14 contains an example of this phenomenon.

The origin of this effect originally was unclear, and has been variously attributed to a difference in emitter and collector bandgaps, improper emitter-base grading, and the differing areas of the emitter-base and base-collector junctions. While these factors do cause an offset voltage, HBT's may exhibit such a characteristic even in the absence of these factors. Because HBT's are isolated either by mesa etching or ion implantation, the perimeter of the base-collector junction is marked by a surface or by damaged material, either of which will have a high concentration of mid-bandgap traps. The base-collector junction is almost certain to have a higher number of these traps than the emitter-base junction, since
the base mesa has a wider perimeter, and the more lightly-doped collector results in a wider depletion region in the collector than in the emitter. These traps act as generation-recombination centers, so that the collector-base junction will have a "softer" forward turn-on than the emitter base junction, particularly if the surface is damaged in some way (such as by reactive ion etching at high power) or if the isolation implant parameters are chosen poorly.

![Graph showing the relationship between collector current and emitter-collector bias](image)

Figure 3.14 Example of the emitter-collector offset voltage for a GaAs HBT: the collector current is negative below an emitter-collector bias of approximately 200 mV

Proper emitter-base grading and a fabrication process which does not unnecessarily expose a large portion of the base-collector junction yield HBT's with negligible offset voltage.\(^{30}\)
3.8 Design trade-offs between the P-n-p and the N-p-n

Because of the high mobility and peak velocity of electrons in GaAs compared to holes, attempts to make high-speed HBT's in GaAs have almost invariably used N-p-n devices. Until the work of Sunderland and Dapkus\textsuperscript{6,44} and Hutchby,\textsuperscript{5} the P-n-p's speed potential was not widely recognized. These authors pointed out several reasons why the P-n-p was not necessarily slower than the N-p-n in GaAs:

1) The longer base transit time ($\tau_B$) of the P-n-p, due to the lower hole mobility, was offset by the lower resistivity of the P-n-p's n-type base layer.

2) The saturated velocities of holes and electrons are not as different as their mobilities, being in a ratio of roughly 4:1 rather than 15:1. Those portions of the device where the carriers move at the saturated velocity (the collector depletion region, and, for large built-in base fields, the base region as well) therefore do not contribute such a large relative delay for the P-n-p.

3) The bulk emitter and collector resistances are usually only small parts of the total resistance, making the higher resistivity of p-type material in the emitter and collector regions not as important.

The first of these points is the most significant for design of high-speed P-n-p HBT's in GaAs. For a given base thickness and doping density, the base transit time of the P-n-p is approximately
\( \frac{\mu_n}{\mu_p} \sim 15 \) times higher than that of the N-p-n, while the base sheet resistance is only \( \frac{\mu_p}{\mu_n} \sim \frac{1}{15} \) that for the N-p-n. Consequently, the base thickness can be substantially reduced for the P-n-p to reduce the base transit time without paying a large penalty in base resistance.

Sunderland and Dapkus performed a thorough theoretical comparison of the speed potentials of N-p-n and P-n-p HBT's in AlGaAs/GaAs, although they did not consider the possibility of LHPNP devices, nor did they include scaling or base lifetime effects. They found that, for typical device dimensions, the P-n-p HBT achieved 70 to 89% of the speed attained by the N-p-n device, measured in terms of \( f_T, f_{\text{max}}, \) and switching time. The base width for the optimized P-n-p was generally between 1/2 and 1/3 that of the optimized N-p-n.

These simulations assumed a maximum donor concentration of \( 5 \times 10^{18}/\text{cm}^3 \) and a maximum acceptor concentration of \( 1 \times 10^{19}/\text{cm}^3 \). However, more recent work indicates that these limits may be somewhat low. Si doping of \( 1.6 \times 10^{19}/\text{cm}^3 \) has been obtained.\(^{45}\) Be doping of \( 4 \times 10^{19}/\text{cm}^3 \) in the base has become fairly routine\(^{4,46}\). Rockwell has successfully fabricated HBT's with Be doping of \( 1 \times 10^{20}/\text{cm}^3 \) using a strained InGaAs base layer.\(^{47}\) This extremely high base doping has not been duplicated, and it is not
clear at this time whether such high doping levels can be obtained in a reproducible manner.

If scaling effects are included, then the P-n-p appears to have a significant advantage for IC applications. Since the limits on lateral scaling are roughly proportional to the thickness of the base layer\textsuperscript{40,43} (see section 3.7.1), P-n-p HBT's should function at smaller device geometries than N-p-n devices with similar speed performance, allowing high-speed operation with much lower power dissipation.

3.9 Strained semiconductor systems

As was described in Chapter 2, one of the reasons that the GaAs/AlAs system is widely used for making heterojunction devices is the fact that this is one of the few ternary semiconductor systems in which the lattice constants of the two compounds are well-matched. However, if small amounts of strain can be tolerated in devices, then great flexibility in choice of materials is gained. In addition, the amount of strain may be viewed as an additional degree of design freedom. Anisotropic strain removes the tetrahedral symmetry of GaAs and related alloys, lifting the degeneracy of the light and heavy hole bands in the valence band; hole mobility can therefore be significantly affected by strain. Also, there is at least circumstantial evidence that strain may reduce dopant redistribution.\textsuperscript{47}
In this section, the effects of strain arising from heterojunctions between semiconductors with differing lattice constants will be examined.

3.9.1 Strain and critical thickness

In the InGaAs/GaAs system, because of the larger lattice constant of InGaAs (5.673 Å for 5% In, compared to 5.653 Å for GaAs -- a difference of 0.36%), the InGaAs layer will be compressed in both directions parallel to the growth interface. This is demonstrated schematically in Figure 3.15a.

Figure 3.15 Exaggerated illustration of strain caused by lattice mismatch between InGaAs and GaAs substrate: a) stable strained layer  b) unstable strained layer which causes misfit dislocations
In response to the compression in the two directions parallel to the growth interface, the InGaAs will expand somewhat in the growth direction (perpendicular to the interface). Thus, the lattice constant is smaller than the unstrained lattice constant parallel to the interface, but larger than the unstrained lattice constant perpendicular to the interface.\textsuperscript{48}

If the strained layer is too thick, then misfit dislocations form to relieve the strain in the layer, as shown in Figure 3.15b. In large numbers, such misfit dislocations adversely affect the performance of minority carrier devices such as bipolar transistors.

The foundational work for stability of strained semiconductor systems was performed by Matthews and Blakeslee,\textsuperscript{49} who investigated the propagation of dislocations in thin strained layers as a function of the magnitude of the strain. They found good agreement between experimental observations and a theoretical model based on energy considerations. Other workers have verified the usefulness of this model for the GaAs/InGaAs system.\textsuperscript{50}

In this model, strained layers are unconditionally stable against propagation of misfit dislocations as long as the thickness of the layer does not exceed some critical value which depends on the amount of strain. Figure 3.16 shows a plot of the Matthews-Blakeslee critical thickness as a function of strain, with the case of In\textsubscript{0.05}Ga\textsubscript{0.95}As/GaAs marked.

The stability of such strained layers has important implications for heterojunction devices: as long as the critical
thickness is not exceeded for a given level of strain, the strained layer need not be treated any differently from unstrained material. Device designers are free to use any combination of strain and thickness which does not exceed the critical limit, without concerns about misfit dislocations.

![Graph](image)

Figure 3.16 Matthews-Blakeslee limit for thickness of a strained epitaxial layer (after J. W. Matthews and A. E. Blakeslee, *Journal of Crystal Growth*, vol. 27, pp. 118-125, 1974).

In certain circumstances, the critical thickness can be exceeded by a substantial amount without adverse effects on material quality. Nonequilibrium growth techniques such as molecular beam epitaxy or metalorganic chemical vapor deposition allow growth of metastable structures which will not relax until sufficient energy is applied (such as through high-temperature annealing). Also, compositional grading of the interfaces between
the strained and unstrained material appears to increase the critical layer thickness;\textsuperscript{51} however, it is not known at this time whether this apparent increase in critical thickness is stable or merely metastable.

Elaborations of the Matthews-Blakeslee analysis have been published, clarifying issues such as the effect of the cap layer on critical thickness. The interested reader is referred to the literature.\textsuperscript{52}

3.9.2 Strain effects on band structure and carriers

The valence band structure of GaAs and related compounds is complex, even in the absence of strain. The top of the valence band is actually two degenerate bands with anisotropic effective masses; a spin-split-off band lies at a slightly lower energy (Figure 3.17a).

InGaAs layers on GaAs substrates are under biaxial compression, since the lattice constant of the InGaAs is larger. This biaxial compression may be thought of as being superimposed hydrostatic compression and uniaxial tension (tension in the growth direction, perpendicular to the InGaAs/GaAs interface). The hydrostatic compression increases the bandgap, while the uniaxial tension splits the degeneracy at the valence band maximum (Figure 3.17b).

Ignoring other effects for the moment, the hole effective mass is significantly affected by the splitting of the degeneracy. In the direction parallel to the growth interface, the average effective
mass is reduced compared to the unstrained case; in the perpendicular direction, the average effective mass is increased.

However, the splitting of the degeneracy in the valence band is not the only effect. The anisotropic strain does not merely shift the various subbands, but also changes their shapes in a complex fashion, resulting in further changes in the hole effective mass\(^{48,53}\).

![Diagrams showing band structures](image)

Figure 3.17 Schematic representation of the band structure of a) an unstrained tetrahedral semiconductor with a direct bandgap, and b) the same semiconductor under biaxial compression. CB=conduction band, HH=heavy-hole band, LH=light-hole band, and SO=spin-split-off band.

Most studies of strained InGaAs layers have focused on quantum wells, so that the effects of strain are combined with quantum confinement effects. However, the light hole/heavy hole
splitting due to strain is of the order of 50 meV for a strain of 1% (corresponding to In content of 15%).\textsuperscript{53,54}
Chapter 4
Initial Results for Strained-Base P-n-p Heterojunction Bipolar Transistors

4.1 Preface to Chapter 4

This chapter describes fabrication and experimental results for the first reported P-n-p HBT's with strained layers. While the device results were not spectacular, valuable information was gained through these initial experiments. In particular, the relatively low current gains and poor uniformity emphasized the need for better understanding of the surface passivation mechanism of graded AlGaAs layers, as well as a better-controlled method for producing such passivation layers. The emitter-base heterojunction grading for these devices used chirp grading cycles of 20 Å, rather coarse for P-n-p devices (see Appendix A); hole transport across the emitter-base junction may have been hindered by the individual barriers in the chirp grading, contributing to the low current gain.

