

University of California
Santa Barbara

Improved Current-Gain Cutoff Frequency
and High Gain-Bandwidth Amplifiers
in Transferred-Substrate HBT Technology

A Dissertation submitted in partial satisfaction
of the requirements for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering
by
Dino Mensa

Committee in charge:
Professor Mark Rodwell, Chairperson
Professor Arthur Gossard
Professor Herbert Kroemer
Professor Stephen Long

October, 1999

The dissertation of Dino Mensa
is approved:

Committee Chairperson

October, 1999

Copyright by

Dino Mensa

1999

Acknowledgments

My graduate advisor, Professor Mark Rodwell, is an extremely hard working and dedicated man. His motivation and enthusiasm have set him apart as a very personable and inspiring employer. The technical knowledge which he demonstrates (often on the spur of the moment) results in a clear, concise answer to whatever the question was. I am indebted to him for a lot of time and attention over the past four years.

The advisory committee (Professor Gossard, Professor Kroemer, and Professor Long) deserves thanks for their time, attention, and especially patience.

I would like to thank Professor Hua Lee for bailing me out of a few sticky situations but mostly for being a good friend.

My lab mates have made this a very interesting (and I mean that in a good way) experience and have given me much assistance with full knowledge that I may not have anything to offer them in return.

My parents have always encouraged me to try my best. I have done that here, and the rewards are many. Thank you dad and mom for all of your love and support, especially since I moved out to UCSB in 1991.

Thank you to my loving wife, Anjanette. You have encouraged me no matter what the circumstance, and it looks like you were right all those times you said I could do it. You are always seeing the best in every situation, and I look forward to our new life together.

Thank you, God, for everything.

Vita

- October 25, 1973: born in Ventura, California.
- June, 1995: B.S., Physics, University of California, Santa Barbara.
- June 18, 1995, Married to Anjanette Ruments.
- June, 1997: M.S., Electrical and Computer Engineering, University of California, Santa Barbara.
- 1995-1999: Research Assistant, Department of Electrical and Computer Engineering, University of California, Santa Barbara.

Journal Publications

1. D. Mensa, S. Krishnan, J. Guthrie, S. Jaganathan, T. Mathew, Y. Wei, and M.J.W. Rodwell, “18 dB, DC - 50 GHz Resistive Feedback Amplifiers”, to be submitted to *IEE Electronics Letters*.
2. D. Mensa, Q. Lee, J. Guthrie, S. Jaganathan, and M.J.W. Rodwell, “Transferred-Substrate HBTs with 254 GHz ft”, *Electronics Letters*, vol. 35, no. 7, pp. 605-606, 1999.
3. D. Mensa, R. Pallela, Q. Lee, J. Guthrie, S.C. Martin, R.P. Smith, S. Jaganathan, T. Mathew, B. Agarwal, S. Long, and M.J.W. Rodwell, “48 GHz Digital ICs and 85 GHz Baseband Amplifiers Using Transferred-Substrate HBTs”, *IEEE Journal of Solid-State Circuits*, vol. 34, no. 9, pp. 1196-1203, Sept. 1999.
4. M.J.W. Rodwell, D. Mensa, Q. Lee, J. Guthrie, Y. Betser, S.C. Martin, R.P. Smith, S. Jaganathan, T. Mathew, S. Krishnan, S. Long, R. Pallela, B. Agarwal, U. Bhattacharya, L. Samoska, “Submicron Scaling of Heterojunction Bipolar Transistors for THz Device Bandwidths”, invited submission for *IEEE Transactions on Electron Devices*, 1999.
5. Q. Lee, S.C. Martin, D. Mensa, R. P. Smith, J. Guthrie, and M. J. W. Rodwell, “Submicron Transferred-Substrate Heterojunction Bipolar Transistors”, *IEEE Electron Device Letters*, vol. 20, no. 8, Aug. 1999.
6. R. Pallela, D. Mensa, Q. Lee, B. Agarwal, J. Guthrie, S. Jaganathan and M.J.W. Rodwell, “48 GHz static frequency-dividers in transferred-substrate HBT technology”, *Electronics Letters*, vol. 34, pp. 1580-1581, 1998.
7. B. Agarwal, R. Pallela, U. Bhattacharya, D. Mensa, Q. Lee, L. Samoska, J. Guthrie and M. J. W. Rodwell, “Ultrahigh f_{max} AllnAs/GaInAs transferred-substrate heterojunction bipolar transistors for integrated circuit applications”, *International Journal of High Speed Electronics and Systems*, vol. 9, no. 2, pp. 643-670, 1998, (invited paper).

8. J. Guthrie, D. Mensa, B. Agarwal, Q. Lee, R. Pallela and M. J. W. Rodwell, "HBT IC process with a copper substrate", *Electronics Letters*, vol. 34, IEEE, pp. 467-8, 1998.
9. B. Agarwal, R. Pallela, Q. Lee, D. Mensa, J. Guthrie and M. J. W. Rodwell, "80 GHz distributed amplifiers with transferred-substrate heterojunction bipolar transistors", *IEEE Transactions on Microwave Theory and Techniques*, vol. 46, pp. 2302-2307, 1998.
10. B. Agarwal, Q. Lee, R. Pallela, D. Mensa, J. Guthrie and M. J. W. Rodwell, "Broadband feedback amplifiers with AlInAs/GaInAs transferred-substrate HBTs", *Electronics Letters*, vol. 34, no. 13, pp. 1357-1358, 1998.
11. B. Agarwal, Q. Lee, R. Pallela, D. Mensa, J. Guthrie and M. J. W. Rodwell, "A Transferred-Substrate HBT Wideband Differential Amplifier to 50 GHz", *IEEE Microwave and Guided Wave Letters*, vol. 8, no. 7, pp. 263-265, 1998.
12. Q. Lee, B. Agarwal, D. Mensa, R. Pallela, J. Guthrie, L. Samoska and M. J. W. Rodwell, "A > 400 GHz f_{max} transferred-substrate heterojunction bipolar transistor IC technology", *IEEE Electron Device Letters*, pp. 77-79, March 1998.
13. B. Agarwal, D. Mensa, R. Pallela, Q. Lee, U. Bhattacharya, L. Samoska, J. Guthrie and M. J. W. Rodwell, "A 277 GHz f_{max} transferred-substrate heterojunction bipolar transistor", *IEEE Electron Device Letters*, vol. 18, no. 4, pp. 228-231, 1997.
14. U. Bhattacharya, L. Samoska, R. Pallela, J. Guthrie, Q. Lee, B. Agarwal, D. Mensa and M. J. W. Rodwell, "170 GHz transferred-substrate heterojunction bipolar transistor", *Electronics Letters*, vol. 32, no. 15, pp. 1405-1406, 1996.

Conference Papers

1. D. Mensa, Y. Wei, J. Guthrie, S. Jaganathan, T. Mathew, S. Krishnan, and M.J.W. Rodwell, "A Transferred-Substrate HBT with 275 GHz Current-Gain Cutoff Frequency", to be submitted to *1999 International Electron Devices Meeting*.
2. D. Mensa, M.J.W. Rodwell, Q. Lee, J. Guthrie, S. Jaganathan, T. Mathew, S. Krishnan, "Bipolar Technologies and Optoelectronics", *IEEE Microwave Theory and Techniques Symposium Workshop: Technologies for the Next Millenium*, June 1999.
3. D. Mensa, Q. Lee, J. Guthrie, S. Jaganathan, and M.J.W. Rodwell, "Transferred-Substrate HBTs with 250 GHz current-gain cutoff frequency", *International Electron Devices Meeting Technical Digest*, pp. 657-660, 1998.
4. D. Mensa, Q. Lee, R. Pallela, B. Agarwal, J. Guthrie, S. Jaganathan, M. Rodwell, "Baseband Amplifiers in Transferred-Substrate HBT Technology", *GaAs IC Symp. Tech. Dig.*, pp. 33-36, 1998.
5. D. Mensa, R. Pallela, B. Agarwal, Q. Lee, J. Guthrie and M. J. W. Rodwell, "Over 220 GHz f_{τ} and f_{max} HBTs in a transferred-substrate IC process", *Workshop on Compound Semiconductors, Materials and Devices*, 1998.
6. Q. Lee, S. C. Martin, D. Mensa, R. P. Smith, J. Guthrie, S. Jaganathan, Y. Betser, T. Mathew, S. Krishnan, L. Samoska, and M. Rodwell, "Submicron Transferred-Substrate Heterojunction Bipolar Transistors with Greater than 1 THz f_{max} ", *Device Research Conference Dig.*, June 1999.
7. Q. Lee, D. Mensa, J. Guthrie, S. Jaganathan, T. Mathew, Y. Betser, S. Krishnan, S. Ceran, and M.J.W. Rodwell, "66 GHz Static Frequency Divider in Transferred-Substrate HBT Technology", *IEEE Radio Frequency Integrated Circuits Symposium Tech. Digest*, pp. 87-90, June 1999.

8. M.J.W. Rodwell, Q. Lee, D. Mensa, J. Guthrie, Y. Betser, S.C. Martin, R.P. Smith, S. Jaganathan, T. Mathew, S. Krishnan, S. Ceran, and S. Long, "Ultra High Frequency Integrated Circuits using Transferred-Substrate Heterojunction Bipolar Transistors", *IEEE International Symposium on Circuits and Systems*, June 1999.
9. M. Rodwell, Q. Lee, D. Mensa, J. Guthrie, Y. Betser, S.C. Martin, R.P. Smith, S. Jaganathan, T. Mathew, S. Krishnan, S. Ceran, and S. Long, "Transferred-Substrate Heterojunction Bipolar Transistor Integrated Circuit Technology", *Eleventh International Conference on Indium Phosphide and Related Materials*, May 1999.
10. Q. Lee, S.C. Martin, D. Mensa, R.P. Smith, J. Guthrie, S. Jaganathan, T. Mathew, S. Krishnan, S. Ceran, and S. Long, "Submicron Transferred-Substrate Heterojunction Bipolar Transistors with Greater than 800 GHz fmax", *Eleventh International Conference on Indium Phosphide and Related Materials*, May 1999.
11. J.R. Guthrie, D. Mensa, T. Mathew, Q. Lee, S. Krishnan, S. Jaganathan, S. Ceran, Y. Betser, and M.J.W. Rodwell, "A 50mm Copper/Polymer Substrate HBT IC Technology for > 100 GHz MMICs", *Eleventh International Conference on Indium Phosphide and Related Materials*, May 1999.
12. Q. Lee, D. Mensa, J. Guthrie, M.J.W. Rodwell, S. Long, S.C. Martin, R.P. Smith, Y. Betser, S. Jaganathan, T. Mathew, S. Krishnan, and C. Serhan, "Ultra High Frequency HBT Integrated Circuits", *1999 Ultrafast Electronics and Optoelectronics Technical Digest*, Apr. 1999.
13. O. Wohlgemuth, B. Agarwal, R. Pallela, D. Mensa, Q. Lee, J. Guthrie, M.J.W. Rodwell, R. Reuter, J. Braunstein, M. Schiechtweg, T. Krems, K. Kohler, "A NLTL-based integrated circuit for a 70-200 GHz VNA system", *28th European Microwave Conference Proceedings*, vol. 1, pp. 104-107, 1998.
14. R. Pallela, D. Mensa, B. Agarwal, J. Guthrie, M. Rodwell, "47 GHz static frequency divider in ultrafast transferred-substrate heterojunction bipolar transistor technology", *1998 International Conference on Indium Phosphide and Related Materials Dig.*, pp. 68-71, 1998.

15. Q. Lee, S. C. Martin, D. Mensa, R. Pullela, R. P. Smith, B. Agarwal, J. Guthrie and M. Rodwell, "Deep sub-micron transferred-substrate heterojunction bipolar transistors", *Device Research Conference Dig.*, pp. 26-27, 1998.
16. M. Rodwell, Q. Lee, D. Mensa, R. Pullela, J. Guthrie, S.C. Martin, R.P. Smith, S. Jaganathan, T. Mathew, B. Agarwal, S. Long, "48 GHz digital ICs using transferred-substrate HBTs", *GaAs IC Symp. Tech. Dig.*, pp. 113-116, 1998.
17. L. Samoska, R. Pullela, B. Agarwal, D. Mensa, Q. Lee, V. Kaman, J. Guthrie, and M. J. W. Rodwell, "InP heterojunction bipolar transistor decision circuits", *1998 MTT-S International Microwave Symposium dig.*, vol. 3, pp. 1843-1846, 1998.
18. R. Pullela, B. Agarwal, Q. Lee, D. Mensa, J. Guthrie, L. Samoska and M. J. W. Rodwell, "Ultrafast transferred-substrate heterojunction bipolar transistor ICs for high-speed fiber-optic transmission", *Optical Fiber Conference Dig.*, 1998, (invited paper).
19. R. Pullela, Q. Lee, B. Agarwal, D. Mensa, J. Guthrie, L. Samoska, and M. Rodwell, "A > 400 GHz f_{max} transferred-substrate HBT integrated circuit technology", *Device Research Conference Technical Digest*, pp. IIB-2, 1997.
20. M. J. W. Rodwell, R. Pullela, B. Agarwal, M. Reddy, Q. Lee, J. Guthrie, D. Mensa, L. Samoska, S. C. Martin and R. P. Smith , "Submicron lateral scaling of vertical-transport devices: transferred-substrate bipolar transistors and schottky-collector tunnel diodes", *Ultrafast Electronics and Optoelectronics Technical Digest*, pp. 192-194, 1997.
21. B. Agarwal, D. Mensa, R. Pullela, Q. Lee, U. Bhattacharya, L. Samoska, J. Guthrie and M. J. W. Rodwell, "A 277 GHz f_{max} transferred-substrate heterojunction bipolar transistor", *International Conference on Indium Phosphide and Related Materials Technical Digest*, pp. 633-636, 1997.
22. U. Bhattacharya, L. Samoska, R. Pullela, J. Guthrie, Q. Lee, B. Agarwal, D. Mensa, M.J. Mondry, G. Hurtz, and M. J. W. Rodwell, "100

GHz Transferred-Substrate Schottky-Collector Heterojunction Bipolar Transistor”, *International Conference on Indium Phosphide and Related Materials Technical Digest*, pp. 145-148, 1996.

Abstract

Improved Current - Gain Cutoff Frequency and High Gain - Bandwidth Amplifiers in Transferred-Substrate HBT Technology

by
Dino Mensa

The transferred-substrate HBT technology developed at UCSB has demonstrated the ability to realize extremely high bandwidth bipolar transistors. Record values of f_{max} over 1 THz have been attained, promising levels of performance in distributed and tuned amplifier circuitry far beyond the current state of the art. In order to fabricate fast digital circuitry as well as analog circuitry of more conventional topologies, it is absolutely necessary to have simultaneously high f_τ as well as f_{max} . It is of paramount importance to improve the f_τ figure of merit in transferred-substrate HBTs via improved MBE growth of the HBT semiconductor, bandgap engineering techniques, and lateral as well as vertical scaling of the transistor structure. An f_τ of 275 GHz has been achieved in transferred - substrate HBT technology; this is the highest f_τ achieved to date for a bipolar transistor.

The high f_{max} values achieved at UCSB so far are the result of aggressive lithographic scaling of the collector dimension and space - charge screening effects in the collector (capacitance cancellation). Achieving high f_τ is more difficult. To minimize the various components of the HBT forward delay (proportional to the inverse of f_τ) quickly requires a combination of lithographic scaling, reduction of contact and access resistances, and reduction of parasitic layout capacitances.

Development of the basic HBT technology has continued in this work. A gas carbon source is now operational in the MBE system, and is able to dope the bases of future HBTs well into the $10^{20} / \text{cm}^3$ range without worry of dopant diffusion. A phosphorous source also has been installed in the MBE system and will soon be operational, giving transferred-substrate HBTs improvement in breakdown voltage and hence output power performance.

This work also demonstrates the highest gain-bandwidth single-stage amplifiers, implemented in the common-collector, common-emitter configu-

ration. These wide bandwidth amplifiers have very low power consumption of tens of milliwatts, and are further evidence of the frequency performance of transferred-substrate HBTs.

Contents

1	Introduction	1
2	MBE Growth	7
2.1	Base Layer Growth	9
2.2	Regrowth	14
2.3	Carbon Doping	17
2.4	Collector Layer Growth	22
2.5	Emitter Layer Growth	24
2.6	Summary	27
3	Processing	29
3.1	Process Flow	29
3.2	Discussion	33
3.3	Heat Sinking and Microstrip Interconnects	37
4	Device Physics	41
4.1	Device Modeling	41
4.2	Base Resistance	44
4.3	Base Transit Time	46
4.4	Current Gain	50
4.5	Collector Transit Time	52
4.6	Emitter Resistance	58
4.7	Kirk Effect	64
5	Device Results	71
5.1	Layout Parasitics	73
5.2	DC Measurements	76
5.3	Emitter Resistance	79

5.4	Discussion of Delay Components	82
5.5	Pulse Doping in the Collector	85
5.6	Collector-Base Capacitance	86
6	Resistive Feedback Amplifiers	89
6.1	First Generation Amplifier Design	92
6.2	First Generation Amplifier Results	96
6.3	Second Generation Amplifier Design and Results	97
6.4	Discussion	102
7	Fiber-Optic Receiver Design	105
7.1	Photodetector	108
7.2	Transimpedance Amplifier	111
7.3	Variable Gain Amplifier	112
8	Conclusion	119
8.1	Achievements	119
8.2	Future Work	121
A	Process Flow	125

List of Figures

1.1	Cross-sections, a) Double Mesa HBT, b) Transferred Substrate HBT	4
1.2	Transferred-Substrate HBTs are scalable	5
1.3	Transferred-Substrate HBTs are scalable	6
2.1	Baseline layer structure for transferred-substrate HBTs.	8
2.2	Band diagram for baseline structure with $V_{ce}=1.0$ V and $I_c=0$ mA.	9
2.3	Layer structure for beryllium-doped SIMS sample	12
2.4	As ₂ flux and temperature profile for beryllium-doped SIMS sample	12
2.5	SIMS results for beryllium pulses in InGaAs	13
2.6	Simplified process flow for regrown base. After emitter mesa definition (a), definition of dielectric mask (b), regrowth (c), then etchback of dielectric mask for emitter contact (d).	16
2.7	Schematic of carbon source gas manifold	19
2.8	Hole mobility as a function of P-type doping concentration.	19
2.9	Layer structure for carbon-doped SIMS sample.	21
2.10	SIMS results showing scans for carbon and silicon.	22
2.11	Collector Layer Structures, a) Baseline Schottky collector, b) Ohmic collector for CML	24
2.12	Layer structure of the baseline base-emitter grade. Alternating layers of InGaAs and InAlAs are used to approximate a linear grade from the InAlAs emitter to InGaAs. The superlattice period is 33.3 Angstroms, and there are nine periods.	26

3.1	Process flow diagram after deposition of emitter metal and base contact etch (a), after deposition of base metal (b), after isolation etch (c), after polyimide etchback (d).	31
3.2	Process flow diagram after deposition of metal 1 (a), after etching of SiN (b), after deposition of metal 2 (c), after etching of BCB (d).	32
3.3	Process flow diagram after gold plating and solder bonding to GaAs carrier wafer (a), after InP substrate removal and deposition of collector metal (M3) of SiN (b).	33
3.4	Passive elements in transferred-substrate process: NiCr resistor with heat sink (a), MIM capacitor (b), microstrip interconnect (c).	34
3.5	Schematic of base mesa and emitter finger for misaligned emitter finger.	36
3.6	Cross-section of emitter-base junction	37
4.1	Hybrid- π model of bipolar transistor, expressions for f_τ and f_{max}	42
4.2	Fit of $1/\text{Re}(Y_{21})$ vs. inverse emitter current to determine R_{ex} and V_t/n	43
4.3	Components of base resistance.	45
4.4	Base layer band diagrams for (a) linear compositional grading and (b) exponential doping grading	49
4.5	Base transit time as a function of compositional base grading strength.	51
4.6	Inverse current gain as a function of periphery to area ratio.	52
4.7	Collector field profile (a) with no injected current and (b) with a sheet charge σ of injected electrons	56
4.8	Velocity field curves for (a) III-V semiconductor and (b) silicon.	57
4.9	$1/R_{ex}$ as a function of emitter width W_e to determine effective undercut.	62
4.10	Emitter cap layers for contact resistance study	63
4.11	Increased current changes electric-field profile in collector.	65
4.12	Band diagram of collector-base junction at equilibrium and with Kirk Effect threshold current density flowing.	66
4.13	Collector - base band diagram with current density above Kirk threshold in a III-V HBT.	67

4.14	Postulated increase in electron velocity at high currents. . .	69
4.15	Pulse dope in collector used to delay onset of Kirk Effect . .	70
5.1	Baseline layer structure for transferred-substrate HBTs . . .	72
5.2	Detail of layouts showing base mesa and interconnects	75
5.3	Common emitter characteristics for highest f_τ device on each sample. Lithographically defined emitter and collector di- mensions and expected severity of undercutting etch are in- dicated.	78
5.4	RF data for $V_{ce} = 1.0$ V and peak $f_\tau I_c$	80
5.5	RF data for $V_{ce} = 1.0$ V and peak $f_\tau I_c$	81
5.6	Normalized C_{cb} vs. J_e at $V_{ce} = 1.0$ V.	83
5.7	Normalized C_{cb} vs. J_e at labeled V_{ce} for samples B, D, F. . .	88
6.1	Resistive feedback amplifier configurations	91
6.2	Circuit diagram of Darlington Amplifier. Line section char- acteristic impedance and electrical length as indicated. . . .	93
6.3	Chip photo of Darlington Amplifier.	94
6.4	Circuit diagram of Mirror Doubler.	95
6.5	Chip photo of Mirror Doubler Amplifier.	96
6.6	Measured S parameters of the Darlington Amplifier.	97
6.7	Measured S parameters of the Mirror Doubler Amplifier. . .	98
6.8	Circuit diagram of high - gain, second generation amplifier. .	99
6.9	Circuit diagram of wide - bandwidth, second generation am- plifier.	99
6.10	Chip photo of high-gain Darlington amplifier.	100
6.11	Measured S parameters of the high-gain Darlington Amplifier.	101
6.12	Measured S parameters of the high-bandwidth Darlington Amplifier.	101
7.1	Fiber-optic receiver block diagram showing transimpedance amplifier (TIA), variable gain amplifier (VGA), automatic gain control (AGC), decision circuit (DEC), clock and data recovery circuit (CDR), and demultiplexer (DMUX).	106
7.2	Photodiode: (a) Top View (b) Cross section	109
7.3	Circuit Diagram of the Transimpedance Amplifier	111
7.4	Simulated results of transimpedance amplifier	113
7.5	Chip photograph of transimpedance amplifier	113

7.6	Gain control stage with in, in, out, out, gain control voltage (Vgc) and reference voltage (Vref) labeled.	115
7.7	Simulated results of variable gain amplifier	116
7.8	Chip photograph of variable gain amplifier	117

Chapter 1

Introduction

This work describes growth, design, and fabrication of high-speed InP based heterojunction bipolar transistors (HBTs) and wide bandwidth HBT amplifiers.

The motivation to improve the frequency performance of semiconductor devices is the requirement for high-speed circuitry. Two key applications for high-speed circuitry will be examined: commercial demand for increased capacity fiber-optic communication networks and military programs which call for high-speed, high-resolution A/D conversion.

The use of fiber optics is increasingly desirable in modern communication networks. Local area networks (LANs) based on fiber optics provide far higher bit rate transmission than traditional twisted-pair or coaxial distribution systems. Optical fiber can simultaneously provide communication with low dispersion and loss. The intricacies of fiber-optic design in order to achieve these objectives are beyond the scope of this thesis, here it is assumed that such fiber exists and the focus is on the transceiver.