4.2 Introduction

Despite the lattice mismatch between AlGaAs/GaAs and InGaAs, successful N-p-N heterojunction bipolar transistors (HBT's) using these materials have been reported. Such devices utilize base layers sufficiently thin that the strain is accommodated elastically. However, the amount of In which is incorporated places an upper limit on the base thickness; In_{0.1}Ga_{0.9}As base layers must
be no more than approximately 400 Å thick, leading to high base resistance for N-p-N HBT's with significant In content.

Due to the much higher mobility of electrons relative to holes, the base resistance of P-n-P HBT's is about an order of magnitude lower for identical base thicknesses and doping levels, allowing the use of a thinner base without the resistance penalty. In addition, as self-aligned process technology decreases lateral dimensions, reduction of the base thickness becomes increasingly important to control lateral carrier diffusion and maintain acceptable gain. P-n-P HBT's, while retaining most of the speed of N-p-N heterojunction bipolar transistors, offer advantages in fabrication: outdiffusion of n-type base dopant is lower at high doping levels, and fast-diffusing acceptors with high solid solubility, such as Zn, may be used to overcompensate the base doping and contact the buried collector in P-n-p structures without a collector contact etch, thus providing a potential planar HBT technology.

4.3 Growth and Processing

Four wafers were grown with varying emitter and base compositions (see Tables 4.1 and 4.2) in a computer-controlled Varian Gen II MBE system. The base was 500 Å wide; the base-emitter junction was parabolically graded over 200 Å to eliminate the heterojunction barrier. To reduce the barrier at the base-collector junction in the InGaAs-base devices, the first 100 Å of the collector next to the base was InGaAs, followed by 150 Å of linear grading to GaAs. Al content of the AlGaAs emitter was 0.3; In
content of the InGaAs base was 0.1. The graded-base devices were linearly graded from 0 to 10% In, resulting in a built-in pseudo-electric field of approximately 2 x 10^4 V/cm. The GaAs growth rate was 1 μm/hr; the substrate temperature was 600^°C for the GaAs collector and AlGaAs emitter. The substrate temperature was reduced to 500^°C during growth of the InGaAs base at a growth rate of 1.11 μm/hr; the heavily-doped GaAs emitter contact cap was grown at 550^°C at 0.5 μm/hr.

**Table 4.1**

<table>
<thead>
<tr>
<th>Epitaxial layer structure</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Contact Cap</td>
<td>50 Å Be 6x10^{19}</td>
</tr>
<tr>
<td>Cap</td>
<td>300 Å Be 1x10^{18}</td>
</tr>
<tr>
<td>Grading</td>
<td>200 Å Be 1x10^{18}</td>
</tr>
<tr>
<td>Emitter</td>
<td>300 Å Be 1x10^{18}</td>
</tr>
<tr>
<td>Grading</td>
<td>200 Å Be 1x10^{18}</td>
</tr>
<tr>
<td>Base</td>
<td>500 Å Si 5x10^{18}</td>
</tr>
<tr>
<td>Grading</td>
<td>250 Å Be 1x10^{17}</td>
</tr>
<tr>
<td>Collector</td>
<td>3000 Å Be 1x10^{17}</td>
</tr>
</tbody>
</table>

Surface recombination significantly reduces the gain of small area HBT's. We have used a thin, depleted AlGaAs mesa guard ring structure to minimize surface recombination (Figure 4.1). This AlGaAs layer (approximately 300 Å thick) was left during wet chemical etching of the emitter mesa. Capacitance-voltage profiling was used to detect the etch endpoint. The thin AlGaAs layer was removed in the base contact area to minimize ohmic
contact resistance. Because of the high doping densities, non-
alloyed metal contacts could be used for emitter and base ohmic
contacts; collector contact was made on the back side of the wafer.
Device isolation was accomplished by mesa etching.

Figure 4.1 Schematic cross-section of P-n-P HBT

4.4 Results and Discussion

Three wafers with AlGaAs emitters all yielded working devices with typical current gains from 3 to 10 (Figures 4.2 and 4.3).
The nonuniformity in gain across a wafer is attributed to variations
in thickness of the thin depleted guard ring layer. The AlGaAs/GaAs
devices and AlGaAs/InGaAs devices with uniform bases had
comparable gains, indicating the high quality of the strained
material; graded-base devices had somewhat higher gains. The
highest observed gain was 35, which was attained by a graded-base
device with a 24 x 30 μm² emitter at I_c = 24 mA.
Figure 4.2  Common emitter I/V characteristics of 24 x 30 \( \mu \text{m}^2 \) graded-base HBT

Figure 4.3  Current gain versus collector current for 24 x 30 \( \mu \text{m}^2 \) graded-base HBT
The GaAs/InGaAs/GaAs devices invariably had gains less than one. For the case of current gain limited by emitter injection efficiency, such a device is expected to have a gain of about 15. Attempts to fabricate GaAs/InGaAs/GaAs P-n-P HBT's with In content ranging from 2 to 13% also produced devices with gain less than one. The mechanism for this apparent reduction in emitter injection efficiency is not understood; however, small discontinuities in the conduction and valence bands in the emitter-base graded region may reduce the flux of relatively massive holes while allowing the less massive electrons to tunnel through.

<table>
<thead>
<tr>
<th>Wafer</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emitter</td>
<td>AlGaAs</td>
<td>AlGaAs</td>
<td>AlGaAs</td>
<td>GaAs</td>
</tr>
<tr>
<td>Base</td>
<td>GaAs</td>
<td>InGaAs</td>
<td>InGaAs-GaAs (graded)</td>
<td>InGaAs</td>
</tr>
<tr>
<td>Collector</td>
<td>GaAs</td>
<td>GaAs</td>
<td>GaAs</td>
<td>GaAs</td>
</tr>
<tr>
<td>Typical gain</td>
<td>5</td>
<td>5</td>
<td>8</td>
<td>&lt; 1</td>
</tr>
</tbody>
</table>

Table 4.2 Comparison of HBT's with different emitter and base compositions

4.5 Conclusion

Strained-base P-n-P HBT's have been demonstrated for the first time using AlGaAs/InGaAs/GaAs, with current gains as high as 35. Despite significant lattice mismatch, current gain was comparable to that of similar AlGaAs/GaAs transistors, indicating
the high quality of the strained material. Comparison with similar GaAs/InGaAs/GaAs devices indicates that an AlGaAs emitter is required for high emitter injection efficiency in a P-n-P HBT with In content of 13% or less.
Chapter 5
Investigation of the Emitter Edge
Passivation Mechanism of Thin AlGaAs Layers

5.1 Introduction

GaAs/AlGaAs heterojunction bipolar transistors (HBT's) are capable of achieving very high current gains because of the wide-bandgap emitter which insures high emitter injection efficiency. However, the scale of integration using HBT's has been limited. At device dimensions suitable for integration, surface effects tend to dominate and current gain degrades considerably. The high density of recombination sites at the surface of GaAs and AlGaAs, coupled with band-bending (due to pinning of the Fermi level in the bandgap at the surface) which accelerates minority carriers toward the surface, lead to a very high surface recombination velocity. Consequently, if the base is exposed along the perimeter of the emitter mesa, then a large component of the base current is due to surface recombination around the emitter periphery in small-dimension HBT's.

One way of reducing this recombination is to leave a thin AlGaAs layer surrounding the active emitter region, as proposed by Lin and Lee. In this scheme, the surrounding AlGaAs layer is wholly depleted due to the interaction of the surface depletion region and the p-n junction depletion region (Figure 5.1). This depleted layer prevents the base surface from being exposed at the emitter periphery, where the minority carrier concentration is large.
Figure 5.1 Passivation of the surface surrounding the emitter mesa by means of a thin, wholly-depleted AlGaAs layer (after H. H. Lin and S. C. Lee, *Applied Physics Letters*, vol. 47, pp. 839-841, 1985)

However, experience has shown that the efficacy of such depleted layers is strongly dependent on layer doping, composition, and thickness; merely insuring that the layer is depleted is not sufficient to control or even reduce surface recombination,
particularly in cases where the heterojunction is graded. This chapter investigates the role of the depleted layer surrounding the emitter, and offers criteria for effective control of surface recombination by means of such layers.

5.2 Analysis

Consider the case of a graded N-p heterojunction (for an N-p-n HBT) with uniform n- and p-type doping of \( N_D \) and \( N_A \), respectively.

![Diagram of a graded N-p heterojunction with AlGaAs layers.](image)

**Figure 5.2** Band diagram for depleted AlGaAs surface passivation layer;

\( x_d \) = width of depleted p+ base layer
If the n-type layer is sufficiently thin that the surface depletion layer extends down to the p-n junction depletion layer, such that the entire n-type layer is depleted, then the band diagram is as shown in Figure 5.2.