Fiber optic transceivers generally can be classified as either time division multiplexing (TDM) or wavelength division multiplexing (WDM) systems. These two categories refer to two different approaches to sending multiple channels of data down one fiber. In TDM, n channels of data are sent at one wavelength down one fiber. The receiver takes all of the data from all of the channels from that one fiber, and must sort the data according to which channel it belongs. It is evident that if 16 channels are to be sent through one fiber, the transmitter and the front end of the receiver must operate at least 16 times the data rate of each channel. Here there is a

direct tradeoff between the carrying capacity per channel and the number of channels, as the limiting component in all this is the speed of the front end receiver electronics.

In WDM systems, data are sent through a fiber at multiple wavelengths. It is not necessary that there be a one to one correspondence between wavelength and channel, a plurality of channels can be carried at one wavelength. In any event, some scheme is used to divide the signal received from the fiber among a number of detectors, each of which is tuned to receive a particular wavelength (channel). Here the distinction between WDM and TDM systems becomes indistinct, as each wavelength may be carrying multiple channels that in turn must be decoded using TDM subsystems. An increase in the TDM data rate for each wavelength means fewer wavelengths are needed for the same overall capacity, reducing crosstalk as well as providing more tolerance to laser diode emission line drifts with changing temperature and lifetime. The key point here is that both TDM and WDM systems stand to benefit from higher speed electronics.

The dominant factor limiting TDM system performance is most often the bandwidth of the front-end electronics. TDM transceivers offering data rates of 10 Gb/sec are now commercially available. There are 40 Gb/sec systems under development in many companies today that should be commercially available within the next few years [1] [2]. The Nyquist criterion is that the data must be sampled at least twice the signal bandwidth in order to accurately recover the signal. In order to satisfy the Nyquist criterion for accurate recovery of the data, the amplifiers must have very flat gain-frequency characteristics at a frequency equal to at least one - half of the data rate. This means that the 3dB bandwidth of the front - end amplifiers must be well above one-half of the data rate. The preamplifier gain should also be high in order to minimize the overall noise figure. Add to this the restrictions on the circuit topology of the front end to reduce thermal noise from biasing resistors and maintain gain flatness, and it becomes evident that the device bandwidth for the currently demonstrated 40 Gb/sec systems (not to mention the proposed 100+ Gb/sec systems) must be extremely high.

Much of the interest in the work of this group is for military applications. High-resolution A/D and D/A converters are required for microwave digital radio and phased array beamsteering applications. Future generations of these applications place enormous demands on the bandwidth, resolution,

and dynamic range of these converters [3]. High-resolution A/D converters of standard circuit topology (e.g. flash ADC) require extremely precise passive component parameters in order to achieve the desired resolution, a precision that has to date not been achievable. Instead Σ - Δ A/D converters can be made with high resolution by the technique of oversampling. Oversampling involves clocking the A/D converter at many times the signal bandwidth in order to achieve a higher "effective" resolution. The combination of high bandwidth and high resolution required will in turn require devices with far higher bandwidth than reported even in this work.

Now that the need for high-speed devices has been motivated, one question remains. Which devices should be used? The frequency performance of HEMTs has long been proven to be a scalable parameter, that is, as the gatelength of the HEMT is reduced, the frequency performance is enhanced. The limits to HEMT scaling are gate-channel separation, source/drain ohmic resistance, and breakdown voltage. A similar scaling phenomenon has only recently been shown with HBTs through the development of the transferred-substrate process at UCSB. The HBT has the benefits of high transconductance and uniformity in the relevant turn-on voltage (V_{be}) when compared with the HEMT. As always, the device to choose depends upon the application.

The transferred-substrate process makes the HBT a scalable device [4] [5]. What is meant here by a scalable HBT is that there is improvement in frequency response as the emitter stripewidth is reduced. So far, the maximum frequency of oscillation, f_{max} , has been shown to be scalable. F_{max} is the maximum frequency of oscillation of the device, power gain at frequencies beyond f_{max} is not possible. Another useful figure of merit is the short - circuit current gain cutoff frequency, f_{τ} . Beyond f_{τ} the device will have current gain less than unity. While it is possible to realize power gain beyond f_{τ} , circuit designs specifically designed to do this must be implemented which are not always practical. Furthermore, digital circuits are strongly limited by f_{τ} . Having an f_{max} that is much greater than an f_{τ} is not very useful in digital circuits. The limit to f_{τ} scaling in the transferred-substrate HBT process is the resistance in the Ohmic contact to the emitter, R_{ex} , which has not been addressed by the transferred-substrate HBT process differently than in any other HBT technology.

A diagram of a mesa HBT process is shown in fig. 1.1. The thickness and dielectric constant of the collector semiconductor and the area of the

base mesa determine the collector-base capacitance (C_{cb}). As the emitter stripewidth of the mesa HBT is reduced, the base spreading resistance decreases. The dominant component of base resistance for highly scaled HBTs is the base contact resistance, not the sheet or spreading resistance. This reduction in base spreading resistance is therefore a second order effect. The size of the base mesa can be reduced as the emitter stripewidth is reduced in order to reduce C_{cb} . In order not to adversely affect the base resistance (R_{bb}), the base mesa should extend at least two transfer lengths from the edge of the emitter stripe on all sides. This lower limit on the size of the base mesa precludes dramatic improvements in bandwidths of scaled mesa HBTs.

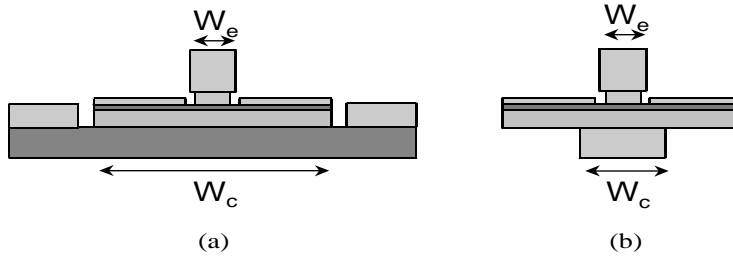


Figure 1.1: Cross-sections, a) Double Mesa HBT, b) Transferred Substrate HBT

The transferred-substrate process allows lithographic definition of both emitter and collector stripes. A schematic is shown in fig. 1.1 (b). The relevant area for C_{cb} is no longer the area of the base mesa, but the area of the collector stripe. The collector stripe need not be much wider than the emitter stripe, hence a large increase in f_{max} as the emitter (and collector) stripewidth is reduced. At the same time, the base mesa can be as wide as is necessary to accommodate the transfer length of the base ohmic. Fig. 1.2 shows the dramatic improvements in f_{max} as emitter stripewidth is scaled for the transferred-substrate HBT.

This motivates the next question: why InP based HBTs? The decision to use InAlAs/InGaAs HBTs over AlGaAs/GaAs HBTs was made for two reasons. The first was the much higher Γ -L separation in InGaAs as opposed to GaAs resulting in improved electron velocity in the high field collector.

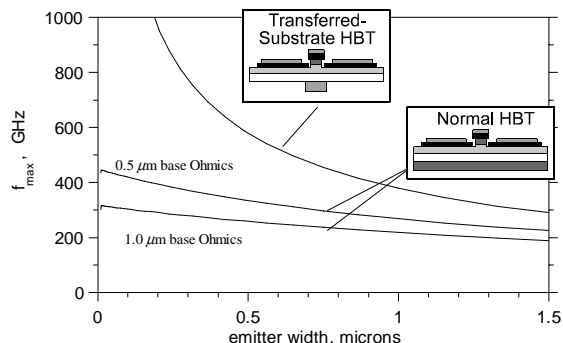


Figure 1.2: Transferred-Substrate HBTs are scalable

But more importantly, the decision was made not to use GaAs/AlGaAs because of the high surface recombination velocity in that material system. Transferred-substrate HBTs were meant to demonstrate scalability of HBTs and it has always been understood that their advantage would be evident only for emitters of small geometry. Without the addition of ledges (an AlGaAs cap on the exposed base surface), the current gain of GaAs/AlGaAs HBTs is severely degraded in narrow-emitter devices. The additional process complications introduced in ledge technology (employing an AlGaAs layer which inhibits electron motion toward the exposed surface) would be an unnecessary distraction in what was to be a demonstration of scalability, and so the InP based material system was chosen.

All of this improvement in f_{max} has had little or no impact on the current-gain cutoff frequency, f_{τ} . In fact, the highly scaled HBTs yielding the record values of f_{max} invariably have substantially lower f_{τ} than the devices of more conservative geometry. The MBE growth and fabrication issues pertaining to the performance of HBTs are discussed herein. The focus tends to be on f_{τ} , however other relevant HBT parameters such as current gain, breakdown voltage, etc. are discussed as they affect the suitability of these devices for IC applications. There was also fabrication and characterization of a carbon doping source, installed in the UCSB MBE facility. Amplifiers demonstrating record gain-bandwidth products have been fabricated; these gain blocks reflect the high bandwidth of the component devices and point to the high bandwidths possible for larger ICs fabricated in this process.

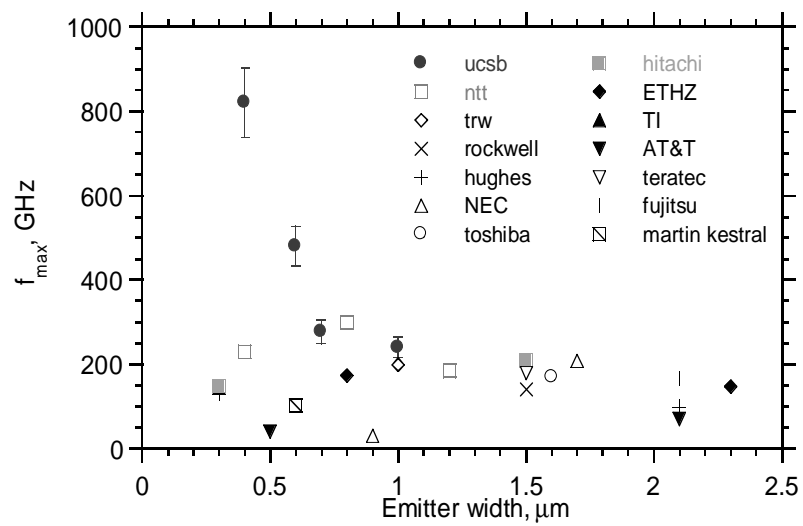


Figure 1.3: Transferred-Substrate HBTs are scalable

Chapter 2

MBE Growth

All material growth for this work was performed on a Varian Gen II MBE system here at UCSB. This particular MBE system is mainly used for growth of optoelectronic devices, such as surface emitting lasers, vertical-cavity lasers, detectors, etc. It has been beneficial to work on a system where the majority of the users were very concerned with the optical quality of the material.

All HBT growths were done on 2-inch Fe doped semi-insulating InP substrates. These substrates were purchased "epi-ready" from Sumitomo and the native oxide present on the wafer before growth seemed to be very thin. RHEED streaks are visible on these wafers before oxide desorption has begun. To desorb the oxide, the wafer is heated to above 530°C as indicated by the pyrometer. After about 30 seconds at this temperature, the typical twofold reconstruction is visible in the $\bar{1}\bar{1}0$ plane. As this twofold reconstruction begins to appear, the wafer temperature as indicated by the pyrometer begins to drop. After another 30 seconds or so, the apparent temperature may drop as much as 10-15 degrees and the twofold reconstruction becomes very distinct. It is presumed that the real temperature is not dropping, but that the emissivity is changing as the oxide is removed. If the temperature of the wafer is increased so that the pyrometer again reads above 530°C and the temperature is maintained at that level, a complete 2×4 reconstruction begins to appear. It has been observed that if the full 2×4 reconstruction appears, the stoichiometry of the surface near the edges of the wafer may have been ruined due to higher temperature present the the edge of the wafer where it is in contact with the block.

A layer structure typical of those used for this work is shown in fig. 2.1. This structure exactly as shown in the figure is referred to as "baseline" material. Typical growth temperatures are also indicated, and a growth rate of $\sim 1\mu\text{m}/\text{hour}$ was used. A band diagram is shown for this layer structure in fig. 2.2, under an applied bias of $V_{ce} = 1.0\text{ V}$, $V_{be} = 0.75\text{ V}$, and with no current flowing.

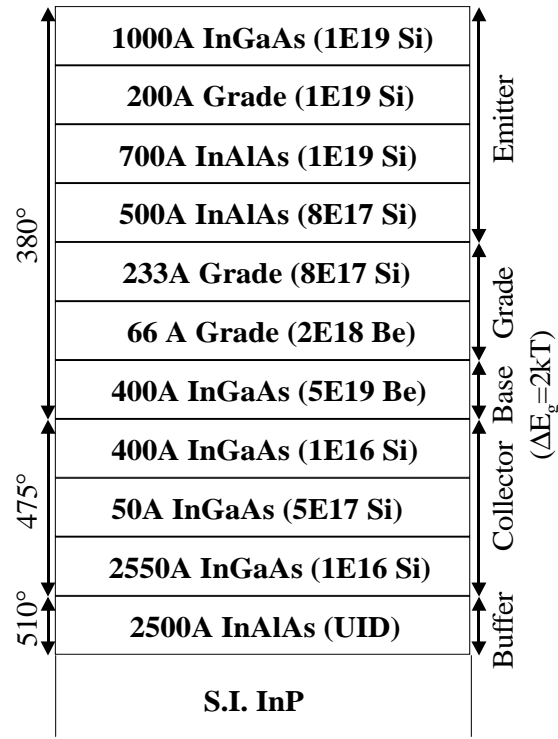


Figure 2.1: Baseline layer structure for transferred-substrate HBTs.

To ensure the epitaxial layers were lattice matched to the InP substrate, RHEED oscillations were measured each growth day for InAs on InAs, GaAs on GaAs, and AlAs on GaAs. The high-capacity gallium and aluminum cells yield a beam flux vs. temperature curve that changes very slowly over time until the cell nearly becomes depleted. The somewhat lower-

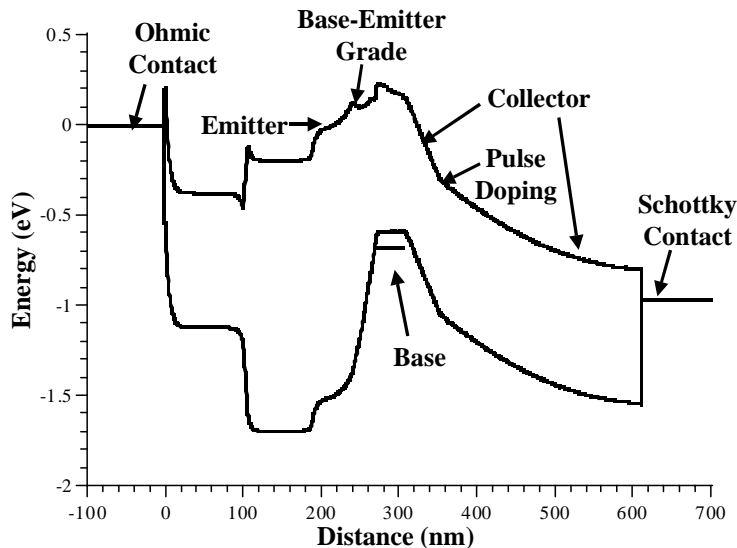


Figure 2.2: Band diagram for baseline structure with $V_{ce}=1.0$ V and $I_c=0$ mA.

capacity indium cell, however, may have its setpoint change substantially after only one full day of growth. A 5% drop in indium beam flux for a given temperature has been recorded after a 12 hour long growth day. This is an important point as it would imply that structures grown later in the day are very likely to be increasingly deficient in indium.

The remainder of this chapter discusses MBE growth issues pertaining to the individual layers of the HBT. The carbon-doping source is discussed, as well as future work on MBE growth of InP layers.

2.1 Base Layer Growth

In 1981, a report from Bell Labs showed that InGaAs could be doped into the low 10^{19} range with beryllium using solid-source MBE with nearly complete activation under certain growth conditions [7]. Beryllium was then widely adopted as the p-type dopant in MBE systems dedicated to growth of III-V compounds. Not long after, in the literature there appeared many papers related to beryllium diffusion mechanisms and growth conditions

that minimize beryllium diffusion. It has been experimentally determined that the diffusion coefficient of the beryllium acceptors (D_{be}) depends on the concentration of beryllium as well as the temperature [8]. Dopants may also diffuse due to electric fields present during the growth process. Beryllium diffusion is discussed herein, and growth conditions reviewed that would tend to favor a high activation efficiency of beryllium.

The layer structure of an n-p-n HBT is such that it is at the mercy of the beryllium diffusion. The p-type base must be very thin and very heavily doped. Beryllium diffusion into the lightly doped collector can render large regions of the collector p-type, affecting the Kirk effect threshold as well as the overall transit time. Beryllium diffusion into the emitter affects the turn-on voltage V_{be} , the DC current gain, the ideality factor, and the base transit time. The degree of beryllium diffusion depends upon the growth temperature, the beryllium concentration, and the III-V flux ratio employed during growth. If the degree of diffusion is reproducible, then a layer structure can be engineered to minimize the effects of the known diffusion profile, but this would require precise control of the substrate temperature and the beryllium doping concentration during growth. Variations in the density of defects and/or dislocations promoting dopant diffusion would complicate matters. Irreproducibility in the beryllium diffusion profile means irreproducibility in the things it affects such as the $V_{be,on}$ and the ideality factor. But reproducible V_{be} and high transconductance are two reasons why many applications require HBTs in the first place!

This beryllium diffusion problem can be handled in one of three ways: take steps to minimize the diffusion itself, minimize the effects of beryllium diffusion, or avoid beryllium altogether. Minimizing the amount of beryllium diffusion will be discussed here. Beryllium was the sole p-type dopant available for most of this work; the advice of previous growers as well as the abundance of information in the literature was therefore greatly appreciated.

The most common remedy for beryllium diffusion is simply to grow the beryllium containing layers at a lower temperature. The beryllium is less mobile at lower temperatures, therefore diffuses less. Enhanced activation at high doping concentrations may also occur with low growth temperatures due to inhibited formation of electrically inactive beryllium precipitates [9]. The lower surface mobility of adatoms at a lower growth temperature makes it more difficult for the beryllium atoms to "find" each other on the

growing surface. The growth temperature used in this work for growth of the base was estimated to be 380°C in the following manner. The desired growth temperature is well below the 450°C low temperature limit of the infrared pyrometer used to measure substrate temperature during growth. Upon commencing growth of the base layer, the substrate was brought to a temperature of 455°C as measured by the pyrometer, then the thermocouple setpoint was dropped by another 75°C to achieve the estimated substrate temperature of 380°C.

An increased As₂ flux during growth may also allow a higher maximum doping concentration. The idea is that the concentration of group III vacancies will increase at the expense of the As vacancies [8]. The Group III site is where the beryllium should incorporate in order to be electrically active, and so it seems natural that Group III vacancies should be preferred over Group V vacancies. For this reason, the V/III flux ratio during MBE growth of the base layer has always been in the range of 20-50. This number is limited by the Group III flux which yields ~1µm/hour growth rate, the temperature of the As₂ sublimator at ~365°C, and having the As₂ valve wide open.

The applicability of these theories regarding beryllium diffusion was tested using a calibration sample upon which Secondary Ion Mass Spectroscopy (SIMS) was performed. SIMS is a technique for vertical profiling of semiconductor layers. All SIMS work was performed by Charles Evans and Associates.

To examine beryllium diffusion under various growth conditions, a test sample was grown with the layer structure shown in fig. 2.3. Note that beryllium-doped layers were bounded by silicon doped "marker" layers. The variations of As₂ flux and growth temperature are shown in fig. 2.4.

The SIMS results are plotted in fig. 2.5, and at first seem to be just as expected. Deeper into the sample, the pulses become less abrupt, also the peak beryllium concentration is seen to have decreased. It is possible, however, that this would be observed even if this beryllium diffusion was not caused by the varying growth conditions, for two reasons. First, the incident ion beam causes intermixing of atoms as it strikes the sample, limiting the resolution of SIMS. This intermixing becomes worse as the sample is probed deeper. As a result, the peaks would naturally be expected to take on a more rounded shape deeper into the sample; simply an artifact of the SIMS technique. Another artifact of SIMS is rounding of the SIMS

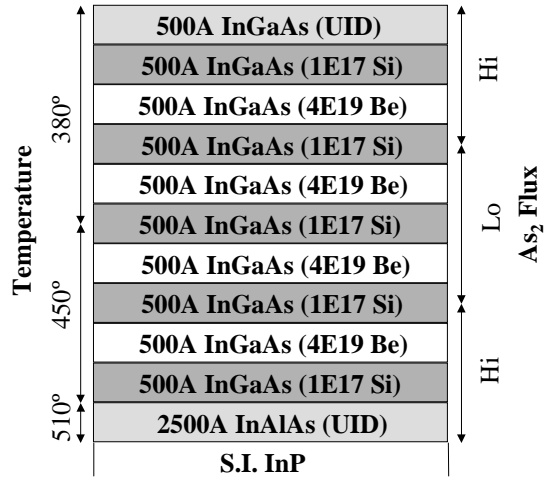
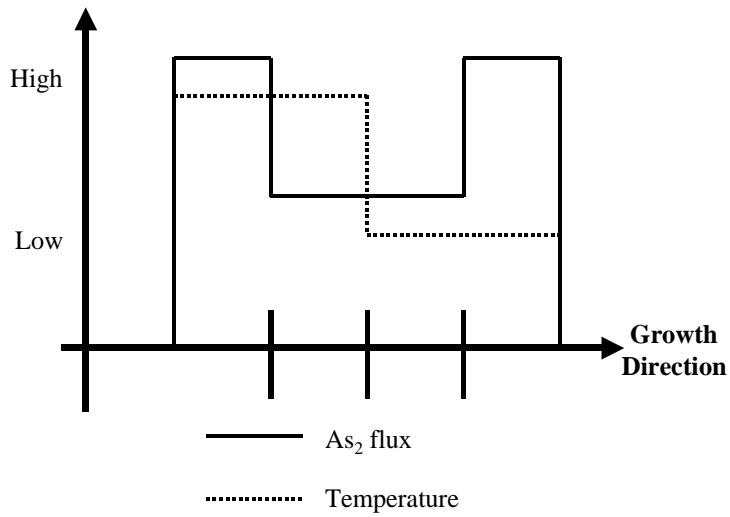


Figure 2.3: Layer structure for beryllium-doped SIMS sample

Figure 2.4: As₂ flux and temperature profile for beryllium-doped SIMS sample

crater. As the beam sputters away layers of the surface, the crater that is generated becomes more rounded with increasing sputter depth. The sputtering process is not perfectly isotropic. The effect of the rounding of the crater is the collection of atoms from layers of the semiconductor that are not as deep as those in the crater floor. This would also result in a SIMS profile that would indicate dopant diffusion when there really was none. The SIMS was redone on a different portion of the same sample using a lighter (Oxygen instead of Cesium) impinging ion in order to reduce intermixing and produce a more uniform crater. The results of this second scan were nearly identical to those of the first, indicating that the variation in apparent beryllium diffusion was not due to the aforementioned artifacts of the SIMS measurement.

Another factor that must be considered is that the deeper beryllium doped pulses were the first to be grown, and were therefore at the growth temperature for a longer time than the last pulse to be grown. It is entirely possible that the growth conditions used for the shallowest doping pulse, which were assumed to cause the least diffusion, would look far less optimal if that pulse were to be held at the growth temperature for 30 minutes while the rest of the sample was grown.

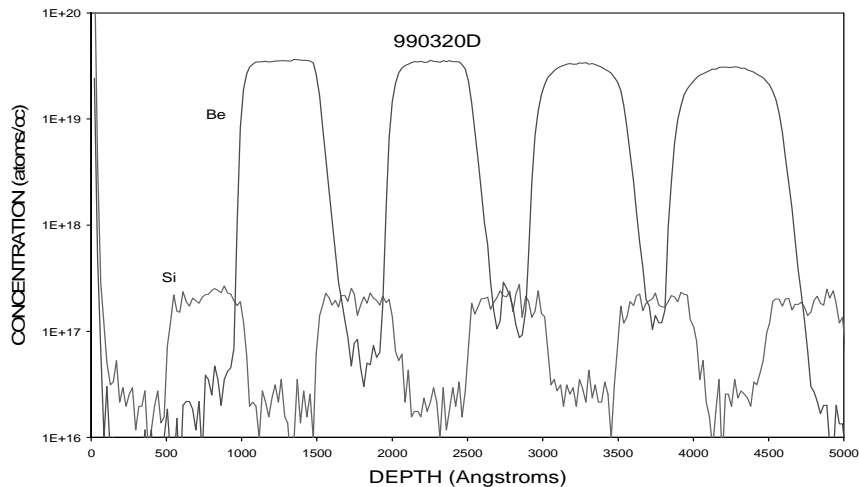


Figure 2.5: SIMS results for beryllium pulses in InGaAs

For all HBTs grown in this work, the base layers were grown with the low-temperature, high-As₂ flux condition represented by the topmost layer of the calibration sample. After the base layer is grown in an HBT, the total thickness of material grown subsequently is 2400 Å. So the length of time at which the base layer is exposed to the growth temperature is between those of the second and third doping pulse in the calibration sample. In any event, it appears that very little beryllium redistribution occurs, at least during the growth. This will be discussed further in the context of the collector design and the base-emitter junction, however, it is an interesting and surprising result in its own right; beryllium is often assumed to be by its very nature a highly diffusive dopant.

The hole concentration in InGaAs as a function of the beryllium cell temperature is determined with the use of Hall measurement on doping calibration samples. These samples consist of an undoped InAlAs buffer layer to smooth out the surface, followed by some thickness of InGaAs doped to some concentration of beryllium. Table 2.1 shows the results of five such doping calibration samples, with layer structure, beryllium cell temperature, and hole concentration indicated. The more lightly doped samples contained thicker InGaAs layers in order to make negligible the experimental error resulting from surface and back depletion of the InGaAs. Note that the $5 \times 10^{19} / \text{cm}^3$ target has not actually been realized in these doping calibrations since 1996. Another interesting point is that for the same beryllium cell temperature, the hole concentration is seen to decrease with time. It is possible that the beryllium cell has been substantially depleted over the course of the 3-4 years spanned by these calibrations. The hole mobility measured in InGaAs for this work is similar to that reported in the literature.

2.2 Regrowth

Epitaxial regrowth refers to the use of epitaxial growth processes after the initial epitaxial growth of the device layer structure following some ex-situ processing. Regrowth has been shown to improve performance in many compound semiconductor devices. It is sometimes used in the fabrication of low resistance ohmic contacts to HBTs as well as HEMTs, and is used extensively in the fabrication of confinement (current and/or index) structures in diode lasers [10]. The pros and cons of regrowth will be examined

Table 2.1: Hall measurements of beryllium doping calibration samples.

Date	Cell Temperature (°C)	Expected Concentration ($\times 10^{19}/\text{cm}^3$)	Measured Concentration ($\times 10^{19}/\text{cm}^3$)	Mobility ($\text{cm}^2/\text{V}\cdot\text{s}$)
7/96	968	5	5.9	48.6
6/97	966.8	5	3.5	54.2
01/98	979.6	5	4.2	50.5
03/99	985	5	2.8	57.2
01/98	864	0.2	0.11	107.2
03/99	880	0.2	0.15	83.6

in this section.

Regrowth has often been proposed as a solution to the tradeoff situation inherent in the base layer of the HBT [11]. The idea is to implement two different base composition/doping profiles within one device. The base layer located beneath the base ohmic contacts, referred to here as the extrinsic base, would have the following features. It would be doped as heavily as possible in order to reduce base contact and sheet resistance. It would be also be thick, in order to reduce sheet resistance. The ideal base layer beneath the active emitter is referred to here as the intrinsic base. The intrinsic base is that through which the minority carriers pass, and it has a different set of constraints. It should be very thin and somewhat more lightly doped, in order to maintain a high current gain and a low base transit time.

A simple schematic of a process flow incorporating regrowth is shown in fig. 2.6. The somewhat lighter doping in the intrinsic base simplifies the epitaxial growth of the HBT; there is no longer a need to confine as high a doping concentration to the thin base layer. An improvement in crystal quality may be seen as a higher growth temperature becomes possible.

So regrowth promises improved device performance, but with a cost. The schematic showing a regrowth process referred to earlier was oversimplified. The MBE chamber or MOCVD reactor in which the regrowth is done should not be exposed to elements such as those present in contact metallization, especially if optical devices are grown in that system. The

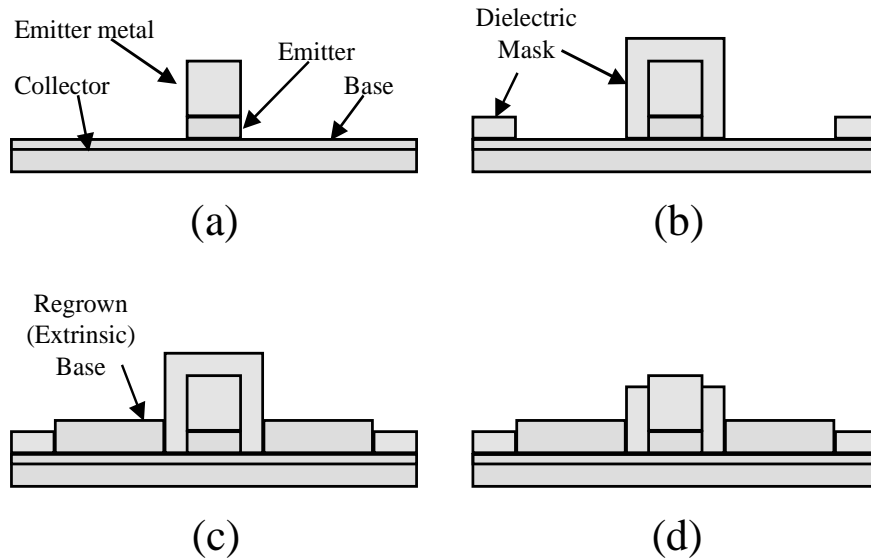


Figure 2.6: Simplified process flow for regrown base. After emitter mesa definition (a), definition of dielectric mask (b), regrowth (c), then etchback of dielectric mask for emitter contact (d).

desire to keep the MBE chamber free from contaminating metals rules out the possibility of placing a wafer with contact metal inside under vacuum and heating to growth temperature. HBT regrowth requires a "dummy" dielectric emitter finger to be deposited as a sort of placeholder for the true emitter, with processing steps subsequent to the regrowth that remove the dummy emitter and replace it with contact metallization. So the first negative consequence is a more complicated process.

The transferred-substrate process yields an HBT with scalable f_{max} . The most dramatic improvement over baseline technologies comes when the emitter finger is laterally scaled to submicron dimensions. At these dimensions, the base contact metal really must be self-aligned in order to maintain a reasonable sheet resistance in the base. In order to avoid depositing base semiconductor on the exposed emitter sidewall, it is necessary to incorporate some sort of spacer layer. But this spacer layer must be de-

fined lithographically, with an alignment. The proximity of the regrown base semiconductor to the emitter sidewall would then be determined by alignment tolerances of the lithographic tools, introducing another critical alignment to the process. It also introduces a certain minimum spacing between emitter sidewall and base contact metal. It is this minimum spacing that will determine the sheet resistance of the base contact, setting a lower bound on the doping of the intrinsic base. The benefits of regrowth mentioned previously are compromised during practical implementation.

In the current state of transferred-substrate HBT technology, the sheet resistance is no longer dominant, the current gain is sufficient for most perceived applications, the base transit time is not the dominant contribution to the delay, and the MBE growth of HBT material often becomes a bottleneck to throughput. All four of these factors serve to dilute the gains that regrowth promises. As this project continues to push to deeper submicron dimensions as well as simplify the processing (goals that at times are in conflict), the use of regrowth in the near future is highly doubtful.

2.3 Carbon Doping

The problems of beryllium diffusion have been discussed earlier in this section, and growth conditions that minimize the diffusion of beryllium have been examined. The best solution to the beryllium diffusion problem, if indeed there is one, is to not use beryllium. Carbon is known to have a diffusion coefficient roughly three orders of magnitude less than beryllium at similar temperatures and concentrations. Carbon doping promises far-reaching benefits for HBT technology, while introducing its own unique set of problems.

In this work, a carbon tetrabromide (CBr_4) source was fabricated and installed on a solid-source MBE. This was a collaborative effort with fellow graduate student Ryan Naone and MBE Lab Supervisor John English. The successful CBr_4 source of Dr. David Miller's (Pennsylvania State University) was used as a model from which to build on for the fabrication of the UCSB carbon source [12]. It was shown that the use of CBr_4 in this way did not degrade optical quality in material grown on that system. If anything, material quality was enhanced.

Graphite filaments are commercially available which will provide carbon doping with a solid source. One of these units is currently in use at

UCSB. This filament source has demonstrated some rather severe drawbacks. Degraded surface morphology occurs at concentrations above the low $10^{18}/\text{cm}^3$ range. No growth conditions were found that would alleviate the problem. Reliability problems surfaced, possibly from the mechanical construction of the source. Occasional loss of electrical continuity through the circuit containing the filament itself would render the doping source useless until the cell could be fixed. Also, the filament radiates enough heat to substantially affect the substrate temperature. The gas source would not introduce any extra heat to the system, and it had been proven in other laboratories to successfully provide high doping concentrations with good surface morphology.

A schematic of the carbon source is shown in fig. 2.7. The CBr_4 itself had been vacuum sublimed into a stainless steel cylinder. CBr_4 vapor passes through a leak valve, which is set by the user to deliver the required flux of CBr_4 . The vapor then is directed either into an ion pump, where the flux of CBr_4 can be measured by monitoring ion pump current, or is diverted to the system for doping. This is the so-called RUN/VENT method, where RUN corresponds to doping the growing sample, and VENT represents diversion of CBr_4 to the ion pump. The ability to monitor ion pump current in order to determine doping level is extremely useful, as ion-gauge beam flux readings change with introduction of a new filament or aging of the current one, and the leak valve exhibits hysteresis.

A series of doping calibration wafers were grown in order to characterize carbon doping in $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$. The layer structure was a 2500 Å unintentionally doped InAlAs buffer layer followed by the carbon doped InGaAs layer. The thickness of the InGaAs layer ranged from 3000 Å (for the samples with the highest doping) to 5000 Å. The results of hole concentration (determined by Hall measurement) vs. hole mobility are plotted in fig. 2.8, along with similar data points obtained with beryllium doped InGaAs. A hole concentration of $1.2 \times 10^{20} / \text{cm}^3$ was obtained with a hole mobility of 43 $\text{cm}^2/\text{V}\cdot\text{sec}$. The hole mobility in the carbon doped material can be seen to be somewhat higher than beryllium doped material for a given doping level.

A concern with carbon doping not faced with beryllium doping is its tendency to become n-type with sufficient In composition in the Group-III sublattice. As a Group-IV element, this amphoteric nature should perhaps not be a surprise. Carbon is known to act as a p-type dopant on (100)

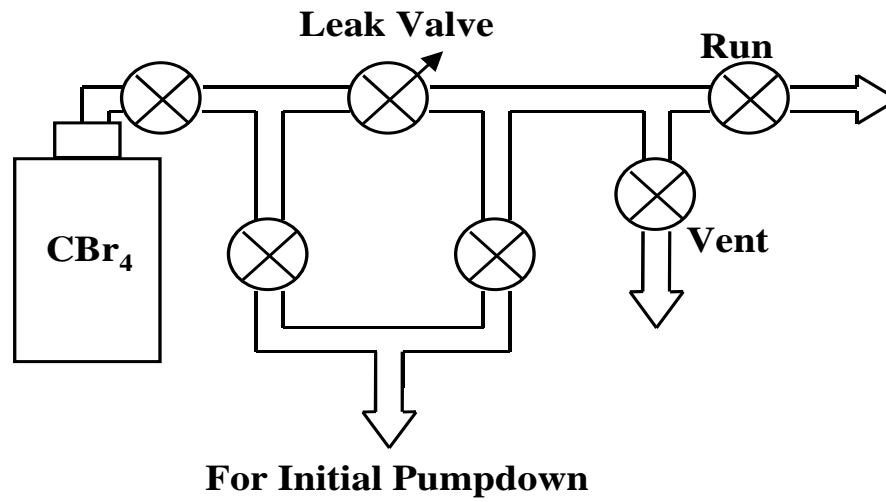


Figure 2.7: Schematic of carbon source gas manifold

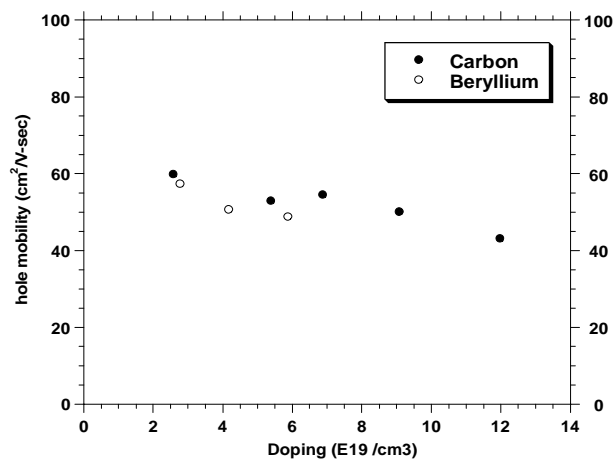


Figure 2.8: Hole mobility as a function of P-type doping concentration.

GaAs, and a rather inefficient n-type dopant in (100) InAs. Fortunately, no substantial compensation was discovered at even the highest doping levels attempted in $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$. Growth conditions for the carbon doped layers were modified from those of beryllium doped layers. The V/III ratio was decreased in order to facilitate carbon incorporation on the Group V site, and the substrate temperature was 420-430°C instead of 380°C. Surface morphology on 0.5 μm thick calibration samples of carbon doped InGaAs was noticeably better than the surface morphology on more lightly beryllium doped samples of comparable thickness. The improved morphology as well as the improved hole mobility are attributed to the modified growth conditions (higher T_{sub} and reduced As_2 flux).

In order to take full advantage of the reduced diffusion coefficient of the carbon acceptor, the source should be able to start/stop doping with carbon over time scales that are short compared to the growth of a few monolayers. The turn-on time of the source is cause for concern due to the physical construction of the source. The RUN/VENT switch is located a few feet from the actual point of injection into the system. By monitoring the beam flux gauge, Ryan Naone determined that the time to reach 90 % of the desired doping level is 10 seconds (10 second rise time), and a similar fall time after switching to the VENT mode. At 1 $\mu\text{m}/\text{hour}$ growth rate, 10 seconds corresponds to about 28 Å. The fall time of the carbon flux is the time taken for the length of line between the point of injection and the RUN/VENT switch to be pumped out by the system when switching from the RUN mode to the VENT mode. The rise time corresponds to filling this line with CBr_4 when switching from the VENT to the RUN mode. This problem would be fixed with the introduction of a mechanical shutter blocking line-of-sight paths from the injector to the wafer. That way, the RUN mode could be entered 10 seconds before the doping was actually required, and switching to the VENT mode could be accompanied by closing the shutter, immediately blocking CBr_4 from the wafer.

The finite fall time response of the carbon source can be circumvented with a growth interrupt, resuming growth after CBr_4 has been pumped out of the RUN line for some time. The effect of the finite rise time can be reduced by beginning carbon doping slightly before initiating growth of the layer to be doped with carbon. If growth has been interrupted to do this, care must be taken not to excessively delta-dope with carbon before growth of the intentionally carbon doped layer by waiting too long to initiate

growth.

The response of the carbon source was also examined by SIMS analysis of a structure with carbon doped pulses and silicon doped marker layers. Plotted in figs. 2.9 and 2.10 is the intended profile along with the SIMS analysis of that profile. The carbon doping edges show a somewhat more gradual slope than the silicon doping does indicating either carbon diffusion or, more likely, finite flux transient response. For this sample, no growth interrupts were implemented, nor was any attempt made to reduce the effect of the flux transients in any way.

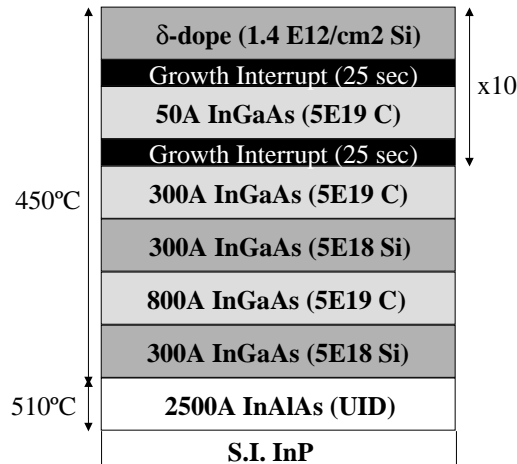


Figure 2.9: Layer structure for carbon-doped SIMS sample.

So far, the carbon source has demonstrated exciting potential for use in transferred-substrate HBTs. But what obstacles remain before it can practically be implemented? The biggest hurdle has been development of a base contact etch process that does not passivate the carbon acceptors. The current hydrogen based dry etch chemistry renders a carbon-doped base useless. Here at UCSB, Dr. Yoram Betser has worked to develop a chlorine-based dry etch process that would be compatible with carbon. This process needs further refinement in order to realize high yield in submicron emitter devices. Another problem is that the method in which the MBE growth

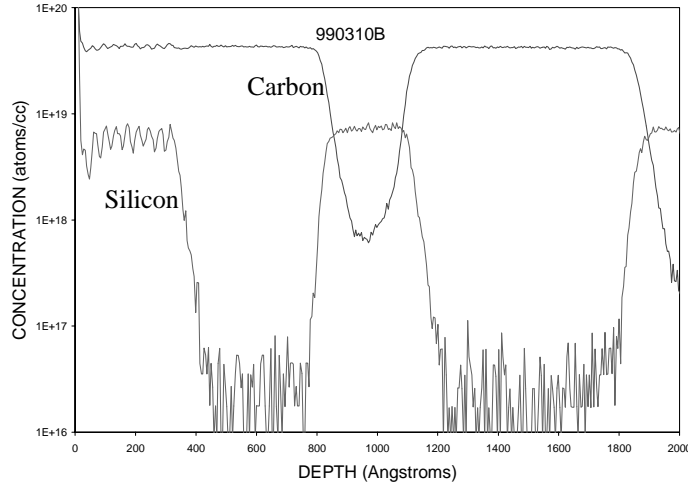


Figure 2.10: SIMS results showing scans for carbon and silicon.

of the graded base is performed is incompatible with the turn-on transient of the source. Recall that the base grading is implemented by a series of 50 Å layers separated by growth interrupts that allow the gallium cell to stabilize at its new setpoint. Each one of these layers takes only 18 seconds to grow, a 0-90 % risetime over 10 seconds would result in a base profile with significant modulations in the base doping. It may therefore be necessary to perform an analog grade in the base with no growth interrupts during base growth or to open the RUN valve early before each layer resulting in a slight delta-doping at each interface.

2.4 Collector Layer Growth

The best growth conditions for the collector are those which tend to reduce the possibility of scattering as the electrons pass through it. A growth temperature from 480°C to 510°C has been used for the InGaAs collector, high enough for good crystal quality. The As₂ flux used should be sufficient to keep the surface just As₂ rich, but not more. For the early part of this work, the As₂ flux used for growth of the collector was the same high As₂

flux used for the base and emitter, but recently it has been reduced by $\sim 40\%$ according to the beam flux gauge. No improvements in device performance have been noted as a direct result of this change, but it is believed that the observed substantial reduction in background pressure results in a decreased incorporation of impurities.

Hall measurements were performed in a similar fashion as described in section 2.1, but for lightly silicon doped layers of InGaAs. The target doping was the desired collector doping of $1 \times 10^{16} / \text{cm}^3$. Such a low doping level requires a very thick layer to be grown if the effects of surface and back depletion are to be neglected. Silicon-doped InGaAs was grown to at least $4 \mu\text{m}$. Due to imperfect lattice matching and/or the natural accumulation of defects in an InGaAs layer that thick, the surface morphology of these doping calibration samples has never been good. With the sample illuminated by a flashlight or a laser beam, the surface roughness is visible to the naked eye. Nevertheless, as can be seen by consulting table 2.2, the electron mobility in these samples approaches $9,000 \text{ cm}^2/\text{V}\cdot\text{sec}$ at room temperature, a respectable value for this doping level in InGaAs.

Table 2.2: Hall measurements of silicon doping calibration samples.

Date	Cell Temperature ($^{\circ}\text{C}$)	Expected Concentration ($\times 10^{16}/\text{cm}^3$)	Measured Concentration ($\times 10^{16}/\text{cm}^3$)	Mobility ($\text{cm}^2/\text{V}\cdot\text{s}$)
7/96	1041	1	1.6	7430
6/97	1025.6	1	1.2	8190
1/98	1026	1	0.79	8850

One variation of the baseline epitaxial structure of fig. 2.1 is to include an n+ subcollector layer, as shown in fig. 2.11. This layer is doped at or above $1 \times 10^{19} / \text{cm}^3$ with silicon and the highest As_2 flux is used for growth of this layer, in order to improve the activation efficiency of silicon.

On the collector side, beryllium outdiffusion can be viewed as simply extending the width of the base at the expense of the collector. This would be expected to increase base transit time and decrease current gain. In order to reduce the effect of beryllium diffusion into the collector, a thin, heavily doped n-type layer was placed near the base. This is the layer

400Å InGaAs (1E16 Si)	250Å InGaAs (1E16 Si)
50Å InGaAs (1E18 Si)	50Å InGaAs (1E18 Si)
2550Å InGaAs (1E16 Si)	1700Å InGaAs (1E16 Si)
2500Å InAlAs (UID)	1000Å InGaAs (1E19 Si)
S.I. InP	2500Å InAlAs (UID)
	S.I. InP

(a) (b)

Figure 2.11: Collector Layer Structures, a) Baseline Schottky collector, b) Ohmic collector for CML

labeled "pulse doping" in fig. 2.1. The intended effect was to increase the voltage dropped between this layer and the base in an effort to deplete much of the outdiffused beryllium. (Comment: for most of the duration of this work, it was simply assumed that beryllium diffuses severely, as this is what journal articles and textbooks indicate. The SIMS data gathered and discussed in section 2.1 indicate that the amount of beryllium diffusion present in this work was grossly overestimated. The implications of this as well as the results of the countermeasures taken such as the pulse doping are discussed in sections 4.4 and 4.8 on the collector transit time and Kirk Effect, respectively).

2.5 Emitter Layer Growth

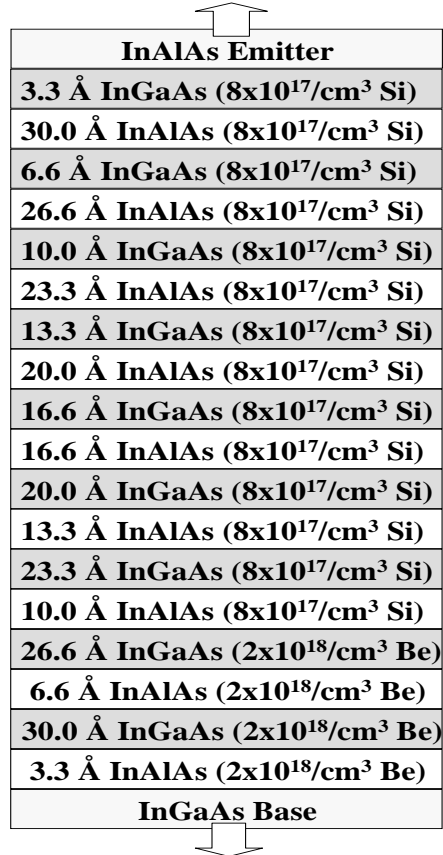
It has been established that the beryllium diffusion is not well controlled, and can lead to irreproducibility of device performance. Accepting that beryllium diffusion will happen, and to varying degree, the layer structure can be designed with reduced sensitivity of device parameters to some beryllium diffusion. In an abrupt base-emitter junction, beryllium diffusion of only a few tens of Angstroms has been shown to drastically degrade current gain and f_τ . Base-emitter junctions which are graded over a few hundred

Angstroms exhibit current gain and f_T that are more slowly varying functions of beryllium diffusion distance [13].