Mobile carriers in the depleted regions may be neglected, since they will have little effect on the potential until their density approaches the emitter doping concentration. If the injected carrier density does approach the emitter doping density, then the device performance will degrade for another reason: the emitter injection will suffer in the intrinsic device region.\textsuperscript{57}

Neglecting then the effect of carriers in the depleted regions, as well as the relatively small difference in dielectric constant between GaAs and AlGaAs, the potential profile may be calculated analytically. Set the x-axis origin at the physical p-n junction; take zero potential as the potential of the conduction band in the undepleted p-type region. Let \( x_d = \) width of depleted p-type region (p-type layer is depleted down to \(-x_d\)), and \( x_s = \) width of n-type region (surface is at \(x_s\)). In the depletion approximation, the electric field and potential for \(-x_d \leq x \leq 0\) are given by

\[
E(x) = -\frac{qN_A(x+x_d)}{\varepsilon},
\]

\[
V(x) = \frac{qN_A(x+x_d)^2}{2\varepsilon},
\]

(5-1) (5-2)
where \( q \) is the electronic charge and \( \varepsilon \) is the dielectric constant of the semiconductor. For \( 0 \leq x \leq x_s \),

\[
E(x) = \frac{q}{\varepsilon} (N_A x_d N_D x) \tag{5-3}
\]

and

\[
V(x) = \frac{q}{2\varepsilon} (2N_A x_d x N_D x^2 + N_A x_d^2) \tag{5-4}
\]

In addition to the electric field, there is a pseudo-electric field arising from the change in bandgap across the heterojunction. Let \( \Delta E_c(x) \) represent the conduction band discontinuity which would occur if an abrupt heterojunction were formed between the semiconductor alloys found at \( x \) and \(-x_d\). Then \( V_c(x) \), the potential of the conduction band as a function of \( x \), is given by

\[
V_c(x) = V(x) - \frac{1}{q} \Delta E_c(x)
\]

\[
= \frac{q}{2\varepsilon} (2N_A x_d x N_D x^2 + N_A x_d^2) - \frac{1}{q} \Delta E_c(x)
\]

Since \( \Delta E_c(x) \) depends only on the heterojunction grading, \( V_c(x) \) is uniquely determined by the value of \( x_d \); conversely, if \( V_c(x) \) is known for some value of \( x \), then \( x_d \) is determined. Solving equation (5-5) for \( x_d \) yields

\[
x_d = -x + \sqrt{x^2 + \frac{N_D}{N_A} x^2 + \frac{2\varepsilon}{qN_D} [\Delta E_c(x) - V_c(x)]}
\]

The value of \( V_c(x_s) \) (the conduction band potential at the surface) is known, since the pinning of the Fermi level at the surface of GaAs
and AlGaAs has been studied in detail. The surface Fermi level is usually reported to pin at about 0.5 eV above the valence band for p-type material, and about 0.75 eV above the valence band for n-type material, independent of the Al content.\textsuperscript{38,39} Inserting the appropriate surface potential into equation (5-6) allows us to calculate \( x_d \), which in turn may be used in equation (5-5) to determine the conduction band potential as a function of \( x \).

5.3 Experiment

The devices described in this section were fabricated and characterized by Won-Seong Lee at Stanford.\textsuperscript{46}

Three wafers were grown in a computer-controlled Varian Gen II MBE system with the epitaxial layer structure given in Table 5.1. The structures differed only in the thickness of the uniform AlGaAs emitter layer, with values of 0, 200, and 500 Å. The GaAs-AlGaAs grading on either side of the emitter layer was the same for all wafers: 300 Å linear grading between the AlGaAs emitter and cap layer, and 300 Å parabolic grading between emitter and base to prevent conduction band barriers.\textsuperscript{30}

WSi was deposited as the emitter contact metal. Emitter patterning was accomplished by Al lift-off followed by reactive ion etching of the WSi using the Al as a mask. After removal of the InGaAs with a 3 : 1 : 50 solution of H\(_3\)PO\(_4\) : H\(_2\)O\(_2\) : H\(_2\)O, the GaAs was selectively etched off using a solution of 200 parts H\(_2\)O\(_2\) to 1 part NH\(_4\)OH by volume. The emitter mesa, including the thin AlGaAs layer surrounding the emitter, was defined by a photoresist-masked etch
down to the base using the H$_3$PO$_4$ solution. Ti/Pt/Au was used for the base contact; collector contact was made to the backside of the wafer. Silicon nitride was then deposited and contact holes were etched before application of interconnect metal.

<table>
<thead>
<tr>
<th>Table 5.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contact Cap</td>
</tr>
<tr>
<td>Grading (linear)</td>
</tr>
<tr>
<td>Cap</td>
</tr>
<tr>
<td>Grading (linear)</td>
</tr>
<tr>
<td>Emitter</td>
</tr>
<tr>
<td>Grading (parabolic)</td>
</tr>
<tr>
<td>Spacer</td>
</tr>
<tr>
<td>Base</td>
</tr>
<tr>
<td>Collector</td>
</tr>
<tr>
<td>Subcollector</td>
</tr>
</tbody>
</table>

Table 5.1 Epitaxial layer structure for testing effect of surface passivation layer thickness

5.4 Results

Simulations of these structures using SEDAN\textsuperscript{46} indicate that the emitter injection efficiency at bias levels of interest is not significantly affected by the thickness of the uniform AlGaAs region. Even for the case with no uniform AlGaAs layer (such that the emitter consists only of the parabolic plus linear graded regions), the barrier in the valence band is sufficiently large that
hole injection into the emitter is effectively suppressed. We therefore assume that differences in current gain arise from changes in surface recombination.

Figures 5.3, 5.4, and 5.5 show the calculated band diagrams for the thin AlGaAs region surrounding the emitter for the three processed HBT wafers, as well as the band diagram that would be expected if the base surface were exposed. From the amount of overetch, we estimate that the selective etch stops at about 10% Al. The three wafers are therefore expected to have depleted AlGaAs layers of 500, 700, and 1000 Å surrounding the emitter.

![500 Å AlGaAs passivation layer](image)

Figure 5.3 Band diagram for 500 Å surface passivation layer (y-axis is in volts relative to Fermi level; x-axis is in Angstroms, taking the physical P-n junction as the origin).
The band diagram for the wafer with the 500 Å AlGaAs surface layer (Figure 5.3) shows that the electric field accelerating minority carrier electrons toward the surface is smaller and does not penetrate as far into the base compared to the case of the exposed base.

For the wafer with the 700 Å AlGaAs surface layer, the electric field has been reduced even more (Figure 5.4). In addition,

![Figure 5.4 Band diagram for 700 Å surface passivation layer (y-axis is in volts relative to Fermi level; x-axis is in Angstroms, taking the physical P-n junction as the origin).](image)

the conduction band is now bending upward at the surface. Although electrons in the base layer still are initially accelerated toward the
surface, the band bending tends to confine them in the depleted layer, away from the recombination sites at the surface. Lifetime should be much longer, and therefore current gain is expected to be higher due to lower base current.

The 1000 Å uniform AlGaAs layer in the third wafer is sufficiently thick that the assumption of total depletion is no longer justified; the simple calculation used to derive the diagram in Figure 5.5 is no longer appropriate. In this case, the AlGaAs has a

![1000 Å AlGaAs Passivation Layer](image)

Figure 5.5 Band diagram for 1000 Å surface passivation layer; the simple analysis is no longer valid (y-axis is in volts relative to Fermi level; x-axis is in Angstroms, taking the physical P-n junction as the origin).

substantial electron population and is a conducting layer. Rather than acting as a surface passivation layer, the AlGaAs is effectively
an extension of the emitter, actively injecting electrons into the base. At the edge of this layer where the base is exposed, surface recombination will occur much as it does in unpassivated mesa structure HBT’s.

Current gain characteristics of HBT’s fabricated with different AlGaAs thicknesses support this analysis (Figure 5.6). The AlGaAs

![Graph showing current gain vs. collector current density](image)

Figure 5.6 Current gain dependence on thickness of AlGaAs surface passivation layer surrounding the emitter perimeter; base doping is $5 \cdot 10^{18}$ and emitter doping is $7 \cdot 10^{17}$ (from Won-Seong Lee, PhD dissertation, Stanford University, 1990).

layer extended 10 µm beyond the emitter edge for all devices with surface passivation. The highest overall current gains were obtained from devices with 700 Å AlGaAs passivation layers, while
devices without the AlGaAs layer had the lowest gains. The 500 Å and 1000 Å layers appeared to be less effective in passivating the surface compared to the 700 Å layer.

The current gain of devices with 1000 Å passivation layers rises more quickly with increasing current density than the gain of devices with thinner layers. Because the 1000 Å layer is not depleted, electrons travel through this layer to the perimeter and are injected into the base near the exposed base surface, where the recombination velocity is high. At higher current densities, however, the series resistance of the thin AlGaAs layer becomes significant, and electron injection at the perimeter becomes less important. Consequently, current gain increases more rapidly at higher current densities.

The HBT with current gain of over 12,000 reported by Lin and Lee is an excellent example of surface passivation.\textsuperscript{41} In this design, grown by liquid phase epitaxy (LPE), the Al\textsubscript{0.5}Ga\textsubscript{0.5}As emitter was doped at $1 \times 10^{17}$ and the base doping was $5 \times 10^{17}$. The depleted AlGaAs -passivation layer was approximately 1000 Å thick. The LPE-grown heterojunction is modeled as linear grading of the Al over 300 Å (the band diagram is not greatly changed by grading other than linear).

Applying equations (5-5) and (5-6) to this structure produces the band diagram in Figure 5.7. The effectiveness of this passivation layer is readily apparent. The electric field which would tend to accelerate minority carriers toward the surface is
overwhelmed by the pseudo-electric field associated with the Al grading. The rate of electron recombination at the surface should be much lower than in the bulk base layer, as is borne out by the unusually large observed current gain.


5.5 Discussion

Factors which appear to determine the effectiveness of thin AlGaAs surface passivation layers for HBT's with graded heterojunctions are completeness of depletion, intensity of field
accelerating minority carriers into the passivation layer, and size of
the confining potential barrier at the surface.

Completeness of depletion may be insured by proper choice of
layer thickness and doping. As was demonstrated for the cases of
the 500 and 700 Å passivation layers, slight variations in thickness
of the passivation layer can cause significant variations in device
performance. Surface passivation layers should therefore be defined
selectively for best uniformity.

For N-p-n HBT's, the intensity of the accelerating field and the
size of the confining potential barrier are both improved by
increasing the Al content. Surface passivation in P-n-p HBT's is not
similarly improved, since the Fermi level is pinned above the
valence band by an amount that is relatively independent of Al
content;\textsuperscript{38} however, effective surface passivation can still be
obtained for P-n-p HBT's (see chapter 7). In practice, the Al content
in N-p-n HBT's is constrained by the need to reduce the
concentrations of deep-donors and DX centers associated with Si-
doped AlGaAs; therefore, the effectiveness of surface passivation
must be traded off against material quality and bulk recombination
considerations.