The graded base-emitter junction is grown with a graded superlattice, as shown in fig. 2.12. Wafer rotation is increased to 20 RPM in order to ensure good uniformity across the wafer. An analog linear grade, where the aluminum and gallium compositions change smoothly from one end to the other, is not straightforward in MBE. It requires precise knowledge of the thermal response of the gallium and aluminum cells as the heater power is adjusted. This thermal response changes as the cell becomes more and more depleted through use, and so reproducibility of the base-emitter junction grading may not be possible. Analog grades are simple to implement in MOCVD, however, as they can be implemented by programming the mass flow controller to increase or decrease the relevant Group-III gas linearly as a function of time.

The graded superlattice has a possible advantage over the analog grade with its multiple interfaces, provided they are of high quality and do not cause excessive scattering [14]. Beryllium diffusion into the emitter may be suppressed if the beryllium preferentially incorporates into one of the layers of which the superlattice is composed.

Hall measurements have been taken of heavily silicon-doped InAlAs and InGaAs layers. The surface morphology of these layers has always been rough, probably due to the low growth temperatures used. The growth of these doping calibration samples is intended to replicate the growth of the corresponding layers in the HBT, and so the temperature at which these samples are grown is 380°C. For the same low substrate temperature, high As₂ flux, and silicon cell temperature, the electron concentration in InAlAs is lower than that of InGaAs by ~20%. A theory proposed in the literature [15] that may explain this is that for InAlAs, higher growth temperatures are required to achieve the highest electron concentrations with the silicon dopant. It is put forth that this phenomenon is not seen in the GaAs/AlGaAs system and is due to two electron trapping centers that appear in InAlAs grown at temperatures in the vicinity of 400°C. The optimum growth temperature found in that work for activation of silicon in InAlAs was 560°C, which the authors admit is a very high temperature for the growth of InAlAs. Upon first inspection, it was suspected that this high growth temperature was simply an anomaly of the temperature measurement apparatus of that particular MBE system. This idea was



InAlAs Emitter
3.3 Å InGaAs ($8 \times 10^{17}/\text{cm}^3$ Si)
30.0 Å InAlAs ($8 \times 10^{17}/\text{cm}^3$ Si)
6.6 Å InGaAs ($8 \times 10^{17}/\text{cm}^3$ Si)
26.6 Å InAlAs ($8 \times 10^{17}/\text{cm}^3$ Si)
10.0 Å InGaAs ($8 \times 10^{17}/\text{cm}^3$ Si)
23.3 Å InAlAs ($8 \times 10^{17}/\text{cm}^3$ Si)
13.3 Å InGaAs ($8 \times 10^{17}/\text{cm}^3$ Si)
20.0 Å InAlAs ($8 \times 10^{17}/\text{cm}^3$ Si)
16.6 Å InGaAs ($8 \times 10^{17}/\text{cm}^3$ Si)
16.6 Å InAlAs ($8 \times 10^{17}/\text{cm}^3$ Si)
20.0 Å InGaAs ($8 \times 10^{17}/\text{cm}^3$ Si)
13.3 Å InAlAs ($8 \times 10^{17}/\text{cm}^3$ Si)
23.3 Å InGaAs ($8 \times 10^{17}/\text{cm}^3$ Si)
10.0 Å InAlAs ($8 \times 10^{17}/\text{cm}^3$ Si)
26.6 Å InGaAs ($2 \times 10^{18}/\text{cm}^3$ Be)
6.6 Å InAlAs ($2 \times 10^{18}/\text{cm}^3$ Be)
30.0 Å InGaAs ($2 \times 10^{18}/\text{cm}^3$ Be)
3.3 Å InAlAs ($2 \times 10^{18}/\text{cm}^3$ Be)
InGaAs Base

Figure 2.12: Layer structure of the baseline base-emitter grade. Alternating layers of InGaAs and InAlAs are used to approximate a linear grade from the InAlAs emitter to InGaAs. The superlattice period is 33.3 Angstroms, and there are nine periods.

discounted by the authors' statement that oxide desorption on InP occurs at 540°C, a similar temperature to that used at UCSB for oxide desorption.

Table 2.3: Hall measurements of heavily silicon-doped calibration samples.

Date	Cell Temperature (°C)	Expected Concentration ($\times 10^{19}/\text{cm}^3$)	Measured Concentration ($\times 10^{19}/\text{cm}^3$)	Mobility ($\text{cm}^2/\text{V}\cdot\text{s}$)
InGaAs	1320.4	1	1.2	2000
InGaAs	1320.4	1	1.2	2000
InGaAs	1320.4	1	0.87	500

The growth of transferred-substrate HBTs is emitter-up; the emitter is grown after the base. This precludes the use of such high-temperature grown InAlAs layers, as the beryllium-doped base will see this temperature for nearly 30 minutes. Carbon doping may make high-temperature emitter growth a possibility.

The advantage of the emitter-up growth is that the emitter cap can be very heavily doped and/or strained without affecting the material quality of the other, more sensitive layers in the HBT. The emitter ohmic contact resistance must be minimized in high-speed devices; heavily doped, narrow bandgap materials can be used for the contact layer. Most of the devices in this work utilized the emitter cap of the baseline layer structure, however, for reasons discussed in section 4.6.

2.6 Summary

It is an oversimplification to say that the transferred-substrate process yields HBTs with superior bandwidth to other HBTs due to lithographic scaling. The high device bandwidths demonstrated by this group would not have been possible were it not for constant attention to and development of the basic MBE growth technology. Beryllium diffusion and Kirk effect (base pushout at high currents, to be discussed in section 4.7) are two examples of past headaches that have, for the moment, been solved by improved growth conditions and a modified layer structure. The new carbon and

phosphorous sources will, in the future, allow much higher base doping concentrations to be achieved and greatly improve the device breakdown voltage, respectively.

This chapter discussed MBE growth conditions and some of the ramifications for device performance. Much of this work is ongoing, as new materials are being incorporated into the layer structure. The ability of this group to grow material, make changes to the MBE system, process material in the cleanroom, and do high frequency testing all at UCSB allows tremendous adaptability. This adaptability has been what allows the transferred-substrate process to maintain the leading edge in device and circuit bandwidth. It has allowed this group to keep pace with other new developments in the field and incorporate the more promising innovations. Constant improvements in the basic technology combined with the aggressive scaling that the transferred-substrate process allows have resulted in device and circuit performances far beyond the present state-of-the-art for HBTs.

Chapter 3

Processing

The transferred-substrate process has evolved from a proof-of-concept vehicle to demonstrate a new scaling law for HBTs into a full-fledged high-speed IC process. The current state of the process provides NiCr resistors, metal-insulator-metal (MIM) capacitors, three levels of interconnect metallization, microstrip wiring, low inductance grounds, and device heatsinking. It will be beneficial to examine the process in detail. The implementation of these features can be explained, along with a "wish list" of things that could be better and may be addressed in future development of this IC process. Along the way it will be made clear how the transferred-substrate process addresses some of the common problems of IC processes that exist today from the standpoints of ease of packaging and thermal management. Some of the specific chemicals and conditions used during the process can be found in Appendix A, the process flow sheets.

3.1 Process Flow

The process begins with the definition of the emitter contact metal. The emitter contact metal listed from the semiconductor up is titanium, platinum, gold, and silicon (Ti/Pt/Au/Si). The titanium promotes adhesion to the semiconductor, the platinum is a diffusion barrier to gold during the heat treatments, the gold is the low resistance contact metal, and the silicon is an etch mask to prevent the gold from sputtering away during the following dry etch. After liftoff of the emitter metal, the base contact etch follows. The thickness of semiconductor above the base is about 2500 Å.

The base thickness itself is typically only 400 Å. In order to expose the base semiconductor without under or overetching, a somewhat complicated but effective etch scheme is employed. The first step is a dry etch in the reactive ion etching (RIE) system. The dry etch removes the bulk of the emitter semiconductor in a controlled fashion. Referring to the layer structure of fig. 2.1, the emitter is etched in RIE with methane/hydrogen/argon and using a laser monitor. The laser monitor signal is used to stop the etch when it nearly reaches the beginning of the base-emitter superlattice grade. When this etch is complete, there is a faint blue ring around the perimeter of the wafer, due probably to nonuniformity in the etch resulting in slightly thicker material near the perimeter. It is supposed that this is due to the nature of the gas flow in the reactor, and to the mechanics of this particular etch process.

After the remaining silicon is removed in CF_4 plasma, the base contact etch process continues. A selective wet etch is then used to undercut the emitter semiconductor and to restore uniformity to the exposed semiconductor. In the vertical direction, this etch stops when it reaches that point in the base-emitter superlattice where there is sufficient gallium content. It will continue to laterally etch the InAlAs under the emitter contact metal, and if allowed to proceed for too long, the emitter semiconductor and metallization will be completely lifted off. Due to the efficiency with which this etch undercuts the emitter, recent process runs have made use of a low temperature bath in order to slow the etch rate down and make it more reproducible. When this etch is complete, the base is exposed via a short, timed nonselective etch that removes the last ~ 100 Å of base-emitter grade above the base.

With the base etch complete, the base metal is deposited self-aligned to the emitter metal, and then the base and emitter metal undergo heat treatment in the rapid thermal annealer (RTA). Device isolation is performed in RIE, which etches down to the InAlAs buffer layer. Polyimide is spun onto the wafer, with the dual purposes of passivating the exposed emitter-base sidewall and planarizing the topography that now exists on the wafer. The polyimide is etched back in RIE until the tops of the emitters are exposed so that they can be contacted with interconnect metal. A schematic flow of the process up to this point is shown in fig. 3.1.

Polyimide etch lithography leaves small islands of photoresist over most of the area of the base mesas, but removes photoresist everywhere else. The

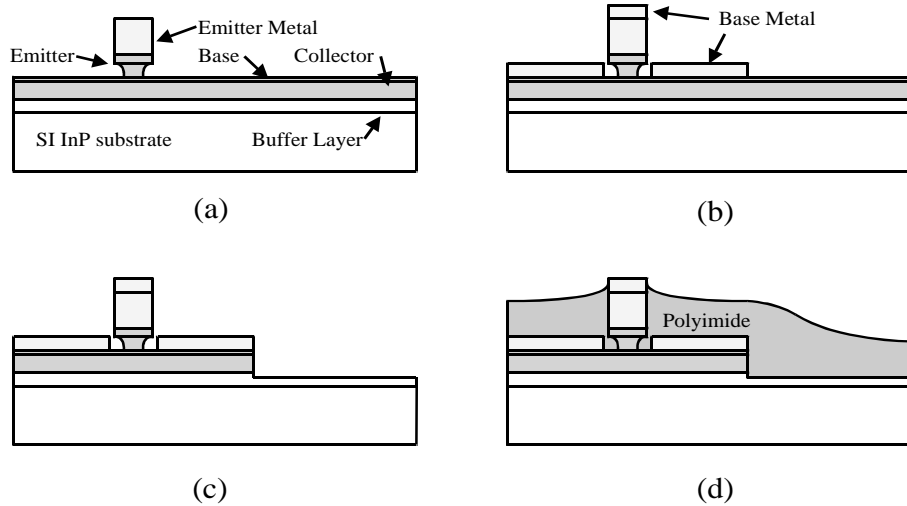


Figure 3.1: Process flow diagram after deposition of emitter metal and base contact etch (a), after deposition of base metal (b), after isolation etch (c), after polyimide etchback (d).

exposed polyimide is then etched completely away to reexpose the InAlAs buffer layer between device mesas and to remove polyimide from the edges of the base mesas themselves. In this way, interconnect metal can also contact the base metallization.

NiCr resistors are deposited to a thickness which gives approximately $50 \Omega/\text{square}$. The first layer of interconnect metal (hereafter called metal 1, or M1) is deposited to contact the device emitters and bases and also to connect to the ends of the resistors. A 4000 \AA thick layer of silicon nitride is deposited to serve as the insulator for MIM capacitors and, as described later, to prevent chosen device emitters from being grounded. Those devices that will have grounded emitters will have the silicon nitride protecting those emitters etched away. The silicon nitride etch also allows for connection of metal 1 to the next layer of interconnect to be deposited, metal 2. Metal 2 serves many purposes, it is the bottom plate of MIM capacitors, it is another layer of interconnect, and it will protect the silicon nitride from attack by the next dry etch which will soon be described. After metal 2 is lifted off, benzocyclobutene (BCB) is spun onto the wafer to a thickness of $5 \mu\text{m}$. BCB acts as the artificial substrate used for the

microstrip dielectric, and was chosen for its low loss and low dielectric constant. After the BCB is cured, the nickel etch mask is deposited and lifted off where it is desired to etch vias into the BCB. Vias are etched into the BCB by RIE, and this etch continues on to attack the silicon nitride if not protected by metal 2. For this reason, emitters not intended to be grounded must have a rectangle of metal 2 above the silicon nitride.

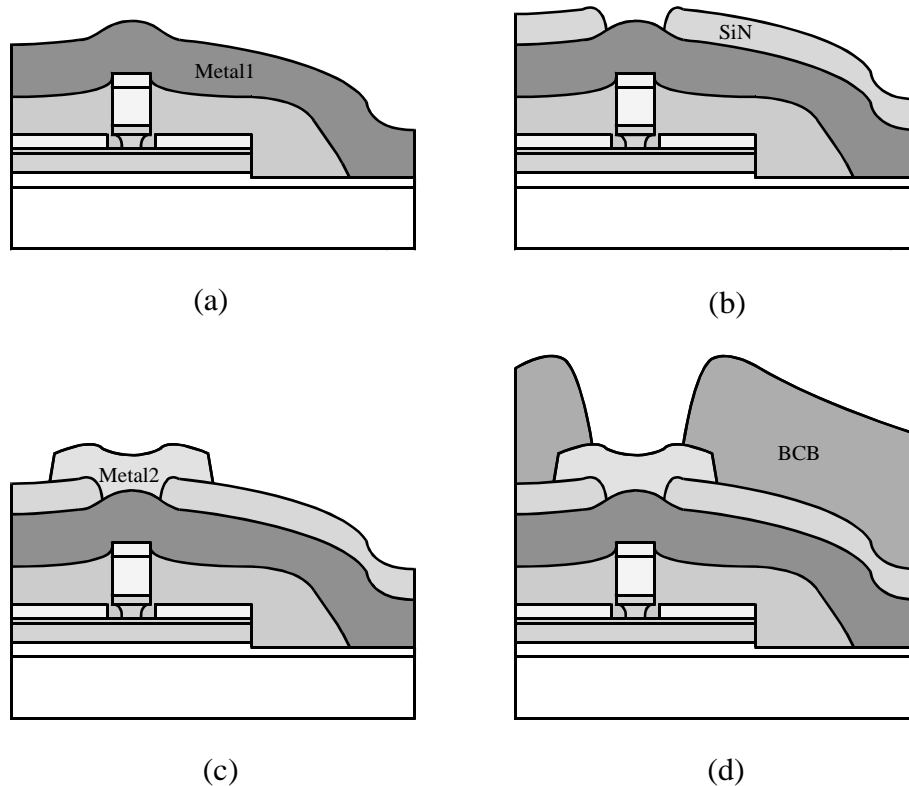


Figure 3.2: Process flow diagram after deposition of metal 1 (a), after etching of SiN (b), after deposition of metal 2 (c), after etching of BCB (d).

Thin layers of Ti/Au are sputtered onto the wafer, the Ti again to promote adhesion, and the gold to form a constant potential surface for gold plating. The gold ground plane is then plated onto the wafer, to an approximate thickness of $10\ \mu\text{m}$. The wafer is then solder bonded to a similar shaped piece of GaAs for mechanical support. The InP substrate is removed in an aqueous HCl solution that does not attack the InGaAs collector at

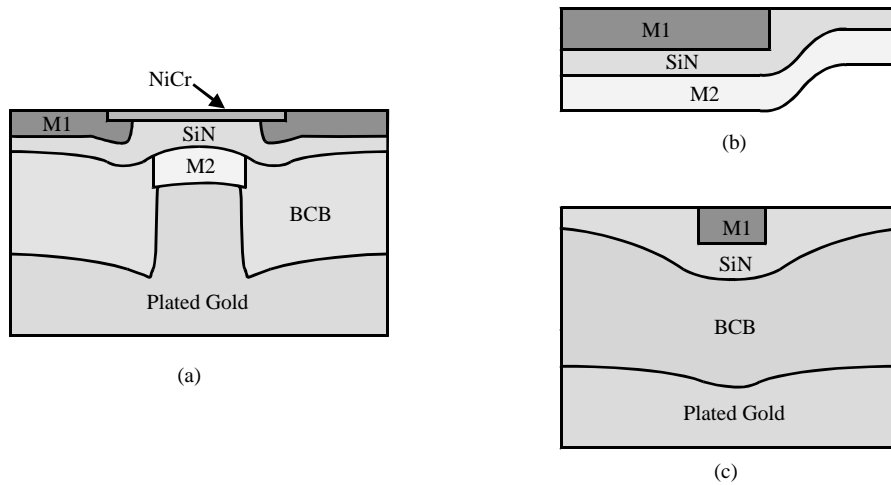


Figure 3.4: Passive elements in transferred-substrate process: NiCr resistor with heat sink (a), MIM capacitor (b), microstrip interconnect (c).

selective wet etch time is sufficient to reach the critical point in the base-emitter grade in the center of the wafer, the result will be uniformity. This is because the selective etch cannot proceed past that critical layer anywhere, at the edges of the wafer which were dry-etched deeper the selective wet etch will simply hit the stopping point earlier.

In spite of the lithographic definition of both emitter and collector stripes, it is still necessary to minimize the size of the base mesa. This has nothing to do with collector-base capacitance, rather, it must be done to reduce base-emitter capacitance. Referring to fig. 3.2(a), it can be seen that the emitter interconnect metal crosses over the base metal, and there is capacitance between them through the polyimide. The intervening polyimide can only be as thick as the sum of the emitter metal and emitter semiconductor, minus the thickness of the base contact metal. The dielectric constant of the polyimide is 3.4, and the thickness is typically 5000 \AA . A $1 \times 8 \text{ \mu m}$ emitter finger has typically been designed to have 2 \mu m of base mesa on either side of it. The width of the interconnect crossing over this base metal is 6 \mu m . The area of overlap is then 24 \mu m^2 , leading to an additional capacitance of 1.5 fF between base and emitter. This capacitance is referred to as C_{la} , as it is a parasitic capacitance resulting from the physical layout of the device. C_{la} is discussed in more detail in section 5.1.

One process run has been performed with the base mesa extending only $1 \mu\text{m}$ out on either side of the emitter. At the time this was determined to be too aggressive as the stepper was consistently aligning only to within $0.3 \mu\text{m}$. In this case, the stepper misalignment was about that, $0.3 \mu\text{m}$. This leaves a $0.7 \mu\text{m}$ wide strip of base metal on one side of the emitter finger, and $1.3 \mu\text{m}$ on the other. The thickness of the base metal is 1200 \AA , 700 \AA of this is gold, the balance is made up of the titanium and platinum contact layers, to simplify this calculation it will be assumed that there is 1200 \AA of gold. The resistivity of gold is $2.44 \times 10^{-6} \Omega\text{-cm}$. The resistance incurred by traveling down the length of the base mesa alongside the emitter finger on either side of the emitter finger is $R = \rho L/A$, where L (here assumed to be $12 \mu\text{m}$ is the length of the emitter, and A is the cross-sectional area through which the current travels. So A is given by the product of the thickness of the base metal and the distance between the edge of the emitter and the edge of the base mesa. The side of the emitter where the base mesa is only $0.7 \mu\text{m}$ wide has an end to end resistance of 3.5Ω . The $1.3 \mu\text{m}$ side has a resistance of 1.9Ω . Is this significant? It is indeed; the R_{bb} of devices with this geometry and a 400 \AA base is approximately 14Ω , the 3.5Ω is 25 % of this, the 1.9Ω is only 14 % of this, resulting in uneven turn-on of the device that progressively becomes worse as the far end of the emitter is approached.

Another possible problem resulting from the stepper misalignment regards base Ohmic transfer length L_t . The L_t is the distance over which carriers tend to spread when entering or leaving Ohmic contact metal; contact metal of more than about two times wider than the L_t is required in order not to increase the resistance. L_t is given by $\sqrt{r_c/R_{sh}}$, where r_c is the specific contact resistance of the base Ohmic, and R_{sh} is the sheet resistance through the base material. Fairly typical values for the 400 \AA , $5 \times 10^{19}/\text{cm}^3$ doped base used in the baseline material are $r_c = 5.5 \times 10^{-7} \Omega\text{-cm}^2$ and $R_{sh} = 800 \Omega$ per square. This gives an L_t of about 2500 \AA and so even with $0.3 \mu\text{m}$ of misalignment, the 7000 \AA of base mesa width on the one side of the emitter finger is 2.8 transfer lengths wide, resulting in a negligible effect on the overall base resistance.

For a device operating under $I_c = 10\text{mA}$ and with a β of 50, the DC base current I_b is $200 \mu\text{A}$. The voltage drop across the base metal along the length of the device is given by this base current multiplied by the parallel combination of the two base metal resistances (one on either side of

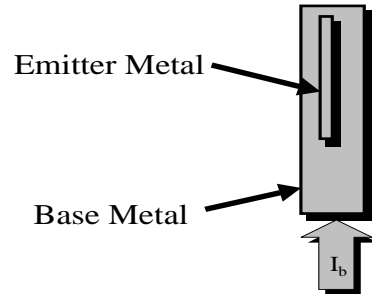


Figure 3.5: Schematic of base mesa and emitter finger for misaligned emitter finger.

the emitter finger for base current traveling through the base metal down the length of the emitter finger). The parallel resistance is 1.2Ω and the resulting voltage drop is 24 mV. This resistance is independent of stepper misalignment, it is dependent upon the total width of the base mesa (minus the width of the emitter finger). A 24 mV drop from one end of the device to the other is very significant given the $e^{qV/kT}$ behavior of the bipolar transistor, mitigated somewhat by the parasitic R_{bb} and R_{ex} Ohmic contact resistances to the base and emitter, respectively. The transistor is not being driven equally on both ends. What has happened here is that the base mesa width has been scaled, but not the length of the device. A lower β will exacerbate this problem, the higher I_b will further increase the voltage drop. Subsequent process runs had the base mesa extending $2 \mu\text{m}$ out from either side of the emitter, reducing the magnitude of this problem by a factor of two.

So why is the base metal so thin? The thickness of the base metal is limited by the thickness of the emitter semiconductor, more specifically, the thickness of the InAlAs. Fig. 3.6 shows a cross-section of the emitter finger and exposed base semiconductor. If base metal is deposited to a thickness greater than the thickness of the undercut semiconductor, there is the danger of base-emitter shorting. At least one HBT process that uses the self-aligned deposition of base metal deposits the base metal to a much greater thickness than the emitter semiconductor. The gap between the adjacent base and emitter metallization becomes as narrow as a few

hundred angstroms at one point. The viability of this technique has not been demonstrated in an IC process as yet, and so for now the base metal is simply kept thin.