5.6 Conclusion

A simple analytical solution for the conduction and valence
band potentials of thin, completely depleted AlGaAs surface layers
has been presented. Such layers can be used to reduce surface
recombination at the emitter periphery in GaAs/AlGaAs
heterojunction bipolar transistors. Features of the calculated band diagrams correlate well with current gain performance of HBT's fabricated with AlGaAs surface passivation layers of varying doping, thickness, and composition. Current gain of these devices correlates with completeness of depletion, intensity of the electric field accelerating minority carriers into the passivation layer, and size of the confining potential barrier at the surface. This is the first explanation given for the surface passivation mechanism of such layers in the absence of abrupt heterojunctions.

In the past, a lack of understanding of the mechanism has required that optimization of such layers be performed empirically; any change in doping or composition required re-optimization. Solving for the band potentials of the surface layers has provided both an explanation of the passivation mechanism and a tool for designing optimum layers.
Chapter 6
Two Selective Etching Solutions for GaAs on InGaAs
And GaAs/AlGaAs on InGaAs

6.1 Introduction

Selective chemical etching is important in the fabrication of semiconductor devices. Etching solutions for selective removal of GaAs\textsuperscript{58,59} from AlGaAs, and AlGaAs from GaAs\textsuperscript{60,61} have been reported.

Recently, increased attention has been given to strained structures using InGaAs on GaAs: lasers\textsuperscript{62,63} optical modulators\textsuperscript{64} pseudomorphic HEMT's\textsuperscript{65} heterojunction bipolar transistors\textsuperscript{51,55,56,9} and resonant tunneling devices\textsuperscript{66} have been demonstrated. Fabrication of HBT's in particular requires strict control of etch depth and would benefit from selective etching.

Selective dry etching of GaAs and AlGaAs relative to InGaAs has been reported\textsuperscript{67} however, this method requires a reactive ion etch system with vacuum load locks for consistent results.

We give the first report of selective wet etching of GaAs on InGaAs and GaAs/AlGaAs on InGaAs. A mixture of H\textsubscript{2}O\textsubscript{2} and NH\textsubscript{4}OH, commonly used for selective etching of GaAs on AlGaAs\textsuperscript{58,59} is shown to etch GaAs on InGaAs selectively as well. The K\textsubscript{3}Fe(CN)\textsubscript{6}/K\textsubscript{4}Fe(CN)\textsubscript{6} redox system investigated by Tijburg and van
Dongen\textsuperscript{60} may be used for selective etching of both GaAs and AlGaAs with respect to InGaAs.

6.2 Sample Preparation

GaAs, AlGaAs, and InGaAs epitaxial layers were grown on (100)-oriented p-GaAs (Zn-doped) wafers by molecular beam epitaxy. InGaAs layers were 600 Å thick, with 150 to 200 Å grading on either side to reduce misfit dislocations.\textsuperscript{51} The InGaAs composition was determined by measuring the beam flux before MBE growth.

For best etching uniformity, we found it necessary to clean the wafer surface before etching by subjecting the wafers to 200 W oxygen plasma at 100 mTorr for 15 seconds. After oxygen plasma, 4:1 \( \text{H}_2\text{O} : \text{H}_3\text{PO}_4 \) for 30 seconds was used to remove all the GaAs native oxide before the selective etching.

Etching was performed in a 400 ml glass beaker with magnetic stirrer at approximately 150 RPM; the wafer was held approximately 30° to the flow with either stainless steel or Teflon tweezers. No differences were noted using steel or Teflon tweezers. The resulting step heights were measured with a Sloan Dektak IIA or Tencor Instruments Alpha Step.

6.3 Experimental results

For testing of the \( \text{H}_2\text{O}_2/\text{NH}_4\text{OH} \) etch solution, three wafers were grown with nominal In compositions of 5, 10, and 15% In in the
InGaAs layer; a fourth wafer was grown with GaAs only. The epitaxial layer structure for the wafers is given in Table 6.1.

### Table 6.1

<table>
<thead>
<tr>
<th>Thickness (Å)</th>
<th>Layer</th>
<th>Type</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>900</td>
<td>GaAs</td>
<td>Be</td>
<td>(1 \times 10^{18})</td>
</tr>
<tr>
<td>170</td>
<td>grading</td>
<td>Be</td>
<td>(1 \times 10^{18})</td>
</tr>
<tr>
<td>600</td>
<td>InGaAs</td>
<td>Si</td>
<td>(5 \times 10^{18})</td>
</tr>
<tr>
<td>150</td>
<td>grading</td>
<td>Be</td>
<td>(1 \times 10^{17})</td>
</tr>
<tr>
<td>3000</td>
<td>GaAs buffer</td>
<td>Be</td>
<td>(1 \times 10^{17})</td>
</tr>
</tbody>
</table>

Table 6.1  Epitaxial layer structure for testing of \(\text{H}_2\text{O}_2/\text{NH}_4\text{OH}\) etch solution

After growth, 1000 Å SiO\(_2\) was deposited by plasma enhanced chemical vapor deposition to serve as an etch mask, since the \(\text{H}_2\text{O}_2/\text{NH}_4\text{OH}\) solution attacks photoresist. The SiO\(_2\) was etched with 6:1 buffered oxide etch using AZ 1370 SF photoresist as a mask. The photoresist was then stripped in acetone.

The selective etching solution was 250 parts \(\text{H}_2\text{O}_2\) (30% concentrated) to 1 part \(\text{NH}_4\text{OH}\) (30% concentrated) by volume; pH of this solution was 7.30 ± 0.05 at room temperature.

During etching, each sample was periodically removed from the etching solution for measurement of step height; the known thickness of the SiO\(_2\) layer was subtracted to obtain the etch depth.
Etch depth versus time is plotted in Figure 6.1 for the four wafers etched at 10° C. An etch rate of 470 Å/min is obtained for GaAs. For the three samples containing InGaAs, once the InGaAs layer was reached (after approximately two minutes of etching), the etch rate was too low to be observed. The samples with 10 and 15% In showed no further change in etch depth (less than 100 Å additional etching) after 15 minutes of exposure to the etchant.

![Etch depth versus time graph](image)

Figure 6.1 Etching of 1000 Å GaAs layer over InGaAs using \( \text{H}_2\text{O}_2/\text{NH}_4\text{OH} \) solution of pH 7.3 at 10° C

However, the sample with only 5% In began to show slight discoloration after 10 minutes; after 15 minutes, the surface was rough and the etch depth was over 4000 Å. We speculate that a porous oxide film was formed on the surface of this sample and subsequently removed, allowing etching of the underlying GaAs layer. The other wafers maintained a smooth surface at all times.
during the etch; no evidence of oxide film cracking or flaking was observed.

The GaAs etch rate was also measured at 20° C and 30° C; the results are plotted in Figure 6.2. The etch rate was found to vary exponentially with temperature, with an activation energy of 11.8 kcal/mole. Because of the uneven nature of the InGaAs etching, no InGaAs etch rate could be obtained.

![Graph showing etch rate vs. 1000/T(K)](image)

Figure 6.2. Temperature dependence of GaAs etch rate in H₂O₂/NH₄OH solution of pH 7.3

The second selective etching solution was prepared by dissolving 19.0 g K₄Fe(CN)₆·3H₂O and 14.8 g K₃Fe(CN)₆ in 200 ml of H₂O. After dissolution, the pH was lowered by the addition of diluted HCl. For pH above ~7, the wafer surface became pitted during etching; between ~6-7, etching was uniform. This solution
differs from the one reported to etch AlGaAs on GaAs selectively, in that added HCl is used to control the pH rather than a buffer solution.

For the tests described here, the pH was \(6.65 \pm 0.05\); approximately 3 ml of 1000:1 \( \text{H}_2\text{O} : \text{HCl}\) was added to obtain this pH value. The solution was between 23 and 25° C during etching.

**Table 6.2**

<table>
<thead>
<tr>
<th>Thickness (Å)</th>
<th>Material</th>
<th>Doping</th>
<th>Concentration (cm(^{-3}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>GaAs</td>
<td>Be</td>
<td>(5 \times 10^{19})</td>
</tr>
<tr>
<td>300</td>
<td>GaAs</td>
<td>Be</td>
<td>(1 \times 10^{18})</td>
</tr>
<tr>
<td>200</td>
<td>grading</td>
<td>Be</td>
<td>(1 \times 10^{18})</td>
</tr>
<tr>
<td>300</td>
<td>(\text{Al}<em>{0.3}\text{Ga}</em>{0.7}\text{As})</td>
<td>Be</td>
<td>(1 \times 10^{18})</td>
</tr>
<tr>
<td>200</td>
<td>grading</td>
<td>Be</td>
<td>(1 \times 10^{18})</td>
</tr>
<tr>
<td>500</td>
<td>InGaAs</td>
<td>Si</td>
<td>(5 \times 10^{18})</td>
</tr>
<tr>
<td>100</td>
<td>InGaAs</td>
<td>undoped</td>
<td></td>
</tr>
<tr>
<td>150</td>
<td>grading</td>
<td>Be</td>
<td>(1 \times 10^{17})</td>
</tr>
<tr>
<td>3000</td>
<td>GaAs buffer</td>
<td>Be</td>
<td>(1 \times 10^{17})</td>
</tr>
</tbody>
</table>

Table 6.2 Epitaxial layer structure for testing of \(\text{K}_3\text{Fe(CN)}_6/\text{K}_4\text{Fe(CN)}_6\) etch solution

Four wafers containing InGaAs layers with nominal compositions of 5, 10, 15, and 20% In (Table 6.2) were etched in this solution, as well as bulk GaAs, \(\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}\), and \(\text{In}_{0.12}\text{Ga}_{0.88}\text{As}\)
samples. AZ 1370 SF photoresist was used as the etch mask. After each etch, the photoresist was stripped in acetone and the step height was measured. Each data point therefore represents an uninterrupted etch for the specified time.

Figure 6.3 shows the results of the bulk GaAs and Al$_{0.3}$Ga$_{0.7}$As etch tests. GaAs and Al$_{0.3}$Ga$_{0.7}$As were found to etch at about 230 and 160 Å/min, respectively. The Al$_{0.3}$Ga$_{0.7}$As etch profiles were uniform steps. However, the GaAs samples were etched more rapidly at the resist edge, indicating that GaAs etching is limited by mass transport.$^{69}$

![Graph showing etch depth and etch time for GaAs and AlGaAs](image)

Figure 6.3. Relation of etch depth and etch time for GaAs and Al$_{0.3}$Ga$_{0.7}$As in K$_3$Fe(CN)$_6$/K$_4$Fe(CN)$_6$ solution at 25$^\circ$C

Figure 6.4 gives etch depth versus time for the wafers with 5, 10, 15, and 20% In. After 7 minutes of etching, the InGaAs layer was
exposed. Beyond this point, there was no significant etching in the samples with 10% or more In. The sample with 5% In continued to etch, but at a reduced rate. The surface remained specular on all samples except the 5% In sample, which exhibited a uniform, finely-grained texture.

![Figure 6.4. Etching of 1000 Å GaAs/AlGaAs over InGaAs using K₃Fe(CN)₆/K₄Fe(CN)₆ solution at 25° C](image)

A bulk In₀.₁₂Ga₀.₈₈As sample was exposed to the etching solution for 20 minutes. Although no step was measurable after the photoresist was removed, the exposed surface was light brown in color. Following a 30 second dip in 4:1 H₂O : H₃PO₄, the discoloration was removed, and a step height of 300 to 400 Å was measured.
6.4 Application

During fabrication of mesa-structure heterojunction bipolar transistors, exposure of the base layer for the purpose of making ohmic contacts requires precise control of etching. We have used the $\text{K}_3\text{Fe(CN)}_6/\text{K}_4\text{Fe( CN)}_6$ solution in fabricating AlGaAs/In$_{0.05}$Ga$_{0.95}$As/GaAs heterojunction bipolar transistors. Despite the low In content of only 5% in the 500 Å base layer, the etch solution was sufficiently selective to obtain devices with high uniformity (see Chapter 7).

6.5 Conclusion

Selective wet chemical etching of GaAs on InGaAs and AlGaAs/GaAs on InGaAs can be achieved using solutions of $\text{H}_2\text{O}_2/\text{NH}_4\text{OH}$ and $\text{K}_3\text{Fe(CN)}_6/\text{K}_4\text{Fe(CN)}_6$. As little as 5% In is required to reduce the etch rate significantly. InGaAs layers with 10% or more In are effective etch stops for these solutions.
Chapter 7
Uniform, High-Gain AlGaAs/InGaAs/GaAs
P-n-p HBT's by Dual Selective Etch Process

7.1 Introduction

The thinner base of the optimized P-n-p HBT compared to the optimized N-p-n HBT makes current gain degradation associated with small dimensions ("emitter size effect") less severe for the P-n-p (see section 3.8). However, the thin base makes fabrication of mesa-structure devices more difficult.

Even for N-p-n mesa-structure HBT's, the etch to expose the base layer is often the most critical step, since current designs typically have base widths of 1000 Å or less. Before the benefits of the thinner-base P-n-p HBT can be realized, the base exposure etch must be made more reliable.

If a small amount of In is included in the base layer, then the base may be exposed by selective etching using a solution of potassium ferrocyanide and potassium ferricyanide (see chapter 6). This chapter describes fabrication of HBT's using this selective etch. Selective etching is also employed to define a thin, wholly-depleted AlGaAs layer for passivation of the emitter periphery, as has been described previously.\textsuperscript{42}

7.2 Device Fabrication

The epitaxial layer structure of the wafer, grown by MBE on a (100)-oriented p+ substrate, is given in Table 7.1. During growth of
the InGaAs, the substrate temperature was reduced to 500° C; growth of the heavily-doped cap layer was at 550° C. All other material was grown at a substrate temperature of 600° C. The last 250 Å of the collector next to the base consisted of a 150 Å GaAs-InGaAs graded layer, followed by 100 Å of InGaAs to insure that no hole barrier existed between the base and collector. The emitter-base junction was graded parabolically\textsuperscript{30,57} over 200 Å. Although emitter-base grading reduces current gain due to increased recombination in the space charge region, such grading lowers the turn-on voltage of the junction by removing the heterojunction barrier in the valence band.

| Table 7.1 |
| Layer structure |
| Contact Cap | GaAs | 100 Å | Be | 1x10\textsuperscript{20} |
| Cap | GaAs | 300 Å | Be | 1x10\textsuperscript{18} |
| Grading | AlGaAs-GaAs | 200 Å | Be | 1x10\textsuperscript{18} |
| Emitter | AlGaAs | 300 Å | Be | 1x10\textsuperscript{18} |
| Grading | InGaAs-AlGaAs | 200 Å | Be | 1x10\textsuperscript{18} |
| Base | InGaAs | 500 Å | Si | 5x10\textsuperscript{18} |
| Grading | GaAs-InGaAs | 250 Å | Be | 1x10\textsuperscript{17} |
| Collector | GaAs | 3000 Å | Be | 1x10\textsuperscript{17} |

Table 7.1 Epitaxial layer structure of dual selective etch P-n-p HBT's
After lift-off of evaporated Ti/Pt/Au emitter contacts, the wafer was etched in a magnetically-stirred 250 : 1 solution of $\text{H}_2\text{O}_2$ / $\text{NH}_4\text{OH}$ at $10^\circ \text{C}$ for 3 minutes to expose the AlGaAs layer (Figure 7.1a). Photoresist was then patterned to protect the AlGaAs passivation region surrounding each emitter while etching down to the InGaAs base. This AlGaAs passivation region extended from 1 to 3 $\mu$m around the active emitter region.

The base exposure etching solution was 19.0 g $\text{K}_4\text{Fe}(\text{CN})_6 \cdot 3\text{H}_2\text{O}$ and 14.8 g $\text{K}_3\text{Fe}(\text{CN})_6$ dissolved in 200 ml $\text{H}_2\text{O}$ with the pH lowered to 6.9 ($\pm 0.1$) with diluted $\text{HCl}$. The etch rate of $\text{In}_{0.05}\text{Ga}_{0.95}\text{As}$ in this solution is roughly 1/3 that of the AlGaAs emitter. The wafer was etched for 5 minutes at room temperature with magnetic stirring (Figure 7.1b).

Figure 7.1 The two selective etching steps: a) Exposure of AlGaAs using $\text{H}_2\text{O}_2/\text{NH}_4\text{OH}$ solution  b) Exposure of InGaAs away from active emitter edge using $\text{K}_3\text{Fe}(\text{CN})_6/\text{K}_4\text{Fe}(\text{CN})_6$ solution
Base contacts were evaporated Ti/Al; collector contact was made to the back side of the wafer with Ti/Al. Device isolation was accomplished by mesa etching, and silicon nitride dielectric was deposited by PECVD before lift-off of interconnect metal.

7.3 Experimental results

The device characteristics were remarkably uniform, with little device-to-device variation in current gain, and very mild emitter size effect. Despite the 500 Å base layer, yields were consistently above 98%.

7.3.1 Uniformity of current gain

Common-emitter current gain ($b$) of the resulting HBT's was uniformly high, with values over 200 at high emitter current densities. As an example of the uniformity obtained, $b$ was measured for the two most common emitter sizes of 3 x 4 and 3 x 10 $\mu$m$^2$ at an emitter current density of 1 x $10^4$ A/cm$^2$. The current gains of all devices (13 total) fall within 9% of the average value of 73, with a standard deviation of 5.1%. Figure 7.2 demonstrates the uniformity of the current gain at a given current density: the measured characteristics of all devices from two dies, with emitters ranging from 24 x 30 $\mu$m$^2$ down to 3 x 4 $\mu$m$^2$, have been overlaid. Note the small spread in current gain at a given current density, despite the wide range of emitter size.
Figure 7.2 Overlaid incremental current gain characteristics for all HBT's on two dies.

This uniformity in current gain is among the best reported for GaAs HBT’s. The use of selective etching for all critical etch steps, the suppression of surface effects, and the absence of base dopant outdiffusion (a severe problem for N-p-n HBT's) contribute to the superior uniformity.

No differences in current gain were noted for devices with varying widths of the AlGaAs passivation layer; 1 μm-wide ledges appeared to be as effective in surface passivation as 2 or 3 μm-wide ledges.

7.3.2 Emitter size effect

The base current of an AlGaAs/GaAs HBT may be thought of as containing two components: bulk recombination of minority carriers, which scales with emitter area, and surface recombination at the
edge of the emitter, which scales with emitter perimeter. Consequently, the slope of \( b \) versus emitter perimeter/area ratio is a measure of the importance of surface recombination in limiting current gain. \( b \) versus emitter perimeter/area ratio is plotted in Figure 7.3 for several different emitter sizes at constant emitter current density of \( 1 \times 10^4 \) A/cm\(^2\). Each data point represents the average of several devices of the same dimensions.

For perimeter/area ratios of 0.8 \( \mu \text{m}^{-1} \) or less, surface recombination has little influence on current gain. In fact, the current gain of HBT's with 21 x 3.5 \( \mu \text{m}^2 \) emitters is actually slightly higher than that of HBT's with 24 x 30 \( \mu \text{m}^2 \) emitters. For perimeter/area ratios above 0.8 \( \mu \text{m}^{-1} \) (representing devices with emitter dimensions of approximately 3 x 4 \( \mu \text{m}^2 \) and 3 x 10 \( \mu \text{m}^2 \)), the gain has dropped roughly 20% compared to the largest devices.

![Graph showing current gain vs. emitter perimeter/area ratio]

Figure 7.3 Emitter size effect for P-n-p HBT's with 500 Å base
The AlGaAs surface passivation layer surrounding the emitter significantly improves the current gain as shown in Figure 7.4. Not only is the current gain higher, but the emitter size effect is less severe for the passivated devices, as shown by the lower slope of the data points for passivated devices compared to unpassivated devices.

![Figure 7.4 Effect of surface passivation on current gain and emitter size effect for P-n-p HBT's](image)

The emitter size effect is less severe for devices with thin base layers, as discussed in section 3.7.1. The significance of this advantage for thin-base HBT's is demonstrated in Figure 7.5. The N-p-n devices have a 1000 Å base, while the P-n-p devices have a 500 Å base, so that these designs should have roughly comparable high speed performance. When the surface passivation is removed from the N-p-n HBT's with $6 \times 6 \, \mu m^2$ emitters (perimeter/area ratio
of 0.667/µm), the current gain drops by roughly a factor of 6; however, the corresponding drop in current gain for P-n-p HBT's of similar dimensions is less than 40%. For very small geometries, or for self-aligned structures where surface passivation is very difficult, the P-n-p HBT is clearly superior.