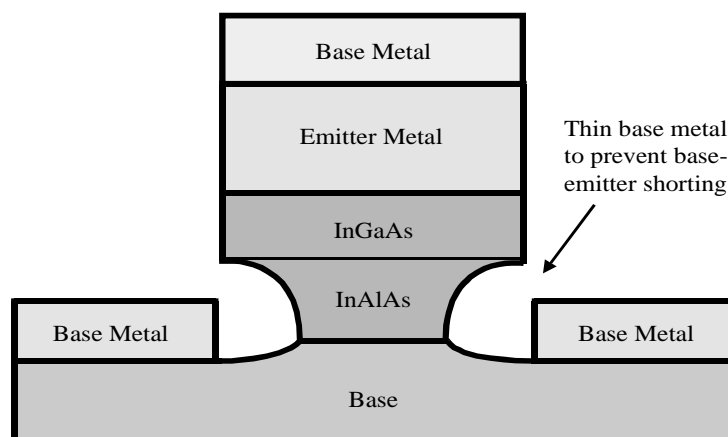


Figure 3.6: Cross-section of emitter-base junction

The capacitance between metal 1 and metal 2 is $0.128 \text{ fF}/\mu\text{m}^2$ through the 4000 \AA of silicon nitride. The parallel-plate capacitance from either metal 1 or metal 2 to ground through the BCB is approximately $.025 \text{ fF}/\mu\text{m}^2$, and substantially more than this for narrow lines due to the fringing fields. Metal 3 (collector) can be used as an interconnect, and polyimide can be left beneath it to reduce its parasitic capacitance to metal 1, metal 2, and ground. These crossover and parasitic capacitances must be accounted for in simulations, especially in large chips with multiple crossovers.

3.3 Heat Sinking and Microstrip Interconnects

The transferred-substrate process provides a via for heatsinking of each device. The heat generated in a grounded-emitter device must exit the semiconductor through the emitter contact metal, out through the interconnect metal to the gold ground plane. In a non-grounded device, the heat must pass through the silicon nitride and the metal 2 before reaching the

ground plane. In either case, the majority of the temperature drop, that is, the majority of the thermal resistance, is incurred across the emitter semiconductor itself. The thermal conductivities of InGaAs and InAlAs are extremely low, on the order of 0.05 W/cm-K, to be compared with 0.46 W/cm-K for GaAs, 0.68W/cm-K for InP, 1.5 W/cm-K for Si, and 0.16 W/cm-K for Si₃N₄.

This is motivation for the following question: Why does the emitter semiconductor need to be so thick? The depletion region of the base-emitter junction is on the order of a few hundred angstroms thick, so why the need for 1400 Å of InAlAs with a 1000 Å InGaAs cap? The 1400 Å of InAlAs is deemed necessary because it sets an upper limit on what is felt to be a safe thickness for base contact metallization per the above discussion on base metal thickness. The InGaAs cap need not be so thick as 1000 Å. It has been kept at this thickness so that a higher temperature sinter could be used on the emitter metal if so desired. A higher temperature sinter would cause the emitter metal to react more with the semiconductor, setting a lower limit to the thickness of this layer. If the minimum thermal resistance is desired, InP should replace the InAlAs, and the InGaAs cap should be made thinner. This idea so far has not gained serious consideration, as it would require complete redevelopment of a currently successful base-emitter etch process.

The heatsinking via adds a parasitic capacitance to the transistor from the emitter to ground. In a non-grounded emitter device, the capacitance is between the metal 2 layer and the ground plane itself, through the 4000 Å of silicon nitride. This capacitance can only be reduced if smaller via holes could reliably be opened in the BCB. The BCB etch is isotropic for the most part, leading to sloped sidewalls. If the via hole becomes much smaller, the sloping sidewalls would require a large amount of overetch in order to open a hole through the BCB. The overetch required would cause shorting between metal 1 and metal 2. If smaller via holes are desired, it will be necessary to use thinner BCB or photo-definable BCB that can be exposed and developed away much like photoresist. A thicker layer of silicon nitride would also reduce the capacitance, but at the expense of added thermal resistance. This capacitance is about 0.128 fF/μm² of emitter area with the current design rules and nitride thickness.

A question often arises as to why the device is heat sunk from the emitter side instead of the collector side. The argument put forth is that the

collector is where the majority of the heat is generated in the device, and heat sinking the collector would be more effective. This is true. But the parasitic capacitance that the heatsink introduces is a significant fraction of the collector-base capacitance and might become the performance limiting parameter in even conservatively scaled devices. The base-emitter capacitance at 1×10^5 A/cm² current density is ~ 20 fF/ μm^2 of emitter area, so the effect of the heatsink there is less noticeable. The most constructive step that can be taken to heatsink the collector will be the use of InP as the collector material.

The grounding of internal nodes and/or devices in an IC is greatly improved by the close proximity of the ground plane. The via inductance that plagues most processes with backside processing and wafer thinning is nearly eliminated in the transferred-substrate process. This proximity of the ground plane does have one important drawback: the increase in parasitic capacitance. All interconnect lines have significant parasitic capacitance to ground. Transmission lines of 80-90 Ω characteristic impedance (inductors) are required to be only 4 μm wide, incurring significant skin loss at high frequencies. At 100 GHz, the skin depth in gold is only 0.25 μm . The 4 μm wide line has a resistance of approximately 500 Ω/cm at 100 GHz. A thicker microstrip dielectric would allow high impedance lines to be wider, reducing the skin losses.

This parasitic capacitance also manifests itself as a high pad capacitance. Input and output pads represent a substantial capacitance to ground. This is an impediment in broadband amplifier design, as the capacitance can only be tuned out over a narrow frequency range.

The question of mechanical support must be addressed in any scheme involving wafer thinning or substrate removal. Solder bonding to a GaAs carrier wafer provides mechanical support after the InP substrate is removed for the wafers discussed here. Although the technique improves with practice, this solder bonding is really not considered a robust process. Ongoing work at UCSB by James Guthrie seeks to remove the solder bonding process altogether [19]. In that work, after the ground plane is plated, a thick layer of copper will be plated to provide mechanical support and superior heat sinking. The substrate, then, will be metal. This copper plating technique may prove invaluable for dense and/or high power circuitry. In the solder process, the heating associated with the high current densities in the devices may cause local melting of the solder, leading to stresses in the material

that may break the interconnects or the devices themselves.

Chapter 4

Device Physics

It has been the aim in this work to improve upon every possible aspect of device performance, but with an emphasis on those things that would improve f_τ . The integrated circuits designed in the transferred-substrate HBT process span a variety of architectures: analog lumped amplifiers such as in this work, reactive-tuned amplifiers for W-band, travelling-wave amplifiers, flip-flops configured as static frequency dividers for high speed and/or low power. High speed Σ - Δ A-D converters and direct digital frequency synthesis circuitry are now in fabrication. Due to the limited supply of material, the baseline layer structure has been used for all of these applications. The intention is to avoid experimental layer structures where the innovation in the layer structure produces undesirable effects. This material might then be completely useless to everyone involved. Material with innovations in the layer structure is processed separately, so as not to jeopardize potential IC results. The innovations, if successful, are then incorporated into the baseline layer structure.

4.1 Device Modeling

The scattering parameters (S-parameters) extracted from device measurements are fit to a hybrid-pi model. A description of the process of modeling an HBT from S-parameter measurements is described. The hybrid-pi model is shown in fig. 4.1. The device current-gain cutoff frequency (f_τ) is given as

$$\frac{1}{2\pi f_\tau} = \tau_b + \tau_c + R_{ex}(C_{cb} + C_{la}) + \frac{kT}{qI_e}(C_{cb} + C_{je} + C_{la}), \quad (4.1)$$

while the power-gain cutoff frequency (f_{max}) is given by

$$f_{max} = \sqrt{\frac{f_\tau}{8\pi R_{bb}C_{cb,i}}} \quad (4.2)$$

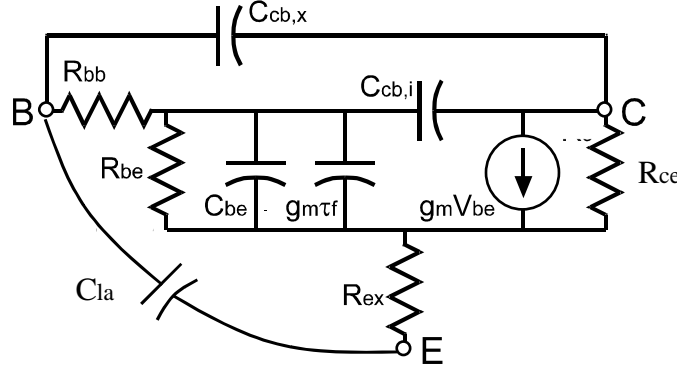


Figure 4.1: Hybrid-pi model of bipolar transistor, expressions for f_τ and f_{max}

In order to construct a hybrid-pi model of a device at a certain bias point, network analyzer measurements are used to determine the S-parameters of the device as a function of frequency. H_{21} and Mason's Unilateral Gain (U) are calculated as functions of frequency. The f_τ and f_{max} of the device are the unity gain frequencies determined by 20 dB/decade extrapolation on a log scale of h_{21} and U , respectively.

The admittance parameter Y_{21} is the inverse of the extrinsic transconductance of the device (measured at 1-2 GHz), given by $g_m/(1 + R_{ex})$. Plotting $1/\text{Re}(Y_{21})$ vs $1/I_c$ gives data that can be fit to a straight line. With $1/I_c$ plotted in units of 1/milliamps, and $1/\text{Re}(Y_{21})$ plotted in units of ohms, the slope of this line yields the product of the thermal voltage and the base-emitter ideality factor (n), in millivolts. The y-intercept of

this line is $1/\text{Re}(Y_{21})$ extrapolated to infinite current, which is R_{ex} . The dynamic emitter resistance r_e is given by nV_t/I_e . Then R_π is βr_e , with β given by the AC ratio of I_c to I_b .

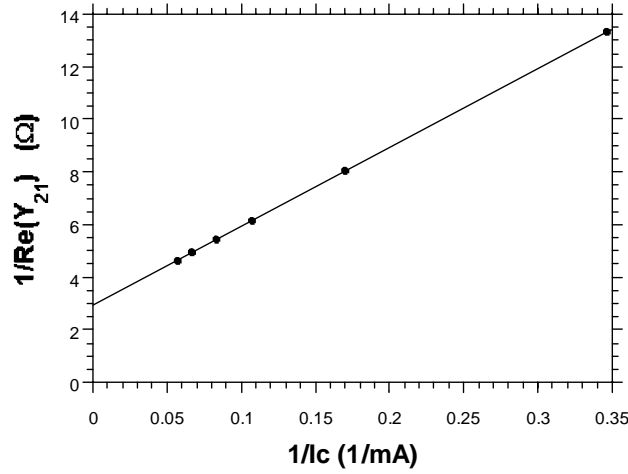


Figure 4.2: Fit of $1/\text{Re}(Y_{21})$ vs. inverse emitter current to determine R_{ex} and V_t/n .

The imaginary component of Y_{12} is the reverse transconductance, and is given by the factor $j\omega C_{cb}$. A plot of $\text{Im}(Y_{12})$ vs frequency will give data that again can be fit to a straight line. The slope of this line gives C_{cb} .

The measurement of the base-emitter depletion capacitance (C_{be}) as well as the sum of base and collector transit times ($\tau_b + \tau_c$) is as follows. The variation in f_τ as a function of collector current should first be measured, and a plot of $1/(2\pi \times f_\tau)$ vs $1/I_c$ generated. A straight line fit to the data will give a slope and a y-axis intercept. Inspection of the defining equation for f_τ as a function of the device parasitics will show that the slope of this line is the sum of C_{be} and C_{cb} multiplied by nV_t . $C_{cb} \times nV_t$ can be subtracted to isolate $C_{be} + C_{la}$, and C_{la} can be removed by hand calculations of $\epsilon A/d$ of the overlap of metal 1 and the base mesa through the polyimide. The intercept is the sum $\tau_b + \tau_c$ plus the sum $(R_{ex} + R_{coll})C_{cb} + R_{ex}C_{la}$. The RC time constant is then subtracted from the value of the y-intercept leaving the sum of base and collector transit times. The split between base and

collector transit time can be determined by processing devices of different base or collector thickness and then comparing the difference. Even this method is flawed, however, as it assumes a linear dependence of transit time on semiconductor layer thickness. This may not be the case if there is significant variation in the velocity of carriers as they pass through the relevant layer (such as if velocity overshoot is present).

Using a hybrid-pi model for the HBT, the base resistance can be estimated by extrapolation of S_{11} vs frequency on a Smith chart. At DC, S_{11} is entirely resistive and consists of the sum of R_{bb} , R_{ex} , and R_{π} . The trace rotates clockwise on the Smith chart as frequency increases due to the capacitive component (C_{be}) of the input port. At the highest measurement frequency, the curve will have stopped somewhere in the lower half plane of the Smith chart. If the trajectory of the S_{11} curve is extended up to the resistive center line, this would be the value of S_{11} at infinite frequency, where C_{be} shorts out the R_{π} . The resistance that is indicated by this extrapolation is then simply the sum of R_{bb} and R_{ex} . R_{ex} can then be subtracted to give R_{bb} .

The output resistance used in the hybrid-pi model can be estimated by the inverse slope of collector current as a function of V_{ce} in the device DC current-voltage characteristics. Output resistance can also be estimated by looking at the DC value of S_{22} on the Smith chart.

4.2 Base Resistance

What are desirable properties for the base layer? A low base resistance is foremost, in order to benefit f_{max} . Base resistance (R_{bb}) is the resistance a hole encounters as it travels from the base lead through the metal - semiconductor interface, into the semiconductor, and through the semiconductor until it reaches its final destination somewhere beneath the emitter stripe. R_{bb} can be separated into three components known as the contact resistance, the sheet resistance (ρ_{bs}), and the spreading resistance.

The contact resistance depends on the metal - semiconductor interface resistance as well as the sheet resistance of the semiconductor beneath the base metal. For the geometry shown in the fig. 4.3 (an emitter stripe surrounded by base contact metal), contact resistance is given by the formula $R_{contact} = \sqrt{\rho_{bc}\rho_{bs}}/2L$, where ρ_{bc} is the vertical base contact resistance in units of ohms-cm² [20]. In an Ohmic contact between metal and heavily

doped semiconductor conduction takes place predominantly through tunneling. Tunneling current as a function of voltage varies exponentially with the square root of the doping in the semiconductor [20]. A higher doping level therefore decreases the contact resistance by facilitating carrier tunneling and thereby reducing ρ_{bc} ; as well as reducing ρ_{bs} , the sheet resistance under the contact. As the contact resistance varies as the square root of ρ_{bs} , an increase in the thickness of semiconductor under the base contact also reduces contact resistance.

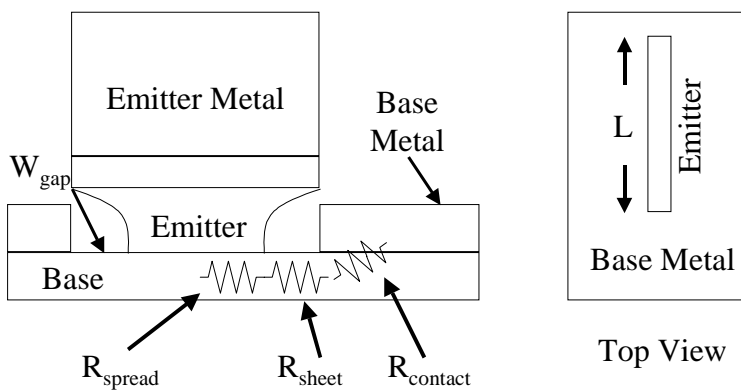


Figure 4.3: Components of base resistance.

The sheet resistance component is that resistance seen by the hole as it leaves the plane of the edge of the base metal until it arrives at the plane of the edge of the emitter stripe. For the geometry shown in the figure, this is given by the formula $R_{sheet} = \rho_{bs} W_{gap} / 2L$. Higher doping and thicker semiconductor reduce the sheet resistance. The parameter W_{gap} is the undercut of the emitter semiconductor in a self-aligned process. In a non-self-aligned process, W_{gap} would be the sum of the undercut and the lithographically defined spacing between the base and emitter metal.

The spreading resistance component is determined by the average length a hole must travel through semiconductor once underneath the active emitter to reach the position where it is needed to ensure charge neutrality. R_{spread} is given by the formula $R_{spread} = \rho_{bs} W_e / 12L$, with direct dependence on the sheet resistance as well as on the width of the active emitter.

R_{bb} decreases with increased semiconductor thickness and increased doping, which is expected. But this analysis has so far neglected the transport of minority carriers through the base. Minority carrier transport consid-

erations point to a thin, lightly doped region. The opposing requirements for good minority and majority carrier transport properties make the base layer the most obvious place where f_τ and f_{max} are balanced. In a self-aligned process with a small W_{gap} , the contact resistance is the dominant component of R_{bb} . In the oversimplified case where R_{bb} is determined completely by the contact resistance, halving the base thickness while doubling the base doping leads to a decrease in base resistance if it can be assumed that the majority carrier mobility is a slowly varying function of doping. (The integral of the concentration of base dopants over the thickness of the base is called the Gummel number. Halving the base thickness and doubling the doping maintains a constant Gummel number). If minority carrier mobility varies sufficiently slowly with doping, the minority carrier transit time also decreases. The inherent advantage of the HBT over the BJT is really contained here: that the base doping can be made arbitrarily high without fear of increased hole injection from the base into the emitter.

This favorable situation can end for many reasons, mostly related to practical growth and processing issues. Dopants generally begin at some point to either compensate or become electrically inactive, either case resulting in rapid degradation of minority and majority carrier mobility with increased doping. The sheet and spreading resistance do not benefit from layers in which the doping is increased as the layer becomes thinner (constant Gummel number). The only ways to improve these resistances are to increase the Gummel number in the base (leading to a degradation in transit time), to define a thinner emitter stripe to reduce spreading resistance or to reduced undercut to reduce sheet resistance at the expense of spreading resistance. From a processing standpoint, a very thin base is difficult to expose during the base etch without etching right through it.

4.3 Base Transit Time

Minority carrier transport through the base remains a limiting factor in the high frequency performance of bipolar transistors. This discussion will focus on the case of electron transport through a p-type base.

The layer structures used in the infancy of this project were such that the base-emitter heterojunction was linearly graded in order to eliminate any conduction band offset there. In this way, so-called hot electron injection was eliminated. After leaving the emitter, electrons traversed the base

via diffusion. The diffusivity of the electrons, if known, can be used to calculate the base transit time. Alternatively, diffusivity can be calculated from measured values of base transit time. In any event, there is a finite transit time taken by the electrons to cross the base.

The base layer in an HBT is thin in order to reduce the transit time. A thinner base layer is usually accompanied by heavier doping so as maintain constant base resistance. The hope here is that the electron diffusivity will decrease at a slow enough rate as doping increases such that the base transit time decreases in the thinner base.

Ultimately there is a limit to how thin the base can be, due to processing or uniformity limitations. A further decrease in base transit time then requires an increase in electron velocity across the base. In practice this has been achieved in one or both of two ways: base grading and hot electron injection.

Hot electron injection refers to the creation of a conduction band step at the base-emitter junction. The most common way of doing this is to simply not grade the p-n heterojunction. As electrons are injected from the wide bandgap n-type emitter into the narrow bandgap p-type base, they gain kinetic energy as the conduction band edge falls out from underneath them. If a significant fraction of electrons traverse a significant fraction of the base before scattering, base transit time is reduced. The conduction band step should be kept smaller than the Γ -L spacing of the base material in order to prevent electrons from scattering into the L valley during their transit through the base.

In fact, Monte Carlo simulations occasionally verified by experiment indicate that the collector transit time is a strong function of the base injection conditions. This makes sense: electrons exiting the base with significant kinetic energy should scatter into the L valley at a point in the collector nearer to the base. It is possible that a moderate conduction band step combined with a moderate amount of base grading would result in a minimum overall transit time. In this way, electrons enter the base with some energy. Those electrons that exit the base without scattering make it somewhat farther into the collector before entering the L valley than if the conduction band step was greater. The electrons that scatter during their transit through the base are accelerated out of the base by the base grading. The result is a slightly higher base transit time but lower overall transit time than when electrons entering the base are "very hot". Hot electron injection

has not been incorporated into transferred-substrate HBTs at this time.

The idea behind base grading is to create a driving force that accelerates electrons from emitter to collector. The base grading itself can be implemented in one or both of two ways. In compositional grading, the bandgap of the base material decreases monotonically from the emitter-base junction to the collector-base junction. Assuming a constant acceptor concentration throughout the base, a constant density of states N_v , and a linear grading profile, the band diagram in the base will look as shown in fig. 4.4(a). The result is a downward slope in the conduction band that tends to accelerate the electrons toward the collector. With a grade in doping, the bandgap of the base material is held constant, but the p-doping concentration decreases monotonically from the emitter-base junction to the collector-base junction. If the acceptor concentration in the base is such that Boltzmann statistics are valid (probably NOT the case in most HBTs, but assumed here for simplicity), an exponential doping grade can be found to give a constant electric field throughout the base. In this way the conduction (and valence) band slopes downward toward the collector resulting in an acceleration of electrons toward the collector as shown in the fig. 4.4(b).

The aim of the transferred-substrate HBT process is to realize a scalable f_{max} . A doping grade in the base is less useful than a bandgap grade for the following reason: a necessary condition for a doping grade in the base is that at least part of the base is doped below the highest possible doping in the material (otherwise there would be no doping grade). The result is increased base resistance when compared with the case where the doping is the maximum possible over the entire base width but the bandgap is graded.

A derivation of the expression for base transit time in the presence of a constant electric field in the base due to compositional bandgap grading begins with the drift-diffusion equation for minority electrons in the base.

$$J_n = qD_n \frac{dn}{dx} + q\mu_n \mathcal{E}n. \quad (4.3)$$

The electric field due to the compositional grade in the base is denoted \mathcal{E} , with $\mathcal{E} = -\Delta E_g/qW_b$. With this substitution and using the Einstein relation for the mobility ($\mu_n = D_n q/kT$), the drift-diffusion equation becomes:

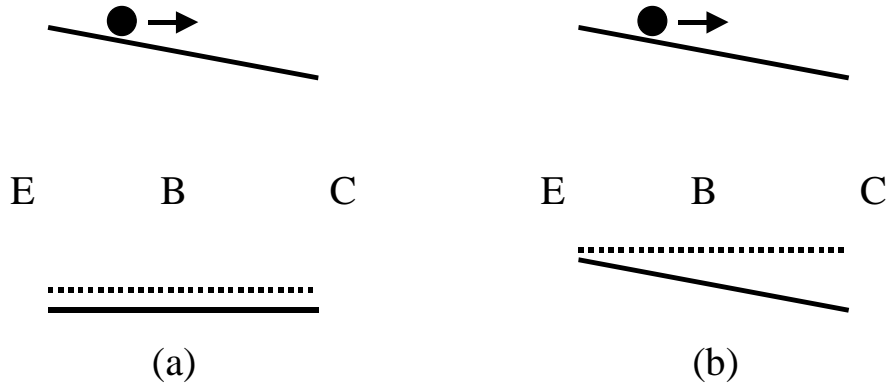


Figure 4.4: Base layer band diagrams for (a) linear compositional grading and (b) exponential doping grading

$$J_n = qD_n \left(\frac{dn}{dx} - \frac{\Delta E_g}{kT} \frac{n}{W_b} \right). \quad (4.4)$$

Another substitution to simplify is the introduction of a characteristic grading length $L_g = T_b kT / \Delta E_g$. The drift-diffusion equation can be rewritten in terms of L_g .

$$\frac{J_n}{qD_n} = \frac{dn}{dx} - \frac{n}{L_g}. \quad (4.5)$$

The solution $n(x)$ can be found by implementing standard techniques for solution of this first order differential equation.

$$n(x) = \frac{-L_g J}{qD_n} + ce^{\frac{x}{L_g}}. \quad (4.6)$$

The solution proceeds with application of the boundary condition at $x = 0$. The carriers are assumed to be moving at a velocity v_{sat} as $x = 0$ is the end of the base and the beginning of the collector drift region. With this assumption, the carrier concentration at $x = 0$ is $n(0) = -J/qv_{sat}$ (the current density J is negative). Substituting into the expression for $n(x)$ and determination of the integration constant c results in the particular solution for $n(x)$.

$$n(x) = \frac{-L_g J}{q D_n} + \left(\frac{L_g J}{q D_n} - \frac{J}{q v_{sat}} \right) e^{\frac{x}{L_g}}. \quad (4.7)$$

Now that the expression has been found for minority electron distribution, the base transit time τ_b can be evaluated.

$$\tau_b = \frac{\int_{-T_b}^0 n(x) dx}{-\frac{J_e}{q}}. \quad (4.8)$$

The base transit time is given by the number of minority carriers present in the base (stored charge) divided by the minority carrier current. Evaluating the integral and performing the division results in an expression for base transit time.

$$\tau_b = \frac{L_g T_b}{D_n} - \left(\frac{L_g^2}{D_n} - \frac{L_g}{v_{sat}} \right) (1 - e^{-\frac{T_b}{L_g}}). \quad (4.9)$$

This result is only for the case of the uniform electric field in the base. General solutions of transit time in the presence of an arbitrary field distribution are given elsewhere [21]. A plot of the predicted τ_b as a function of base width and for a few different values of base grading is shown in fig. 4.5 assuming $D_n = 40 \text{ cm}^2/\text{sec}$ and $v_{sat} = 4 \times 10^7 \text{ cm}/\text{sec}$.