![Graph showing current gain vs. Emitter Perimeter/Area ratio](image)

Figure 7.5 Comparison of emitter size effect for N-p-n and P-n-p HBT's, with and without surface passivation (data for N-p-n devices provided by Won-Seong Lee)

### 7.3.3 Yield

The use of selective etching, both for definition of the AlGaAs surface passivation layer and for exposure of the base layer, resulted in yields of 98.5% for over 500 devices probed on four different wafers. Despite the very thin 500 Å base layer, all failed devices appeared to be due to gross lithographic defects. These defects resulted from a process which was not yield-optimized, as well as a research-oriented laboratory environment. Elimination of
a step where household aluminum foil is placed directly on the photoresist-covered wafers (for a bulk exposure to produce some unpassivated devices on each wafer) would undoubtedly produce higher yields.

7.3.4 Material quality considerations

Figure 7.6 shows base and collector current as a function of base-emitter bias at a constant emitter-collector bias of -3 V; emitter size is 3 x 10 μm². The strong dependence of the gain on current density is due to recombination in the emitter-base space charge region. The collector current has an ideality factor of 1, as

![Graph of base-emitter bias vs. current](image)

Figure 7.6 Base and collector currents versus base-emitter bias for P-n-p HBT at a constant emitter-collector bias of -3 V (emitter size is 3 x 10 μm²).
expected for diffusion current (see section 3.4.1). However, the base current has an ideality factor close to 3, indicating very high recombination current in the emitter-base space charge region. This recombination is probably due to the overlap of In and Al grading in the heterojunction, since GaAlInAs grown by MBE tends to have a high density of traps. The decrease in gain at high current levels is due to thermal effects.7

An improved grading design with InGaAs and AlGaAs separated by a thin GaAs layer reduces the trap density and improves the ideality factor of the base current. Figure 7.7 shows the dependence of current gain on current density for an HBT structure with the In and Al separated in the emitter-base grading layer. The ideality

![Graph showing collector current vs current gain with n=3 behavior](image)

Figure 7.7 Improvement of base ideality factor from 3 to 2 by separation of

In and Al in the emitter-base space charge region
factor for the base current has been reduced to 2, indicating that the In-Al trap has been eliminated. Since AlGaAs/GaAs HBT’s with base current ideality factors of 1.1 are routinely obtained by MBE, other traps are still present. This indicates that the material quality (and therefore device current gain) can still be substantially improved.

One known source of material quality problems during growth of these wafers was a contaminated indium source. InGaAs quantum wells grown with the In source used for these HBT wafers had very poor photoluminescence characteristics. Photoluminescence, or PL, is a semi-quantitative technique for measuring the density of carrier traps; the higher the trap density, the lower the PL intensity. Compared to InGaAs quantum wells previously grown in the same MBE growth chamber (but with a different In source), the PL intensity of InGaAs quantum wells grown with the contaminated In source was approximately 1000 times lower, indicating very low minority carrier lifetime.

This lowered lifetime was also evident from HBT characteristics. Devices fabricated on wafers with GaAs base layers (no In) exhibited maximum current gains over 450, despite emitters as small as 2.5 x 4 μm². However, current gain dropped rapidly and uniformly as In content increased from 0 to 20% (in increments of 5%).
CHAPTER 7. UNIFORM, HIGH-GAIN ALGaaS/INGAAS/GAAS P-N-P HBTS...

This decrease in current gain is not the effect of strain; current gain should be independent of In content until the critical strain is exceeded, at which point the current gain drops off drastically due to large numbers of misfit dislocations. This independence of current gain on In content has been observed experimentally for N-p-n HBT's.\textsuperscript{56}

7.4 Conclusion

High-gain P-n-p HBT's with excellent gain, uniformity, and yield have been demonstrated in AlGaAs/InGaAs/GaAs. A fabrication process relying on selective etching for formation of the AlGaAs surface passivation layer, and also for exposure of the base, has produced devices with standard deviation of current gain of 5%, possibly the best yet obtained for GaAs HBT's.

Despite the very thin 500 Å base, a yield of 98.5\% was obtained; this yield was limited by gross lithographic defects common in a research-oriented laboratory environment. Very high yields should be obtainable in a production facility.

Gain was not strongly dependent on device size, indicating that surface effects do not dominate device performance, and that high-gain devices should be easily obtained at emitter sizes of 1 × 1 \textmu m\textsuperscript{2} or smaller.

The high uniformity and yield, as well as the very small dimensions which can be reached while maintaining high current gain, make the P-n-p HBT a very attractive candidate for larger scales of integration.
Chapter 8
Conclusion

8.1 Summary

This dissertation has investigated the DC characteristics of P-n-p heterojunction bipolar transistors in AlGaAs/InGaAs/GaAs. While there is much interest in P-n-p HBT's for complementary circuits, P-n-p devices with adequate uniformity and current gain characteristics have previously been lacking.

However, P-n-p HBT's are not only of interest because of their requirement for complementary circuits; the P-n-p's potential for high speed has generated interest as well.

While the high-speed potential and several significant technological advantages of the P-n-p have been recognized, this dissertation contains the first proposal as well as demonstration that P-n-p HBT's have superior lateral scaling properties. As discussed in chapter 3, while the optimized N-p-n device may be roughly 15% faster than the optimized P-n-p device for a given device geometry, the optimized P-n-p will have a thinner base layer due to the greatly lower resistivity of n-type GaAs. Since the severity of the emitter size effect roughly scales with base thickness, the P-n-p can be scaled down to smaller lateral dimensions before surface effects begin to dominate. Consequently, the slight speed disadvantage of the P-n-p is outweighed by the fact that very small devices -- probably 1 x 1 \( \mu \text{m}^2 \) or smaller -- can be
made without the severe degradation in current gain experienced by N-p-n HBT's at such dimensions. The smaller emitter area leads directly to reduced power consumption at high speed, allowing an advantageous trade-off between power and delay.

The scaling advantage of the P-n-p is significant. Using only the results actually demonstrated in this thesis (2.5 x 4 \( \mu \text{m}^2 \) emitter HBT's with current gain of 100 at 100 \( \mu \text{A} \)), a chip with a 1 W power dissipation could contain over 10,000 HBT's; scaling down to 1 x 1 \( \mu \text{m}^2 \) increases that number to 100,000.

The uniformity of P-n-p HBT's is greater than that of N-p-n devices as well, due to the N-p-n's problem with outdiffusion of the base dopant during epitaxial growth. The current gain uniformity of the devices reported in this dissertation is among the best reported for GaAs HBT's.

The issue of yield for thin-base devices has been resolved through the use of selective etching for exposure of the base layer. The first reported wet chemical etching solutions for selective etching of GaAs and AlGaAs/GaAs with respect to InGaAs (chapter 5) produced device yields above 98%, limited by lithographic defects rather than by etching. These selective etching solutions have applications beyond HBT fabrication; many devices can benefit from selective etching of GaAs/AlGaAs with respect to InGaAs. For example, fabrication of resonant tunneling transistors employing InGaAs quantum wells should be substantially easier using selective etching.
The P-n-p HBT's in this dissertation are the first P-n-p devices using strained layers. The selective etching used for exposure of the base layer was made possible by the inclusion of small amounts of In, making the use of InGaAs in the device very attractive despite the lattice mismatch to GaAs. This is an excellent demonstration of the design flexibility to be gained if device designers are willing to work with strained systems.

Finally, this dissertation provides the first explanation of the surface passivation mechanism of thin AlGaAs layers. In the past, lack of understanding of the mechanism has required that optimization of such layers be performed empirically; any change in doping or composition required re-optimization. However, solving for the band potentials of the surface layers has yielded both an explanation of the passivation mechanism and a tool for designing optimum layers.

8.2 Suggestions for future work

The results of this work indicate that P-n-p HBT's with epitaxial structures capable of high speed give very good DC performance, with promise of even higher performance if material quality is improved. Several areas of investigation are promising:

- The limits on scaling of P-n-p HBT's should be investigated. Despite the material quality limitations in this dissertation, HBT's with 2.5 x 4 μm² emitters were obtained with current gains in excess of 450; current gain of 100 was obtained at a current density of 1 x 10³ A/cm².
Since a current gain of 50 is sufficient for many applications (including digital IC's), HBT's with emitter dimensions of 1 x 1 μm² or even smaller seem entirely reasonable. At 1 x 10³ A/cm², a 1 x 1 μm²-emitter HBT would consume approximately 10 μW when turned on, so that a 1 W chip could contain 100,000 or more HBT's.

- The high-frequency performance of optimized P-n-p HBT's should be measured. The only high-frequency measurements on P-n-p's to date were for devices which were not optimized for speed, but rather were based on N-p-n designs. The base layer was therefore too thick, and had no compositional grading.

- The 1/f noise of P-n-p HBT's should be measured to verify whether elimination of DX centers allows lower-noise operation compared to N-p-n devices.

- The possibility of light-hole P-n-p devices (LHPNP's) should be investigated (section 3.5.3). If proven feasible, LHPNP transistors will have significantly higher values of f_max than either N-p-n or P-n-p HBT's, making them very attractive for applications requiring low levels of integration (higher levels of integration are less attractive because of the higher operating voltage and wider spread in turn-on voltage).

- Small-scale complementary circuits using the demonstrated P-n-p process should be undertaken.
Low-resistance ohmic contacts are required for the base contacts of HBT's, so much work has been done for contacting thin p-type layers. However, ohmic contact technology for thin n-type layers (below 2000 Å) is a relatively unexplored area. At least two options should be investigated:

The metal spiking of AuGe/Au contacts is not as severe when the Ni conventionally used is omitted. Although the resistance is higher without the Ni, AuGe/Au has been used to contact n-type layers as thin as 500 Å.\textsuperscript{72}

Pd/In/Ge gives ohmic contacts with very low specific resistivity ($10^{-6}$ Ω-cm$^2$ or better), but the penetration depth has not been studied.\textsuperscript{73}

As an alternative to the alloyed contacts described above, inverted (emitter-down) HBT's with graded InGaAs base should be investigated. The emitter side of the base would be GaAs, graded to InGaAs with 30% or more In next to the collector. Such a structure offers several advantages. First, the etch down to the base could be very selective. Second, the contact between n-type InGaAs and practically any metal is ohmic with very low resistance if the In content is high enough. Third, the surface recombination velocity of the exposed InGaAs base would be much lower than for GaAs. Because the In content would be continuously graded, the critical thickness of such a layer
is impossible to predict. However, 200 Å is a conservative estimate; 300 to 400 Å layers may be stable enough for device purposes.
Appendix A
Grading of the Emitter-Base Heterojunction

In applications where obtaining uniform threshold voltage and/or the lowest possible operating voltage are important, the emitter-base heterojunction must be graded in such a way that the conduction and valence bands make a smooth, monotonic transition from emitter to base at all bias levels. Bumps or spikes resulting from discontinuities will hinder the flow of carriers (especially the more massive holes, which do not tunnel easily), resulting in higher operating voltages and sensitivity to slight variations in growth conditions; dips in the bands will tend to confine carriers, causing more space charge recombination. In addition, the grading should take place entirely within the region which is depleted at the operating bias; otherwise, less than the full bandgap difference is being utilized. A more complete description of the emitter-base heterojunction grading has been given by Tiwari and Frank.57

Mere linear grading does not insure monotonic transition of the bands, as was demonstrated in section 3.5.

A solution for optimal grading was proposed by J. R. Hayes and coworkers at Bell Laboratories,30 who recommended grading the composition according to the "complementary function of the electrostatic potential over the depletion layer width at a forward bias equivalent to the bandgap of the narrow-bandgap material." An analogous method, employed in this dissertation, avoids some ambiguities associated with that of Hayes et. al. In this method, the alloy composition at depth t is given by a well defined function:
\[ \frac{\text{Al content at t}}{\text{Maximum Al content}} = \frac{V(t)}{V_{\text{max}}} \]

where \(V(t)\) = electrostatic potential at depth \(t\), and \(V_{\text{max}}\) = total electrostatic potential across the junction (built-in voltage minus applied voltage) when the junction is forward-biased at the operating point (typically 1 to 1.3 V).

The procedure for designing such an optimally graded heterojunction begins with a calculation of the potential as a function of depth for a homojunction with doping identical to that for the heterojunction case. The potential should be calculated at the bias condition for which optimal grading is desired.

For a built-in voltage of \(V_{\text{bi}}\) (approximately equal to the base bandgap for doping levels of interest for HBT's) and applied bias of \(V_{\text{app}}\), the width of the depleted p- and n-type regions is given by

\[ w_p = \sqrt{\frac{2 \varepsilon (V_{\text{bi}} - V_{\text{app}} - \frac{kT}{q})}{q N_A \left(1 + \frac{N_A}{N_D}\right)}} \]

and

\[ w_n = \sqrt{\frac{2 \varepsilon (V_{\text{bi}} - V_{\text{app}} - \frac{kT}{q})}{q N_D \left(1 + \frac{N_D}{N_A}\right)}} \]

For the case of a P-n-p HBT, if the metallurgical junction is taken as the origin, then the base is depleted down to \(-w_n\), while the emitter is depleted to \(+w_p\), as shown in Figure A.1.
The electric field between \(-w_n\) and \(w_p\) is given by

\[
E(t) = \frac{q}{\varepsilon} N_D (w_n + t) \quad \text{for} \quad -w_n \leq t \leq 0
\]

and

\[
E(t) = -\frac{q}{\varepsilon} N_A t + E(0)
\]

\[= -\frac{q}{\varepsilon} N_A t + \frac{q}{\varepsilon} N_D w_n \quad \text{for} \quad 0 \leq t \leq w_p.
\]

Integrating the electric field produces an expression for potential as a function of depth, \(t\). If we define zero potential as the potential of the undepleted base region, the potential between \(-w_n\) and 0 is given by

\[
V(t) = -\frac{q N_D (w_n + t)^2}{2 \varepsilon},
\]
while the potential between 0 and \( w_p \) is given by

\[
V(t) = -\frac{q}{\varepsilon} \left( N_D w_n t - \frac{1}{2} N_A t^2 + \frac{1}{2} N_D w_n^2 \right).
\]

The optimum emitter-base grading is obtained if we set

\[
\frac{\text{Al content at } t}{\text{Maximum Al content}} = \frac{V(t)}{V_{bi} - V_{app}},
\]

where "maximum Al content" refers to the Al content of the undepleted emitter (typically 30%).

The procedure just described yields satisfactory grading designs for typical AlGaAs/GaAs HBT levels of doping and Al content. For unusually high doping levels, the built-in voltage \( V_{bi} \) may differ significantly from the bandgap and should be explicitly calculated. For high Al content, or in systems other than AlGaAs/GaAs where band discontinuities are greater, \( V_{bi} \) for the calculation should be \( V_{bi}(\text{homojunction}) + \Delta E_c \) for N-p-n HBT's, or \( V_{bi}(\text{homojunction}) + \Delta E_v \) for P-n-p HBT's.

**Heterojunction grading during MBE growth**

When growing epitaxial layers of constant alloy composition by MBE, the composition is controlled by the relative fluxes of the constituents, which in turn are determined by the temperatures of the effusion cells. During growth of the emitter-base graded heterojunction layer, one must change the composition continuously from pure GaAs (or InGaAs) to the desired AlGaAs alloy. At typical compositions and growth rates, achieving such grading through
altering the fluxes requires precise temperature control during ramps of approximately 100° C per minute, which is beyond the capability of current technology.

Since the shutters on the effusion cells can be opened or closed in roughly 0.1 sec, emitter-base grading of ternary compounds such as AlGaAs or InGaAs is accomplished by opening and closing the Al or In source shutter with a varying duty cycle, forming a "chirped" superlattice. With computer control of the MBE source temperatures and shutters, optimal grading is easily accomplished by this method.

Once the desired grading profile is known (such as from the design procedure of the previous section), the graded layer growth is divided up into segments of equal thickness. Each segment is assigned an alloy composition according to the grading profile, and then shutter open and close times are calculated as follows: let $A =$ growth rate with shutter open, $B =$ growth rate with shutter closed, $a =$ time shutter is open, $b =$ time shutter is closed, and $T =$ thickness of a grading segment. If $x =$ desired alloy composition, then $a = \frac{Tx}{A}$ and $b = T \frac{1-x}{B}$. The shutter open and close time calculations, together with the design algorithm of the previous section, lend themselves well to computer solution.

The segment thickness $T$ should be chosen small enough to insure that the carrier of interest can easily tunnel through the barriers in the superlattice. For N-p-n HBT's, thicknesses of 10 to 20 Å are appropriate. For P-n-p HBT's, however, smaller grading
segments should be used, since the more massive holes tunnel much less readily. The HBT's in this dissertation used grading segments of approximately "3 Å", so that grading was actually continuous down to the atomic level.
Appendix B
Dual Selective Etch HBT Fabrication Process

All of the HBT's described in this dissertation were fabricated with a process sequence similar to that described in this appendix, with the following exceptions:

1) The first devices, described in chapter 4, used a non-selective phosphoric acid etch solution for all of the etch steps: formation of the emitter mesa, removal of the AlGaAs surface passivation layer away from the emitter edge for exposure and contacting of the base layer, and formation of the base mesa.

2) The devices described in chapter 5 used the non-selective phosphoric acid solution for exposure of the base layer.

In the following section, "phosphoric etch" refers to a solution consisting of 3 parts phosphoric acid (H₃PO₄), 1 part hydrogen peroxide (H₂O₂), and 50 parts deionized water (H₂O) by volume. At 20° C, this solution etches GaAs uniformly at 700-750 Å/min; the etch rate of Al₀.₃Ga₀.₇As is between 1000 and 1400 Å/min.

"PR" refers to AZ-1370 SF positive photoresist; "DI" refers to deionized water. Unless stated otherwise, oxygen plasma is 50 W at a pressure of 100 mTorr with 25 sccm oxygen. All photoresist exposures were done with 400 nm UV radiation.
First mask level: Emitter ohmic contact

1) Singe bake wafers at 120° C for 20 min to desorb H₂O
2) Spin PR at 3500 RPM for 30 sec
3) Pre-bake 85° C for 20 min
4) Soak in chlorobenzene for 10 min to harden the top of the PR; this effectively makes a two-level resist for undercutting during developing to improve metal lift-off.
5) Bake 85° C for 5 min
6) Expose 11 sec (for chlorobenzene process) at 12 mW/cm²
7) Develop for 2 min; DI rinse
8) Inspect under microscope
9) Postbake 120° C for 20 min
10) O₂ plasma for 15 sec to remove PR scum
11) 4:1 H₂O:HCl dip 30 sec, rinse (to remove native oxide formed by O₂ plasma)
12) Deposit Ti 500 Å/Pt 200 Å/Au 1000 Å
13) Liftoff metal in acetone; rinse in methanol, then DI
14) Inspect under microscope

Figure B.1 Lift-off of emitter contact metal
Emitter edge thinning etch (formation of AlGaAs surface passivation layer using the emitter contact metal as a mask)

The 250:1 $\text{H}_2\text{O}_2 : \text{NH}_4\text{OH}$ etch, while very selective to AlGaAs, is sensitive to the cleanliness of the surface and the thickness of the native oxide. The oxygen plasma is required to remove even the slightest traces of photoresist residue or other surface contamination, while the acid dip is needed to remove the native oxide (probably < 100 Å) formed by the oxygen plasma.

15) $\text{O}_2$ plasma for 1 min

16) 4:1 $\text{H}_2\text{O}:\text{H}_3\text{P}0_4$ dip 30 sec, short DI rinse

In the following step, the etching solution should be vigorously rotated for best uniformity; holding the wafer parallel to the flow while magnetically stirring at 150 RPM gives excellent results. Under these conditions, at $10^\circ$ C, an etch rate of 500 Å/min is obtained for GaAs, while the AlGaAs etch rate is very low. [While only one minute of etching is required to expose the AlGaAs layer for the wafers in this dissertation, the AlGaAs layer was slightly thicker than optimal for surface passivation. The additional two minutes of etching was found to remove about 200 Å of AlGaAs, leaving a thickness more appropriate for good surface passivation.]