4.4 Current Gain

Electrons are lost in the base due to recombination. Recombination can occur in the bulk, at a surface, or at the base contact metal. Bulk recombination is simply an electron meeting a hole in its transit through the base; the rate is therefore proportional to the base doping which is very high in HBTs. An electron diffusing through the base may travel laterally and reach the base contact metal where they are lost, or the surface may efficiently conduct electrons to the base contact metal with the same result. As the device dimensions are scaled, the surface to volume ratio of the base increases rapidly and with it electron recombination due to the exposed surfaces.

To ascertain the dominant electron - loss mechanism, data can be taken and fit to the following formula [22], where the fraction P/A is the periphery to area ratio of the emitter stripe:

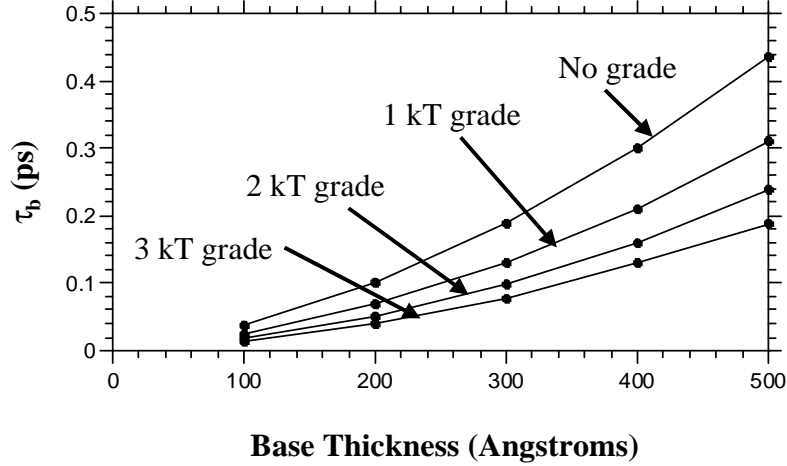


Figure 4.5: Base transit time as a function of compositional base grading strength.

$$\frac{1}{\beta} = \frac{1}{\beta_0} + \frac{J_{b,p} P}{J_c A}. \quad (4.10)$$

The current gain is expressed as the sum of two components: the bulk recombination term which has no dependence on the emitter geometry and a term involving the periphery to area ratio of the emitter stripe. A plot of this function allows separation of the two components. In fig. 4.6, this function is plotted for 1, 0.7, and 0.5 μm wide devices which have been undercut to what is estimated to be 0.9, 0.6, and 0.4 μm respectively. The length of these devices are all 12 μm . The results for two wafers are shown, one for a 400 Å base, the other for a 300 Å base, both with $\Delta E_g = 2kT$ base bandgap grading and identical emitter layer structures (baseline emitter layer structures).

The bulk current gain β_0 is 168 for the 300 Å base sample, and 87.3 for the 400 Å base sample. The periphery recombination current density is 4.56 mA/ μm for the 300 Å base sample and 4.91 mA/ μm for the 400 Å base sample. As expected, the thinner base shows a large improvement in

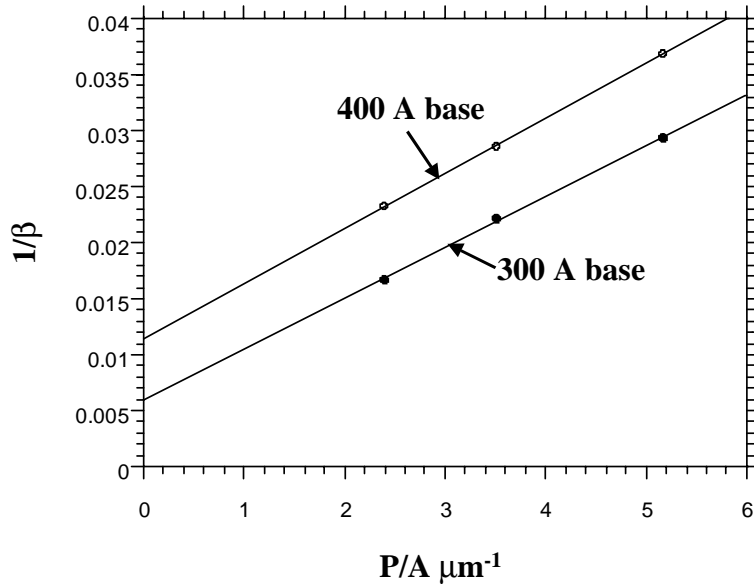


Figure 4.6: Inverse current gain as a function of periphery to area ratio.

the bulk recombination term, as well it shows a small improvement in the periphery recombination term. The current gain in the 300 Å base sample is dominated by the periphery component for the devices measured in this work, as even the widest ($0.9 \mu\text{m}$) emitter device on this wafer had a β of only 60, roughly a third of the bulk recombination term. The 400 Å base sample does not show a one sided dominance of current gain from either effect, the $0.9 \mu\text{m}$ device had a current gain of 43, half that of the bulk recombination term.

4.5 Collector Transit Time

The thin bases and highly scaled lithographic dimensions of modern bipolars have left the collector transit time (τ_c) as one of the dominant time constants limiting f_T . Simply thinning the lightly doped collector semiconductor is often not an acceptable solution for two reasons. For one, the

breakdown voltage of the collector-base junction decreases as the collector semiconductor is thinned. The minimum required breakdown voltage for an application sets a lower limit on the collector thickness. Secondly, the collector-base capacitance (C_{cb}) increases with decreased collector thickness, reducing f_{max} . One limitation of highly scaled transferred-substrate HBTs is the high f_{max} : f_{τ} ratio, an increase in f_{τ} at some expense to f_{max} is acceptable. The problem here is that the increased C_{cb} reduces f_{τ} as well, due to the $R_{ex}C_{cb}$ time constant. So as the collector thickness is reduced to hopefully improve f_{τ} , the increase in the $R_{ex}C_{cb}$ time constant subtracts from that improvement. Not surprisingly, large reductions in f_{max} result from attempts to improve f_{τ} by thinning the collector and with sufficiently high emitter resistance f_{τ} actually reduces as collector thickness decreases.

Examination of the impulse response of the base-collector junction will give important insight into what is meant by transit time in the collector. In particular, it will be most useful to look at the case of a p-i-n diode such as exists in most bipolar transistors today, with a p-type base, an intrinsic or lightly doped collector, and an n-type subcollector. The collector current is not the result of electrons entering the subcollector or collector contact, but rather is due to the displacement current resulting from electron motion through the collector depletion region. The collector transit time is therefore not equal to the time it takes for an electron to travel from base to collector (time-of-flight). If there is a constant electron velocity in the collector, then the collector transit time is one half of the time-of-flight. In III-V semiconductors that exhibit velocity overshoot negative differential mobility, constant velocity can be a poor approximation.

The current flowing in the collector lead will have the same profile as a function of time as the velocity of a traveling electron in the collector as a function of time. This is because of the Maxwell's relation for the displacement current given in equation 4.11.

$$J_c = \epsilon \frac{d\mathcal{E}}{dt}. \quad (4.11)$$

The time derivative of the electric field (at the collector - subcollector edge) is directly related to the velocity of the approaching electrons.

If the velocity of the electrons in the collector is known as a function of time, then, the derivation given by Bhattacharya [23] can be used to calculate the transit time, summarized briefly here for reference. The im-

pulse response of a sheet of electrons injected into the collector is denoted $i_c(t)$. The system transfer function will be calculated, then put into the form $1 - j\omega\tau_c$; the collector transit time is given by the term τ_c .

The system transfer function is given by the Fourier Transform of the impulse response, and is a function of the angular frequency ω .

$$I_c(j\omega) \equiv \int_0^{t_f} i_c(t) \exp(-j\omega t) dt. \quad (4.12)$$

The term t_f in the limit of the integral is the time - of - flight, or the time it takes for the electrons to pass through the drift region. The exponential in the integral can be expanded to first order as $\exp(-j\omega t) \simeq 1 - j\omega t$. Rewriting with this change results in

$$I_c(j\omega) \simeq \int_0^{t_f} i_c(t) dt - j\omega \int_0^{t_f} t i_c(t) dt, \quad (4.13)$$

and the prefactor containing the integral of the collector current is the impulse response with a value of unity. It can be removed without affecting the result,

$$I_c(j\omega) \simeq 1 - \frac{j\omega \int_0^{t_f} t i_c(t) dt}{\int_0^{t_f} i_c(t) dt}. \quad (4.14)$$

This is now in the form that was originally sought: $I_c(j\omega) \simeq 1 - j\omega\tau_c$. The collector transit time τ_c is given by:

$$\tau_c = \frac{\int_0^{t_f} t i_c(t) dt}{\int_0^{t_f} i_c(t) dt}. \quad (4.15)$$

This is the centroid or "center of mass" of the displacement current as a function of time, and therefore is the centroid of the time - dependent velocity function of electrons in the collector, and due to the equivalence of the two, the $i_c(t)$ can be replaced by $v_e(t)$ in the above equation for calculation of the collector transit time.

A charge control analysis can be used to derive the collector transit time if the electron velocity in the collector is expressed as a function of position. Transit time will be found by solving the charge control relation given in equation 4.16.

$$\tau_c \equiv \frac{dQ_b}{dJ_c}. \quad (4.16)$$

This expresses the transit time in the collector as the derivative of the modulated charge in the controlling electrode (the base) with respect to the controlled charge (collector current). It is necessary to impose the condition that $\delta V_{cb} = 0$. This ensures that the modulation in base charge is entirely due to the injected collector current. The mathematical expression for this condition is

$$\delta V_{cb} = \int_0^{T_c} \delta \mathcal{E}(x) dx = 0. \quad (4.17)$$

Note that this does NOT say that $V_{cb} = 0$, it just implies that the change in V_{cb} is zero as a result of the injected charge. The sheet of injected electrons will be denoted as $\sigma(x)$. Poisson's equation can be written $\Delta \cdot \mathcal{E} = \sigma(x)/\epsilon$ for the sheet of charge in the collector that is at position x . The charge $\sigma(x)$ must be compensated by the opposite charge appearing at the edge of the base adjacent to the collector, given by

$$\sigma_{comp} = -\sigma(x)\left(1 - \frac{x}{T_c}\right). \quad (4.18)$$

The electric field profile in the collector is shown schematically in fig. 4.7 with and without electrons in the collector; this field profile changes as the electrons travel in compliance with the condition imposed by equation 4.17.

The amount of compensating charge is equal to the amount of injected charge modulated by the position of the injected charge in the collector in order to satisfy the condition of equation 4.17. This is the modulated charge in the base electrode which is the numerator in equation 4.16 for the transit time. To find the collector transit time, the collector current must be found in terms of the relevant parameters.

$$J_c = qn(x)v(x) = \sigma(x)v(x). \quad (4.19)$$

Rearranging this equation the modulated charge density in the base $\sigma(x)$ can be expressed as $J_c/v(x)$. Now the cumulative modulated charge in the base upon completion of the injected electrons transit through the base can be written as:

$$\delta Q_{base} = \int_0^{T_c} \sigma(x)\left(1 - \frac{x}{T_c}\right) dx = \int_0^{T_c} \frac{J}{v(x)}\left(1 - \frac{x}{T_c}\right) dx. \quad (4.20)$$

The final step is division of this modulated charge by the collector current J_c to give the collector transit time.

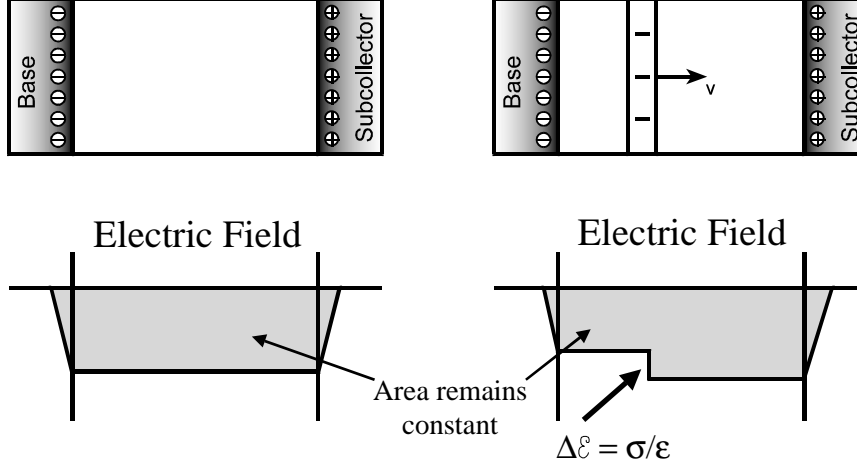


Figure 4.7: Collector field profile (a) with no injected current and (b) with a sheet charge σ of injected electrons

$$\tau_c = \int_0^{T_c} \frac{1}{v(x)} \left(1 - \frac{x}{T_c}\right) dx. \quad (4.21)$$

This result has been derived in a rather different manner by Laux and Lee [24]. For a given time-of-flight, minimum signal delay is the consequence of the highest electron velocity occurring in that region of the collector depletion region nearest the base. This can be seen by examination of either equation 4.15 or equation 4.21 for the collector transit time. Both of these expressions apply weighting functions to the electron velocity in the collector. The velocity near the base edge of the collector is most important, progressively decreasing as the electron travels in position (or time) through the collector.

This result for the signal delay has different implications for silicon bipolars than it does for III-V HBTs, due to the difference in the velocity-field characteristics of the semiconductors. Recall the monotonically increasing electron velocity in silicon as a function of electric field. The portion of the collector near the base should have the strongest electric field in order to minimize signal delay in a silicon bipolar.

The velocity of electrons as a function of applied electric field in III-V semiconductors increases, peaks at some value E_{opt} , and then decreases

asymptotically to some v_{sat} as the field is increased. An electric field of E_{opt} near the base would maximize electron velocity there. High-speed HBT designs employ thin collector semiconductor regions in order to minimize time-of-flight and hence signal delay. The built-in voltage of the base-collector junction alone is enough to create an electric field across the collector that is much greater than E_{opt} . Attempts to reduce the electric field near the base should lead to reduced signal delay and as a result, higher f_τ .

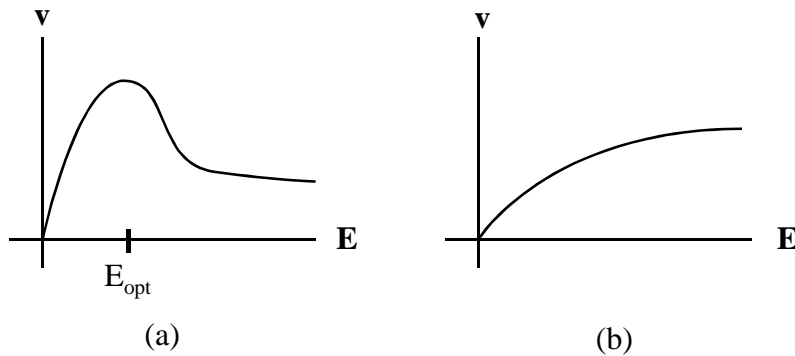


Figure 4.8: Velocity field curves for (a) III-V semiconductor and (b) silicon.

In III-V semiconductors, the reduction in average electron velocity at electric fields greater than E_{opt} is due to scattering into the satellite valleys. If an electron acquires sufficient kinetic energy, it may scatter via interaction with a phonon into a much higher effective-mass valley, where it will respond far more slowly to applied electric fields. In graded base structures, the electrons move under the influences of both diffusion gradients and an electric field. These electrons may exit the base with significant kinetic energy. For a given electric field in the collector these electrons will more readily scatter into a satellite valley, and collector signal delay will suffer. The result is that the reduced transit time from the graded base may be offset by an increased signal delay in the collector. F_τ depends on the total signal delay from base and collector; collector design should therefore accommodate injection conditions from the base.

Tuning the collector semiconductor to minimize signal delay in a circuit implementation is complicated by voltage and current excursions. The voltage across the collector-base junction is only constant in the case of purely small-signal analog circuits. Voltage swings and/or circuits with transistors

biased at different V_{cb} mean the collector can only be engineered to reduce signal delay in an average sense. Changes in current are reflected as changes in the space-charge densities in the collector, which in turn affect the electric field profile there. In many logic families, the switching transistor in a digital circuit switches between two modes. One is the high V_{cb} , low I_c condition, where electric field is strong throughout the depleted collector. A collector design that reduces the electric field near the base will reduce the signal delay. But when the transistor switches to the low V_{cb} , high I_c condition, reduced electric field near the base may impair the frequency performance in this mode of operation.

4.6 Emitter Resistance

The extrinsic emitter resistance R_{ex} is a significant impediment to further improvement in f_τ . The collector transit time (τ_c) contributes significantly to the forward delay. The simplest way to reduce τ_c is to thin the collector semiconductor. To maintain a high f_{max} , the thinning of the collector semiconductor is accompanied by a reduction in the area of the collector contact. Reduced collector dimensions require reduced emitter dimensions. And the reduced emitter dimensions drive up the resistance of the emitter ohmic contact, so while the collector transit time is reduced, the $R_{ex}C_{cb}$ time constant increases. The result of thinning the collector semiconductor is a nearly constant f_τ and a reduction in f_{max} . The result of thinning the collector semiconductor and scaling the lateral dimensions of the device is a reduced f_τ . An examination of the equation for forward delay (equation 4.22) will make this more clear, again with inclusion of the term C_{la} which indicated parasitic capacitance resulting from the physical layout.

$$\frac{1}{2\pi f_\tau} = \tau_b + \tau_c + R_{ex}(C_{cb} + C_{la}) + \frac{kT}{qI_e}(C_{cb} + C_{je} + C_{la}). \quad (4.22)$$

To simplify the discussion, let it be assumed that a 2:1 improvement in f_τ is sought. One way to do this is to reduce all terms in the forward delay by a factor of two. With the assumption of $\tau_c = T_c/2v_s$ (T_c is collector thickness, v_s is the saturation velocity), a 2:1 reduction in τ_c requires a 2:1 reduction in T_c . The following equation expresses $R_{ex}C_{cb}$ in terms of the relevant parameters:

$$R_{ex}C_{cb} \propto \frac{\rho_e}{A_e} \frac{A_c}{T_c}, \quad (4.23)$$

where ρ_e is the extrinsic emitter resistance per unit area. With the 2:1 increase in T_c , a 2:1 improvement in ρ_e is needed in order to maintain a constant $R_{ex}C_{cb}$ and a 4:1 improvement is needed in order to reduce $R_{ex}C_{cb}$ by two. Hence the assertion that R_{ex} limits f_τ scaling of transferred-substrate HBTs at this time.

The R_{ex} of a $1 \times 12\mu m$ emitter stripe is 3-4 Ω in the baseline layer structure. A useful starting point is to determine where this emitter resistance is coming from. Electrons must travel through the pad metal into and through the emitter contact metallization, across the metal-semiconductor interface, then vertically through the cap layer, a heterojunction, and finally through the widegap emitter itself. Each one of these layers or interfaces contributes to the resistance. The metal-semiconductor interface resistance can be measured with transmission line model (TLM) measurements on emitter material. The mobility of electrons in the semiconductor is known from the Hall measurements on the doping calibration samples, so the resistance incurred in the semiconductor can be calculated with knowledge of the device dimensions. Without test structures that force current flow vertically through the stack of emitter metallization and semiconductor, the resistance contributed by the emitter metal itself and by the InGaAs/InAlAs heterojunction cannot be estimated.

Table 4.1 charts the relevant semiconductor parameters and the contribution to emitter resistance from the InGaAs cap layer, the heavily doped InAlAs layer, and the InAlAs emitter itself. The resistivity (ρ) has been calculated from the measured electron mobility (μ_n) and the electron concentration (N_d) measured in doping calibration samples. The device dimensions assumed are a $1 \times 12\mu m$ lithographically defined emitter with the InAlAs layers undercut to $0.7 \times 12\mu m$. The InGaAs layer is supposed to be unaffected by the undercutting etch.

Table 4.1 indicates that about 25% of the R_{ex} is contributed by the semiconductor layers for the $1 \times 12\mu m$ geometry device, a contribution of 0.85 Ω out of 3.5 Ω . The same calculations have been done in table 4.2, but for an $0.5 \times 12\mu m$ lithographically defined emitter with the InAlAs layers undercut to $0.2 \times 12\mu m$.

In table 4.2 the contribution of the emitter resistance from the semicon-

Table 4.1: Contributions to bulk semiconductor resistance from emitter semiconductor layers for a $1 \times 12\mu m$ lithographically defined emitter undercut by $0.3\mu m$.

Layer	Doping ($/cm^3$)	ρ ($\Omega \cdot cm$)	Length (\AA)	Area (μm^2)	Resistance (Ω)
InGaAs	1.2×10^{19}	2.7×10^{-4}	1000	12	0.022
InAlAs	8.5×10^{18}	1.8×10^{-3}	900	8.4	0.19
InAlAs	7.8×10^{17}	1.1×10^{-2}	500	8.4	0.64

Table 4.2: Contributions to bulk semiconductor resistance from emitter semiconductor layers for a $0.5 \times 12\mu m$ lithographically defined emitter undercut by $0.3\mu m$.

Layer	Doping ($/cm^3$)	ρ ($\Omega \cdot cm$)	Length (\AA)	Area (μm^2)	Resistance (Ω)
InGaAs	1.2×10^{19}	2.7×10^{-4}	1000	6	0.044
InAlAs	8.5×10^{18}	1.8×10^{-3}	900	2.4	0.67
InAlAs	7.8×10^{17}	1.1×10^{-2}	500	2.4	2.2

ductor layers is estimated to be about 25% of the total R_{ex} measured for a device of $0.5 \times 12\mu m$ geometry.

The resistance incurred by the metal-semiconductor contact can be estimated by the result of TLM measurements on emitter material. Contact resistivity as measured by TLM is determined to be $\sim 2 \times 10^{-7} \Omega \cdot cm^2$. For a $1 \times 12\mu m$ lithographically defined emitter stripe, this contact resistivity corresponds to a resistance of 1.7Ω . This is significant, but does not make up the remainder of the resistance not accounted for by the semiconductor bulk. To get the measured $R_{ex} = 3.5\Omega$ an extra resistance of 1Ω must be coming from somewhere not accounted for. For the $0.5 \times 12\mu m$ lithographically defined emitter stripe, this contact resistivity corresponds to a resistance of 3.4Ω . Herein lies the problem. The device with the $0.5 \times 12\mu m$ lithographically defined emitter stripe has an R_{ex} of 11.3Ω , so there is 6.3Ω not accounted for in this device. So the $R_{ex}C_{cb}$ term is doing worse than not scaling as the device dimensions are reduced; it is actually an "anti-scaling" term that becomes larger as the device is scaled.

It has been determined that a lithographically defined $0.5\mu m$ emitter stripewidth of a certain length has *more* than twice the R_{ex} of a lithographically defined $1\mu m$ wide emitter. The data to this effect is plotted in fig. 4.9. This is data extracted from devices of 0.5, 0.7, and $1.0 \mu m$ emitter stripewidth processed on a typical wafer.

The x-intercept of the plot of $1/R_{ex}$ is the effective undercut. That is, this is the lithographically defined emitter stripewidth at which the R_{ex} goes to infinity. If R_{ex} goes to infinity, then the effective area of contact must go to zero, and so the x-intercept is a measure of the undercut of the emitter contact. But this is not just the undercut of the InAlAs. The calculations of the contributions to emitter resistance have shown that the InAlAs emitter layers contribute somewhat less than 25% of the total R_{ex} . If the plot of fig. 4.9 were of $1/R_{InAlAs}$ vs. W_e then the result of fig. 4.9 would be expected. The plot of fig. 4.9 is showing that R_{ex} scales as if the bulk semiconductor AND the contact interface were being undercut; the excellent linearity of this plot indicates that the bulk and the contact interface are being undercut to the same extent. Alternatively, the metal itself may be attacked instead of the contact interface. The metal would have to be attacked severely to account for the effect of fig. 4.9 as the titanium and platinum have a moderate conductivity but are very thin and the gold is thick but has very high conductivity compared to the semiconductor.

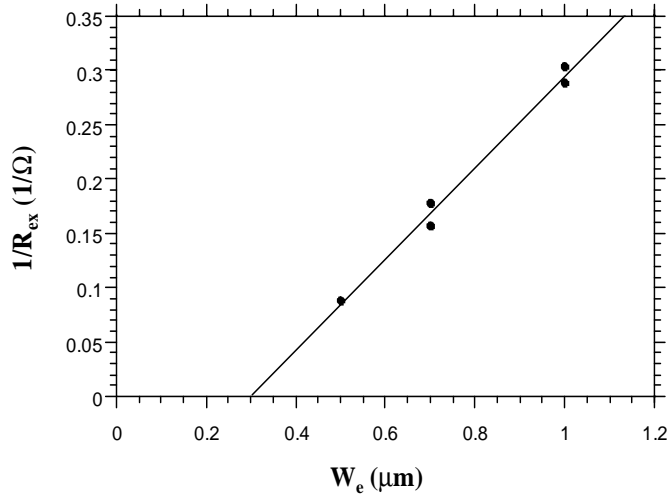


Figure 4.9: $1/R_{ex}$ as a function of emitter width W_e to determine effective undercut.

A study of contact resistance to various cap layers has been done in an attempt to reduce the specific contact resistance component of R_{ex} . TLM measurements were done on four different emitter structures, shown as layer structures A-D in fig. 4.10.

The metallization used for all samples was Ti/Pt/Au of thicknesses 200/500/8000 Å, and all TLM measurements were performed after a 300°C, 1 minute rapid thermal anneal (RTA). The results displayed in table 4.3 were as expected, that increased doping or strained layers with extra In content reduce the contact resistance. However, incorporation of the clear winner of the contact resistance contest, layer structure B, have NOT shown a noticeable decrease in R_{ex} when this emitter cap layer structure is applied to transistors. Some improvement in the R_{ex} was expected for the 80 % reduction in contact resistance; no significant improvement was observed. Unfortunately, changing the emitter cap semiconductor does not address the anomalous R_{ex} problem.

The resistance presented by the heterojunction between the InGaAs cap layer and the InAlAs is not believed to be significant, as the degenerate dop-

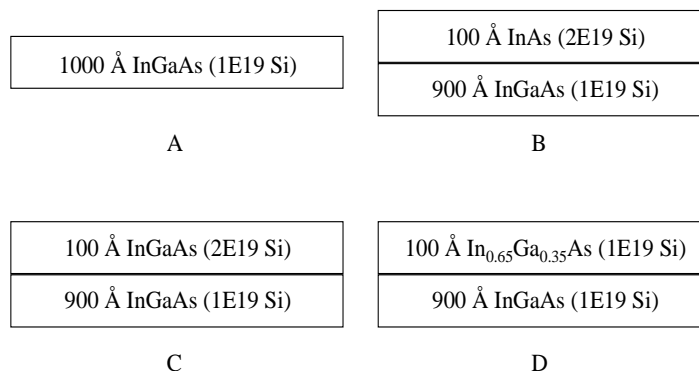


Figure 4.10: Emitter cap layers for contact resistance study

Table 4.3: Sample structure and corresponding specific contact resistance.

Structure	Contact Resistance ($\times 10^{-7} \Omega - cm^2$)
A	2.0
B	0.4
C	1.2
D	1.3

ing density on either side ensures that the peak of the conduction band that occurs at the heterojunction is well below the Fermi level. As a check on this, that heterojunction has long since been graded with a graded superlattice of varying thickness, and no effect on R_{ex} has been observed. The resistive drop from the interconnect metal is negligible, it has been verified by four point probe measurement to be a fraction of an ohm across a distance of some hundreds of microns.

So why the anomalous behavior of R_{ex} ? During the process, the emitter metal and metal-semiconductor junction are exposed to the harsh chemistry of the base contact etch. It is possible that these etches may cause some lateral etching of the metal-semiconductor interface or even one or more of the metal layers themselves. A TLM measurement would not be a good indicator of this, as the area of the contact pad is so large for ease of probing and measurement (about $50 \times 50 \mu m$) that a few thousand Angstroms of

undercut would be insignificant. An emitter stripe $1\ \mu\text{m}$ wide, however, would be greatly affected if the stack of contact metal or the interface were being undercut substantially. During the development of the base-etch process by Michelle Lee at UCSB, scanning electron microscope (SEM) images were taken of the base-emitter junction after the base contact etch. The InGaAs cap layer is not undercut, as expected, and the InAlAs layer is undercut, also as expected. The possibility that the interface or one of the thin metal contact layers is being attacked by the base contact etch cannot be ruled out, however. It is not possible with the available equipment to resolve the $200\ \text{\AA}$ thick titanium contact layer, for example. A phenomenon such as the attack of the metal-semiconductor interface could be responsible for the nonlinear increase of emitter resistance as the emitter stripewidth is scaled.

The reduction of R_{ex} is an important piece of future work. Traditional TLMs fabricated on emitter material have proven to be an ineffective tool to estimate the device R_{ex} . TLM patterns are under development that will force current to flow vertically through the emitter epitaxy and then laterally. These new test structures will allow quick and precise characterization of the emitter resistance and, more importantly, allow quicker evaluation of proposed techniques to reduce R_{ex} . The mask sets used in the latter stages of this work provide for an array of device sizes to be simultaneously processed, and should be used for device characterization as they quickly reveal departures from the scaling laws.

4.7 Kirk Effect

The Kirk effect represents a severe performance obstacle in many HBT technologies. The collector region in most HBTs is lightly doped n-type, that is, there is positive space charge in the collector. If a sufficient number of electrons is injected into the collector, the negative charge of these traveling electrons will cancel the positive space charge, the addition of more electrons will overwhelm it resulting in a net negative charge in the collector. The concavity of the collector band diagram then switches, with serious implications for device performance.

So with injected electron current and the assumption of constant electron velocity v_{sat} , the electric field profile in the collector evolves with increasing current density as shown in fig. 4.11. As the space charge in the

collector begins to be neutralized by the electrons, the electric field near the base reduces, while the electric field near the subcollector increases. A threshold current density can be computed where the electric field at the base-collector junction reaches zero, hereafter referred to as the Kirk threshold. Integration of Poisson's equation reveals that this current density is given by the following:

$$J_c = v_{sat} \left[\left(\frac{2\epsilon}{T_c^2} \right) (V_{cb} + \phi_{bi}) + qN_d \right]. \quad (4.24)$$

Increasing the doping in the collector (N_d) or reducing its thickness (T_c) can increase the Kirk threshold.

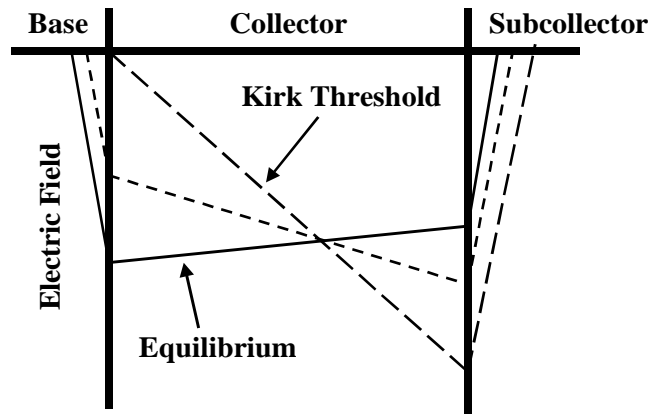


Figure 4.11: Increased current changes electric-field profile in collector.

The band diagram of a p-i-n diode is shown in fig. 4.12, representing the base-collector junction in a bipolar transistor in equilibrium and with the Kirk Effect threshold current density flowing. The depleted donors in the collector drift region give rise to a concave-up shape to the band there at equilibrium, but are overwhelmed by the charge of the traveling electrons when sufficient current is flowing.

Device performance will begin to suffer at this current density. Beyond this current density, the distance over which the electric field goes to zero increases at the expense of the thickness of the depleted collector. The thickness of the base increases, the thickness of the collector decreases.

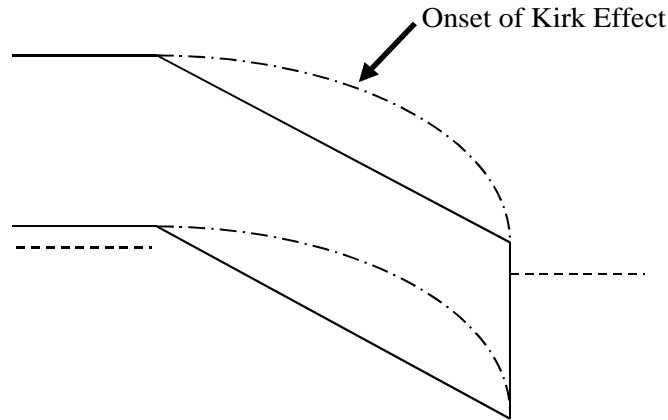


Figure 4.12: Band diagram of collector-base junction at equilibrium and with Kirk Effect threshold current density flowing.

The results of this are increased base transit time, reduced current gain, slightly reduced collector transit time, and increased collector-base capacitance. These results lead to a collapse in both f_τ and f_{max} .

And this is the case for silicon bipolar transistors. These transistors typically have base doping in the $10^{16}/\text{cm}^3$ to $10^{17}/\text{cm}^3$ range. The velocity of electrons in the collector is lower than for III-V transistors, especially at the lower fields due to the lower mobility. The lower electron velocity results in a higher electron density for a given current as defined by $\rho(x) = J/v(x)$. At the current density given in equation 4.24, the electric field will have just reached zero at the edge of the base, and the concentration of holes in the collector near the base will approach the concentration of acceptors in the base. This results in a lower current gain as the base effectively becomes wider; for this reason the Kirk Effect is sometimes called "base pushout".

It has recently been argued that Kirk effect is much more subtle in III-V based HBTs than it is in silicon bipolars [25]. The degenerate p-doping in the base causes the band diagram in the collector to look somewhat different than is shown in fig. 4.12. The band diagram in the III-V HBT collector takes on the shape shown in fig. 4.13. The difference is that the density of holes in the collector between the true base and the current induced base (region Δx_2) cannot be neglected. The high electric field present at the

base - collector junction is due to the presence of high concentrations of holes at distances of tens of Angstroms past the edge of the base. These holes cause the band diagram to take on the concave - up curvature shown in region Δx_1 in fig. 4.13. Electrons will enter the collector with a high kinetic energy due to this curvature. The energy drop ΔE_1 is given by

$$\Delta E_1 = kT \left(\ln\left(\frac{N_a}{p}\right) + \frac{1}{\sqrt{8}} \left(\frac{N_a}{p}\right) + \dots \right), \quad (4.25)$$

and the Joyce - Dixon expansion must be used as the base is degenerately doped and Boltzmann statistics are no longer valid [26].

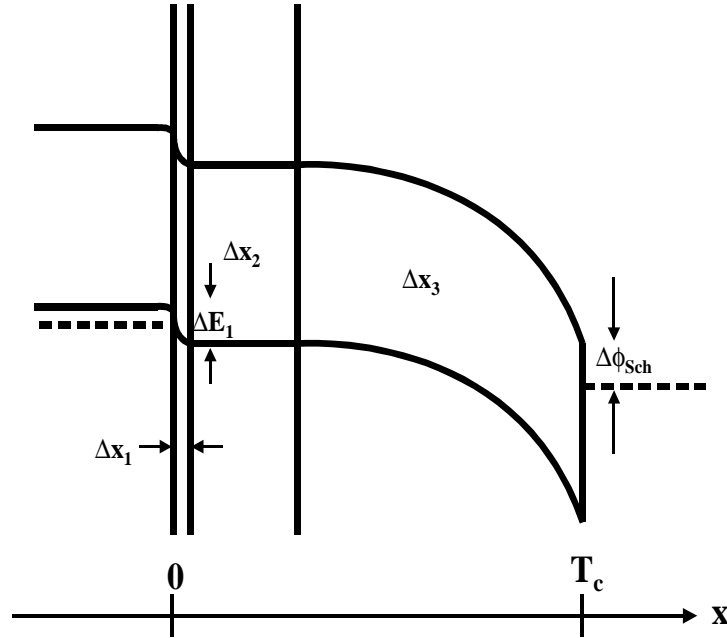


Figure 4.13: Collector - base band diagram with current density above Kirk threshold in a III-V HBT.

Region 2 of thickness Δx_2 is the region over which the current induced base forms. The definition of this current induced base is that region in which the hole density p is given by $p \simeq J/qv_e(x) - N_d$. With this expression, equation 4.25 becomes:

$$\Delta E_1 = kT \left(\ln\left(\frac{N_a}{\frac{J}{qv_e(x)} - N_d}\right) + \frac{1}{\sqrt{8}} \left(\frac{N_a}{\frac{J}{qv_e(x)} - N_d}\right) + \dots \right). \quad (4.26)$$

The band becomes very flat here, but on average the electrons have entered region 2 with significant kinetic energy imparted to them in region 1. Also, the low field mobility of most III-V semiconductors is very high, especially for $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$.

Most of the voltage dropped across the collector in the presence of a current - induced base is in region 3. In this region the hole density has fallen off to negligible values, and the traveling electrons are the dominant charge in this region. The density of electrons is given by $\rho(x) = J_c/v_{sat}$, where the saturation velocity is assumed since the electric field is high. Poisson's equation in region 3 can be solved for the voltage drop as

$$\Delta\phi = (qN_d - \frac{J}{\epsilon v_{sat}}) \frac{\Delta x_3^2}{2}; \quad (4.27)$$

the quantity $\Delta\phi$ also is equal to the following as per fig. 4.13:

$$\Delta\phi = \phi(T_c) - \phi(0) = -\left(\frac{E_{gap} + \Delta E_a - \Delta E_1}{q} - \phi_{Sch} + V_{cb}\right). \quad (4.28)$$

So the distance Δx_3 can be found by substituting equation 4.28 into equation 4.27 and solving for Δx_3 ,

$$\Delta x_3 = \sqrt{-\left(\frac{E_{gap} + \Delta E_a - \Delta E_1}{q} - \phi_{Sch} + V_{cb}\right)\left(qN_d - \frac{2\epsilon v_{sat}}{J_c}\right)}. \quad (4.29)$$

The collector thickness for the case of fig. 4.13 is given by $T_c = \Delta x_1 + \Delta x_2 + \Delta x_3$, so the thickness of the current induced base is given by $\Delta x_2 \simeq T_c - \Delta x_3$, since Δx_1 is very small.

Operation of the device at current densities resulting in the band diagram of fig. 4.13 may actually lead to enhanced f_τ . The reduction in electric field in region 2 leads to a higher electron velocity in that region of the collector. Recall from the introduction to this chapter that a regression plot of $1/2\pi f_\tau$ vs $1/I_c$ is used in order to estimate the forward transit time in an HBT. The result of the modulation of the band diagram in the collector due to the high electron density would appear as a departure in linearity of this plot at high current densities. The increased electron velocity due to the reduction in field should lead to a downward bending shape to the regression plot as indicated in fig. 4.14. A particular design of

GaAs/AlGaAs HBTs has experimentally shown an increased electron velocity with a reduction in the electric field in the collector. The concept has been titled "Ballistic Collector HBTs", and utilizes a p-type delta dope near the subcollector side of the collector drift region [27]. The result is a reduced electric field throughout most of the collector, and the reduction in τ_c is measured, although the reduction in the electric field near the base hastens the flattening of the conduction band near the base.

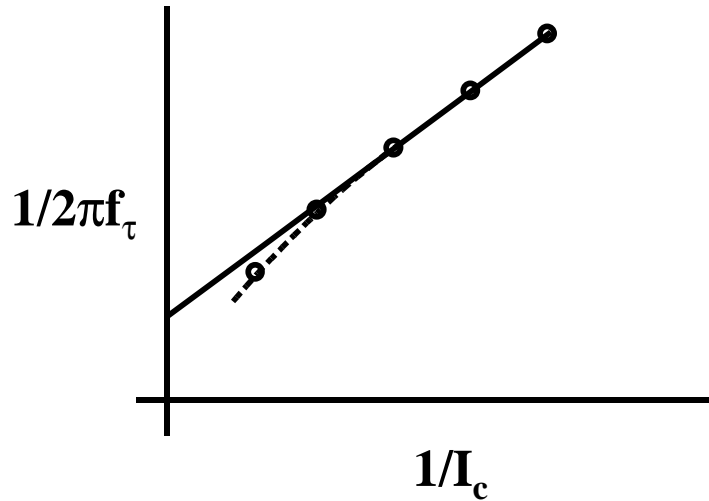


Figure 4.14: Postulated increase in electron velocity at high currents.

It has also been postulated that the decrease in current gain that is supposed to occur due to the onset of Kirk effect will not appear [25]. The number of electrons injected into the collector can never be enough to compensate the high density of holes near $x = 0$, hence the number of holes in region 1 of the collector will remain nearly unaffected. The concentration of electrons would not approach the high concentration of holes until a substantial distance away from the base, out in region 2. At this point in the collector, the concentration of both electrons and holes is more likely to be in the $10^{16}/\text{cm}^3$ range (given by $n(x) = J_c/qv(x)$), electron concentration will certainly never approach the $\sim 10^{19}/\text{cm}^3$ concentration that is found in the base.

In the work done by Uddalak Bhattacharya at UCSB, transferred - substrate HBT performance was hindered by the onset of Kirk Effect. Current

densities of $1 \times 10^5 \text{ A/cm}^2$ were rarely exceeded due to the rapid degradation of f_τ and f_{max} at higher current densities. A solution has been implemented as a short section of strong n-type doping in the collector near to the base. The baseline layer structure has a 50 \AA thick layer of InGaAs doped at $1 \times 10^{18} / \text{cm}^3$ at a distance of 400 \AA from the base. This n-type doping in the collector is to be contrasted with the p-type doping implemented in the collector of the "ballistic collection transistor" mentioned before. In this case, the goal is to improve the useful range of current densities over which the device can be biased, not necessarily to improve the transit time. Fig. 4.15 shows the effect this pulse dope is intended to have against the back-drop of the simplified collector band diagram for the case of equilibrium and under moderate current injection.

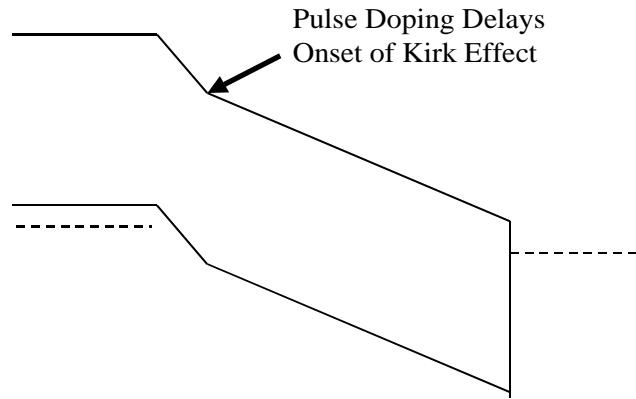


Figure 4.15: Pulse dope in collector used to delay onset of Kirk Effect

Chapter 5

Device Results

Six layer structures have been fabricated that incorporate many of the variations in the layer structures that have been described before. The layer structures are labeled A through F for reference in this chapter. A summary of the layer structures follows.

Fig. 5.1 is the baseline layer structure, designated layer structure A. The layers are numbered so that the other layer structures can be described, as they are different from it. Layer structure B has no pulse doping; instead there is a uniform 3000 Å thick lightly doped collector region. A thinner collector is used in layer structure C, 2000 Å, and there is also no pulse doping in this one. This 2000 Å collector and no pulse doping is retained for layer structure D, but with a 300 Å base. The grading in this base still constitutes a 2kT energy drop, so the effective electric field is stronger. Layer structures E and F are substantially different. Layer structure E has a 2000 Å collector and a somewhat weaker pulse doping than the baseline. The base thickness is 500 Å the first 400 Å of which are graded with a 2kT energy drop, the last 100 Å is doped more lightly at $3 \times 10^{19}/\text{cm}^3$. The n-type portion of the base-emitter grade is doped more lightly than the baseline layer structure, at $5 \times 10^{17} / \text{cm}^3$ instead of the baseline $8 \times 10^{17} / \text{cm}^3$. And layer structure F has the same collector structure as layer structure E, but with the same base and emitter structure as the baseline. The remaining difference in layer structure F is the 100 Å thick InAs cap. These layer structures are summarized in table 5.1.

The following discussion is based on the highest f_τ devices found on the samples, which except for the 0.7 μm wide emitter device of sample

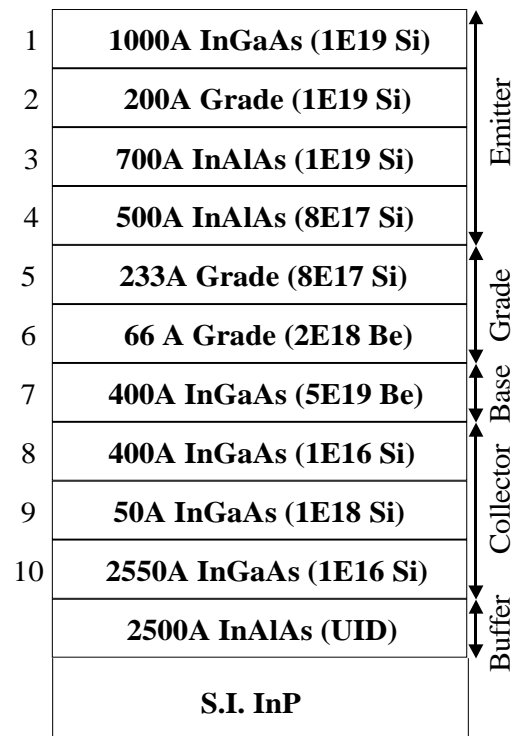


Figure 5.1: Baseline layer structure for transferred-substrate HBTs

Table 5.1: Quick reference for device characterization layer structures.

Sample	Layer(s) Different	Explanation
A	N/A	baseline
B	8,9,10	8,9 removed. 10 3000Å.
C	8,9,10	8,9 removed. 10 2000Å.
D	7,8,9,10	7 300Å. 8,9 removed. 10 2000Å.
E	6,7,8,9,10	6 Si doping 4E17 /cm ³ . 7 500Å. 8 250Å. 9 Si doping 7E17 /cm ³ . 10 1700Å.
F	1,8,9,10	1 100Å InAs, 900Å InGaAs. 8 250Å. 9 Si doping 7E17 /cm ³ . 10 1700Å.

D, happened to be the 1 μm emitter stripe width devices. This is no coincidence - the devices with emitter stripewidth less than $1\mu\text{m}$ had the much higher R_{ex} than expected for their size, this caused a reduction in f_τ .