17) Etch in 250:1 $\text{H}_2\text{O}_2 : \text{NH}_4\text{OH}$ (pH 7.25), $10^\circ$ C for 3 min; rinse in DI
The two minutes overetch of the previous step leaves a significant thickness of oxidized AlGaAs on the surface. In order to promote PR adhesion and prevent undercutting in later etch steps, this native oxide is removed by the following step.

18) 4:1 H₂O:H₃PO₄ dip for 45 sec, short DI rinse
19) Inspect

---

Figure B.2 Formation of AlGaAs surface passivation layer by selective etching

**Second mask level: Emitter mesa definition**

20) Singe bake 120° C for 20 min to desorb H₂O

   *PR adhesion promoter was used for the second and fourth mask levels because of some early problems with poor adhesion.*

21) Spin adhesion promoter, then PR at 5000 RPM for 30 sec
22) Bake 85° C for 20 min
23) Expose 7 sec at 12 mW/cm²

   *In the following step, all of the PR is removed from one section so that one section of the wafer consists of HBT's with no surface passivation.*

24) Second exposure using Al foil as crude mask to remove all PR on one die
25) Develop about 45 sec; DI rinse
26) Inspect under microscope
27) Postbake 120° C for 30 min
   Once again, etching is preceded by oxygen plasma and then an acid dip to remove the native oxide formed by the plasma; this insures a clean surface for etching.
28) O₂ plasma for 15 sec
29) 4:1 H₂O:HCl dip for 30 sec, DI rinse
   Because the etch rate of the K₄Fe(CN)₆/ K₃Fe(CN)₆ solution varied somewhat from mixing to mixing, the etch rate was roughly calibrated by etching a sample of bulk Al₀.₃Ga₀.₇As for 5 min (the relative etch rates of GaAs, AlGaAs, and InGaAs appeared to remain constant, however). From this rough calibration, the time required to etch 400 Å of AlGaAs was calculated, and then this etch time was used in the following step.
   The adjustment of pH requires some care. Although any pH between 6.5 and 7.0 is probably acceptable, the unadjusted pH is around 9. Addition of 1 ml of 1000:1 H₂O : HCl only slightly lowers the pH. However, as pH = 7.0 is approached, the pH begins to drop more rapidly with added HCl solution.
30) Etch wafers in 0.225 molal solution K₄Fe(CN)₆/ K₃Fe(CN)₆ of pH=7.0. (19.01 g K₄Fe(CN)₆·3H₂O + 14.81 g in 200 ml H₂O with about 2.5 ml 1000:1 H₂O:HCl) at 24° C, rinse (etch for time required to etch 400 Å of Al₀.₃Ga₀.₇As -- about 2 min)
The selectivity of the previous etch is based on formation of some product containing In which protects the surface from further etching; this product is removed by an acid dip in the following step.

31) 4:1 \( \text{H}_2\text{O} : \text{H}_3\text{PO}_4 \) dip for 30 sec, rinse

32) Strip PR in acetone; rinse in methanol, DI

33) Check step height

---

Figure B.3 Exposure of base (away from emitter edge) by selective etching

**Third mask level: base ohmic contact metal**

34) Singe bake 120° C for 20 min to desorb \( \text{H}_2\text{O} \)

35) Spin adhesion promotor, then PR at 3500 RPM for 30 sec

36) Bake 85° C for 20 min

37) Soak in chlorobenzene for 10 min

38) Bake 85° C for 5 min

39) Expose 11 sec (for chlorobenzene process) at 12 mW/cm²

40) Develop 1:1 AZ for 2 min; DI rinse

41) Inspect under microscope

42) Postbake 120° C for 30 min

43) \( \text{O}_2 \) plasma for 15 sec to remove PR scum

44) 4:1 \( \text{H}_2\text{O} : \text{H}_3\text{PO}_4 \) dip 30 sec to remove native oxide, short DI rinse
45) Deposit metal: 500 Å Ti, 200 Å Pt, 500 Å Au
46) Lift off in acetone; rinse in methanol, DI
47) Inspect under microscope

![Figure B.4 Lift-off of base contact metal](image)

**Fourth mask level: base mesa define**

48) Singe bake 120° C for 20 min to desorb H₂O
49) Spin adhesion promoter, then PR at 5000 RPM for 30 sec
50) Bake 85° C for 20 min
51) Expose 7 sec at 12 mW/cm²
52) Develop about 1 min; DI rinse
53) Inspect under microscope
55) Postbake 120° C for 30 min
56) O₂ plasma 100 W for 30 sec
57) 4:1 H₂O₂:HCl dip 30 sec, short DI rinse
58) Etch in phosphoric solution at 20° C for 1 min 30 sec
59) Strip PR in acetone; rinse in heated methanol, DI

![Figure B.5 Formation of base mesa by nonselective etching](image)
Fifth mask level: silicon nitride deposit and etch

60) Deposit 1500 Å nitride
   Adhesion promoter is particularly important for the last two mask levels, since PR does not adhere well to silicon nitride.

61) Spin adhesion promoter, then PR at 5000 RPM for 30 sec

62) Bake 85° C for 20 min

63) Expose 7 sec at 12 mW/cm²

64) Develop about 45 sec; DI rinse

65) Inspect under microscope

66) Postbake 120° C for 30 min

67) Etch silicon nitride with 100 sccm C₂F₆/10 sccm O₂ at 150 W, 100 mTorr, for 4 min

68) Inspect under microscope

69) Strip PR in acetone; rinse in methanol, DI
   After being exposed to plasma, PR is not completely stripped by acetone. Many commercial PR strippers etch GaAs, so oxygen plasma is used instead.

70) Strip remaining PR residue in O₂ plasma for 1 min at 150 W

71) Rinse wafer in acetone, then methanol, then DI

72) Inspect for remaining PR residue

---

Figure B.6 Deposition of silicon nitride by PECVD, then opening of contact holes by RIE
Sixth mask level: interconnect metal

73) Singe bake 120° C for 20 min to desorb H₂O
74) Spin adhesion promoter, then PR at 3500 RPM for 30 sec
75) Bake 85° C for 20 min
76) Soak in chlorobenzene for 10 min
77) Bake 85° C for 5 min
78) Expose 11 sec (for chlorobenzene process) at 12 mW/cm²
79) Develop for 2 min; DI rinse
80) Inspect under microscope
81) Postbake 120° C for 30 min
82) O₂ plasma 15 sec
83) Deposit metal: 300 Å Ti, 2700 Å Al
84) Deposit backside metal for collector contact: Ti 300Å, Al 1000Å
85) Liftoff in acetone; rinse in methanol, then DI
86) Inspect under microscope

Figure B.7 Patterning of interconnect metal by lift-off
References


10. D. G. Hill, K. L. Lear, and J. S. Harris, Jr., "Two selective etching solutions for GaAs on InGaAs and GaAs/AlGaAs on InGaAs," to be published in *Journal of the Electrochemical Society*.


12. S. Strite, M. S. Unlu, K. Adomi, G. B. Gao, and H. Morkoç, "AlGaAs/Ge/GaAs heterojunction bipolar transistors grown by


14 G. L. Patton, J. H. Comfort, B. S. Meyerson, E. F. Crabbe, G. J. Scilla, E. de Fresart, J. M. C. Stork, J. Y.-C. Sun, D. L. Harame, and J. N. Burghartz, "75-GHz $f_T$ SiGe-base heterojunction bipolar transistors," *IEEE Electron Device Letters*, vol. 11, pp. 171-173 (1990). It should be noted that the 75 GHz figure is somewhat misleading. High values of $f_T$ may be obtained by making conventional bipolar transistors with very thin base layers, which are not useful devices due to the extremely high base resistance. For this reason, both $f_T$ and $f_{max}$ are customarily quoted when discussing device speed, since $f_{max}$ includes the effect of base resistance. However, $f_{max}$ was very low for the devices in this paper (probably below 1 GHz), although no figure is given.


16 Grading of GaAs to InGaAs or InAs before application of contact metal reduces or eliminates this Schottky barrier and yields very good ohmic contacts; however, this is not a direct contact between GaAs and metal.


19 The actual flux varies somewhat from the measured value, since the flux is measured by an ionization gauge with different sensitivities to various elements. This effect, of course, is taken into account when calibration is performed.


35 For a p-n junction, $I \sim e^{kT/q}$, or $V \sim \frac{kT}{q} \ln(I)$. The effective resistance, $dV/dI$, is therefore $\frac{kT}{q}$. 


52. P. S. Peercy, B. W. Dodson, J. Y. Tsao, E. D. Jones, D. R. Myers, T. E. Zipperian, L. R. Dawson, R. M. Biefeld, J. F. Klem, and C. R. Hills, "Stability of strained quantum-well field-effect transistor structures," *IEEE Electron Device Letters*, vol. 9, pp. 621-623 (1988). Although the analysis is valid, it should be noted that the experimental verification is flawed in this paper: the use of SiO₂ as a cap layer during an 800° C anneal destroys the top few thousand Angstroms of GaAs due to absorption of Ga into the SiO₂.


61. X. S. Wu, L. A. Coldren, and J. L. Merz, "Selective etching characteristics of HF for Al\textsubscript{x}Ga\textsubscript{1-x}As/GaAs," *Electronics Letters*, vol. 21, pp. 558-559 (1985).


64. J. M. Ianelli, J. Maserjian, B. R. Hancock, P. O. Andersson and F. J. Grunthaner, "Optically controlled absorption modulator based on state filling of In\textsubscript{x}Ga\textsubscript{1-x}As/GaAs quantum wells," *Applied Physics Letters*, vol. 54, pp. 301-303 (1988).


70 Michael Kim, private communication.

71 Comparison of Figures 7.6 and 7.7 reveals that despite the improved ideality factor for base current, the current gain has actually decreased. This is probably due to outgassing of the arsenic cracker during growth of the wafer with improved heterojunction grading structure. This was one of the first wafers grown with a new arsenic cracker, before any detailed characterization of the cracker was performed.

72 M. Heiblum, private communication.