5.1 Layout Parasitics

Some differences in the mask layout for these samples should be noted. Fig. 5.2 shows three mask layouts used for the 1 μm emitter width devices. There are differences in the area of overlap of both metal 1 and metal 2 interconnect with the base contact metal that do not scale with the difference in device area. The different overlaps result in differing values of parasitic capacitance from base to emitter. As this capacitance lies outside of the intrinsic transistor, the time constant introduced by this extra capacitance is $(R_{ex} + r_e)C_{la}$, where C_{la} stands for the capacitance introduced by the base metal layout. The value of C_{la} will be pointed out when relevant. The modified f_τ expression becomes

$$\frac{1}{2\pi f_\tau} = \tau_b + \tau_c + R_{ex}(C_{cb} + C_{la}) + \frac{kT}{qI_e}(C_{cb} + C_{je} + C_{la}). \quad (5.1)$$

The last term must be accounted for when extracting the transit time and base-emitter capacitance of the device, as it contributes to the forward delay.

There are three contributions to emitter-base capacitance from the layout parasitics. The first to be discussed is the overlap between the metal 1 (M1) interconnect and the base metal. The area of overlap here is simply the product of the width of the M1 interconnect and the width of the base metal minus the width of the emitter finger. In fig. 5.2(a), the width of the base mesa is $6.7 \mu\text{m}$, the M1 interconnect is $6 \mu\text{m}$ wide, and the width of the emitter stripe is $1 \mu\text{m}$. The area of overlap is therefore $6 \cdot (6.7 - 1) \mu\text{m}^2$, or $34.2 \mu\text{m}^2$. Between the M1 and the base metal lies the polyimide, the parasitic capacitance is $0.06 \text{ fF}/\mu\text{m}^2$. This $34.2 \mu\text{m}^2$ overlap results in an extra capacitance of 2.2 fF . There is also overlap between metal 2 (M2) and the base mesa. But between the M2 and the mesa lies not just polyimide, but also silicon nitride. The parasitic capacitance per unit area is the series combination of that from polyimide ($0.06 \text{ fF}/\mu\text{m}^2$) and from silicon nitride ($0.128 \text{ fF}/\mu\text{m}^2$), or $0.041 \text{ fF}/\mu\text{m}^2$. The relevant area here is the area of overlap between M2 and the base mesa, minus the area of overlap between M1 and the base mesa and also minus the area of the emitter finger that is not covered by M1. For the case of fig. 5.2(a), this area is $54 \mu\text{m}^2$, which gives another 2.2 fF .

Now, in fig. 5.2(a), the length of the emitter stripe is $8 \mu\text{m}$. The consequence of the emitter finger being longer than the width of the M1 interconnect is a small series resistance from the ends of the emitter finger to the M1 interconnect. This was assumed to be negligible due to the thick emitter metal, but a seemingly negligible series resistance may affect the fraction of V_{be} available at the ends of the emitter finger when compared with the center. The result would be uneven current distribution along the length of the device. Also, there will be misalignment of the M1 interconnect to the emitter finger; the resulting asymmetry will cause the resistance to be higher at one end of the emitter finger than the other.

Fig. 5.2(b) contains some obvious differences from the layout of fig. 5.2(a). Here the width of the M1 has been made much wider in order to eliminate the aforementioned series resistance. In order to compensate for the increase in overlap, the width of the base mesa has been reduced from 6.7 to $3.7 \mu\text{m}$. The capacitance due to overlap of M1 with base metal is 2.5 fF , and that due to M2 overlap with base metal is 0.6 fF . There is an additional contribution to the parasitic capacitance due to nothing other than a mistake in the layout. It is the extension of the M2 past the edge of the polyimide etch mask. In fig. 5.2(a), which has been done correctly, the

M2 is completely contained within the Poly Etch mask. This means that the parasitic capacitance from M2 to the base mesa is through the series combination of polyimide and silicon nitride. In figs. 5.2(b) and (c), the Poly Etch mask is not large enough, which means that polyimide has been etched away below a stripe of the M2 which is $1\ \mu\text{m}$ wide. This means that the intervening dielectric between M2 and the base metal in this region is only the silicon nitride. At this point, the base mesa is wider, back to $6.7\ \mu\text{m}$, and so the parasitic capacitance from this narrow stripe of M2 is $0.86\ \text{fF}$. In fig. 5.2(c), the base mesa is also $6.7\ \mu\text{m}$ wide, so the extra capacitance is the same.

The mask layout shown in fig. 5.2(c) yields the largest excess capacitance of $7.0\ \text{fF}$, and unfortunately was used for samples B, D, and F. The base mesa was narrowed to reduce this capacitance as per fig. 5.2(b) for use in sample C. The layout of fig. 5.2(a) was used for samples A and E.

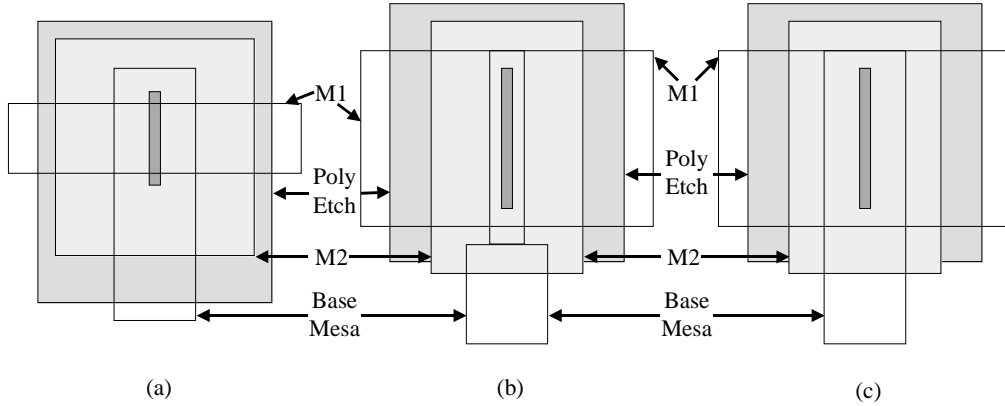


Figure 5.2: Detail of layouts showing base mesa and interconnects

Careful attention must be paid to these layout parasitics when the device is scaled. The parasitic capacitance from the layout, C_{la} is extrinsic to the device. This means that in calculation of the RC time constant contributed to the forward delay, the relevant product is $(R_{ex} + r_e)C_{la}$. The R_{ex} is many times higher than the r_e at higher current densities, and especially so for those scaled devices with the anomalously high R_{ex} . So even though C_{la} is a small fraction of C_{je} (typically about 5%), this percentage is multiplied by a factor of 5 or more in its contribution to forward delay when compared to C_{je} .

The dimensions shown in fig. 5.2 and discussed above are for 1 μm wide devices. If the device is scaled down the area of overlap must also be scaled down if the capacitance per unit area of the device is not to increase. This can be done by reducing the width of the base mesa. One reason for the reduced f_τ generally observed in the submicron devices besides R_{ex} is precisely this lack of scaling of parasitics in the layout. The dimensions of the base mesa shown in the figure and used in this work still maintain a mesa width on each side of the emitter finger that is many times the transfer length (L_t) of the base ohmic metallization; further reduction in C_{la} is still possible, and must be employed in highly scaled devices.

5.2 DC Measurements

The common-emitter DC current-voltage data for the highest f_τ device on each sample is shown in fig. 5.3, along with the mask dimensions of the collector and emitter stripes and the current gain β at $5 \times 10^4 \text{ A/cm}^2$. An important difference among the samples should be noted. Samples A and E were processed with an emitter undercut etch that was allowed to proceed at room temperature. The other samples were all processed with a cooled emitter undercut etch designed to improve reproducibility in the undercut as well as allow smaller undercuts. It is estimated that the amount of emitter semiconductor undercut for the room temperature etch is approximately $0.2\mu\text{m}$ on each side of the emitter stripe for the standard etch time used, e.g. a 1 μm emitter contact stripe would have the semiconductor underneath undercut to $0.6 \mu\text{m}$. This is referred to in fig. 5.3 as "deep undercut". For the cooled etch and the standard time used, the amount of undercut on each side of the emitter stripe is between 0.05 and 0.1 μm , and this is referred to as "shallow undercut" in fig. 5.3. This affects the β of the device as the distance from the emitter semiconductor to the base contact metal is determined by this etch. (The amount of emitter undercut is relevant to current gain - it determines the likelihood of electron recombination at the base metal. If the undercut distance is small compared to the length over which electrons could diffuse in their transit through the base, a reduction in current gain would be expected due to electrons diffusing out to the base metal). The base-emitter depletion capacitance (C_{je}) can be compared in order to determine the amount of undercut one sample experienced relative to others of similar geometry, but this comparison should only be made if

the base - emitter grade and the emitter are doped similarly between the two samples to be compared. (The only sample with a difference in the base - emitter grade or emitter is sample E). Table 5.2 contains the extracted C_{je} for each sample per μm^2 .

Sample B shows a reduction in the DC current gain near $1 \times 10^5 \text{ A/cm}^2$ current density. In this plot the cause seems to be the apparent device heating evidenced as a reduction in I_c as V_{ce} is increased. This reduction in β has been observed again in another HBT process carried out on the same wafer on which no device heating characteristics were observed in the DC current - voltage plot. The reduction in β due to Kirk Effect was suspect, especially as this is the sample with no pulse doping in the collector. Against this, however, is the fact that there is no catastrophic decrease in f_τ or f_{max} at these bias currents, at least not unusually more than on any other of the samples, also, the C_{cb} does not drastically increase at this current density as discussed in section 5.5.

Sample C shows an unusually low β (~ 30) for a $1 \mu\text{m}$ emitter geometry device. Comparison of the base-emitter capacitance with other wafers with similar geometry emitters does not seem to indicate a substantially smaller undercut. The low β is common to other devices on this wafer as well. This effectively rules out collector to emitter alignment issues that are often blamed for low current gain, as the device on this wafer with a $1 \mu\text{m}$ emitter width and a $2.5 \mu\text{m}$ collector width has low β as well. The reason for this low β is not known.

Sample D shows a higher β of 49, and this for an $0.7 \mu\text{m}$ emitter stripewidth device. The reason for this is undoubtedly the 300 \AA base. Due to the large scatter in the current gains observed between samples B, C, and F with the same emitter geometry, no conclusive statements can be made about the dependence of current gain on base thickness, apart from the obvious statement that a thinner base with the same bandgap grading results in higher current gain.

The higher β observed in sample E when compared with sample A is surprising when considering the thicker base of sample E (500 \AA as opposed to 400 \AA). A hint as to why the β is higher in sample E is offered in comparison of the base-emitter junction capacitance between A and E. The capacitance of sample E is expected to be lower due to the lighter n-type doping in the base-emitter grade, however, it was not expected to be as low as it is. Since the n-type doping in the base-emitter grade was reduced to

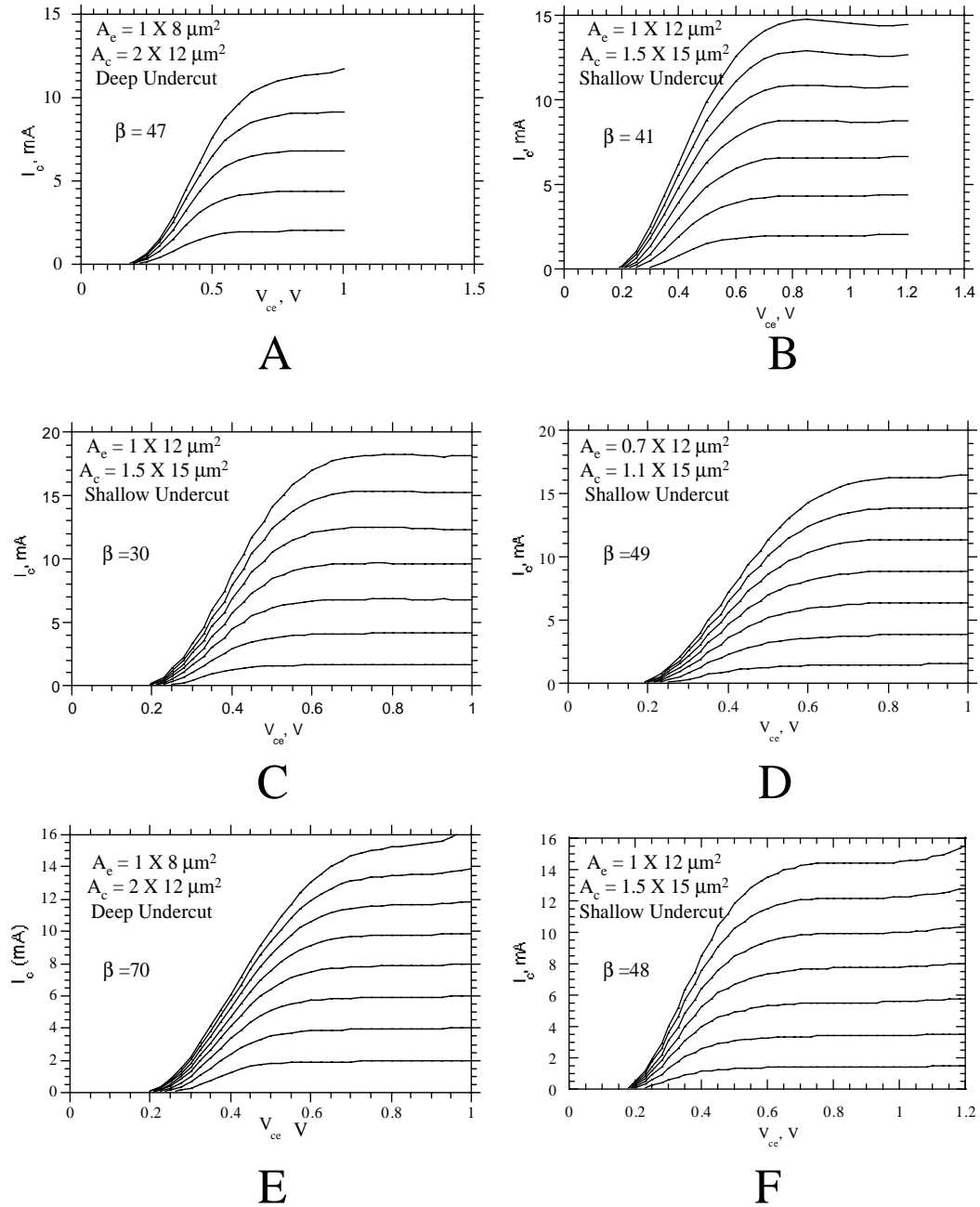


Figure 5.3: Common emitter characteristics for highest f_τ device on each sample. Lithographically defined emitter and collector dimensions and expected severity of undercutting etch are indicated.

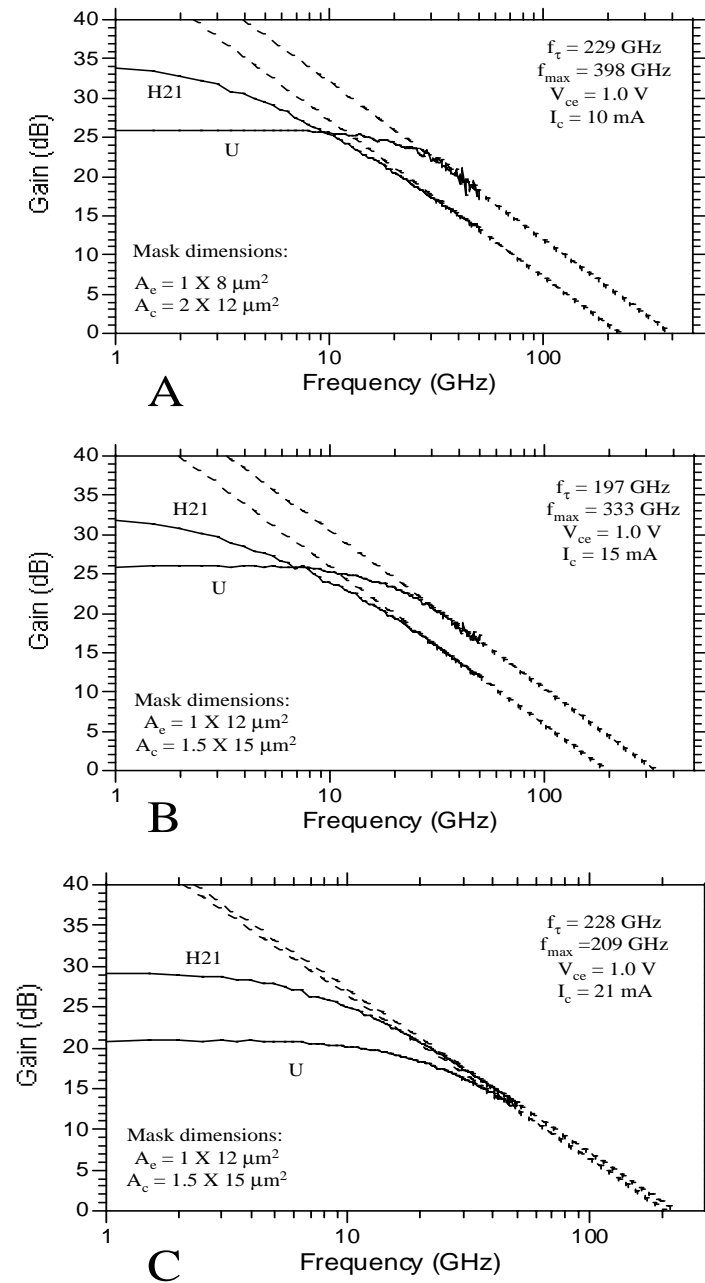
five-eighths of the baseline value of sample A, the depletion capacitance was expected to be reduced by at most $\sqrt{3/8}$, or about 20%. But C_{je} of sample E is less than half that of sample A. This departure from the expected value of C_{je} in junction capacitance indicates a smaller junction area, that is, the undercut was substantially more on sample E than on sample A.

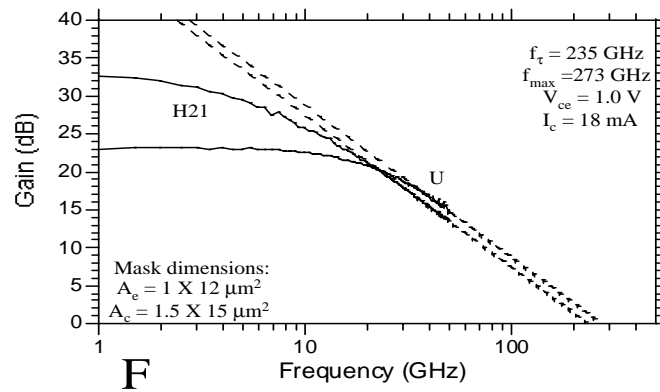
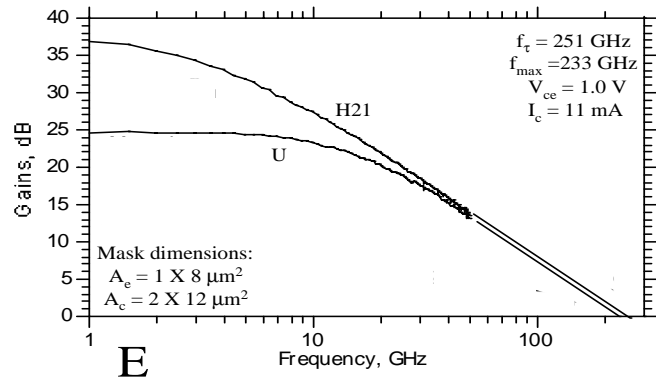
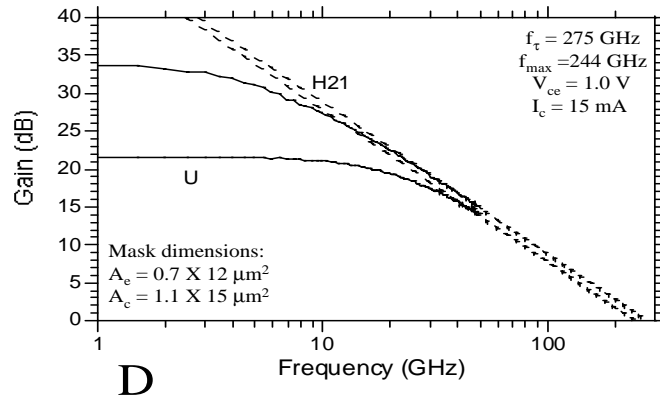
Sample F shows no unusual behavior.

5.3 Emitter Resistance

R_{ex} data is also presented in table 5.2 normalized to emitter contact area. Noteworthy is the fact that the two samples with increased exposure to the undercutting etch show the two highest values of R_{ex} . Also, sample F with the InAs emitter cap, shows only a slight reduction in R_{ex} that may simply be due to other process variations. Both of these observations contradict the expectations generated by calculation of the contribution of the various components of emitter resistance done in section 4.6. When this data is taken together with the nonlinear increase in emitter resistance with decreased emitter stripewidth also presented in section 4.6, the only conclusion that can be drawn is that the current knowledge of what makes up the measured emitter resistance is insufficient. There is some other effect.

It has been suggested that the exposed sidewall of the emitter-base grade gets oxidized, reducing the area of the junction further, but comparison of samples A and E tends to discount this. Sample A has a much higher C_{je} and lower current gain indicative of a larger junction area, yet it has the higher R_{ex} as well. The other possibility is that the etchant or some other chemical in the process attacks either the emitter metal or the metal - semiconductor interface. This could result in a rapid increase in R_{ex} with scaled emitters as well as the increased resistance with increased exposure to the etch. The anomalous R_{ex} problem must be solved in order to achieve a high f_τ on the highly scaled submicron devices that yield the highest f_{max} . Even the 0.7 μm device on which the 275 GHz f_τ was achieved would benefit greatly from reduced R_{ex} . Elimination of the R_{ex} would result in a 17% reduction in forward delay, a decrease of 0.095 ps. The f_τ would then not be 275 GHz, but 330 GHz.

Figure 5.4: RF data for $V_{ce} = 1.0$ V and peak $f_t I_c$.

Figure 5.5: RF data for $V_{cc} = 1.0$ V and peak $f_t I_c$.

5.4 Discussion of Delay Components

The individual terms that comprise the f_τ of the device are summarized for the six layer structures along with some other relevant device parameters in table 5.3. A natural comparison to make is between samples A and B. The difference between these two lies in the collector; while sample A is the baseline layer structure, sample B has no pulse doping at all, just a 3000 Å thick lightly doped collector. Sample B was expected to have a decreased collector transit time due to the lower electric field that would be present at the base-collector junction, and was also expected to have a lower Kirk effect threshold current density, also due to this lower electric field. The sum of base and collector transit time is seen to be higher for sample B than for sample A, and since the base and emitter layers are identical for the two structures, it can only be assumed that it is the collector transit time that is larger for sample B than for sample A. The pulse doping seems to reduce collector transit time.

The plot of C_{cb} as a function of current density at $V_{ce} = 1.0$ V for samples A and B is shown in fig. 5.6 and will be discussed in section 5.5. The Kirk effect is not obvious in either sample, as the C_{cb} changes little over the whole available range of current density. The highest current density shown in the plot is that for which the devices were beginning to break down; if there is significant Kirk effect at higher current density, it is irrelevant to the performance of these devices. However, the f_τ of the device on sample A (with the pulse doping) is higher, and the sum of base and collector transit time is lower. The absence of a strong Kirk Effect in sample B is somewhat puzzling and may be a sign of the successful suppression of beryllium diffusion into the collector.

Comparing samples B and C the expected differences in transit time and C_{cb} are noticed. The sum of base and collector transit time is 0.09 ps lower in sample C than sample B, due to the 2000 Å thick collector in sample C, also with no pulse doping. The terms in forward delay resulting from charging of the C_{cb} are not as much higher in sample C than sample B as would be expected due to the higher current density at which sample C could be measured.

Sample D contained a 300 Å base in order to reduce transit time further. The ability to bias the $0.7 \mu\text{m} \times 12 \mu\text{m}$ device at 18 mA (2.1×10^5 A/cm²) resulted in low charging times for the capacitors, although the layout of fig.

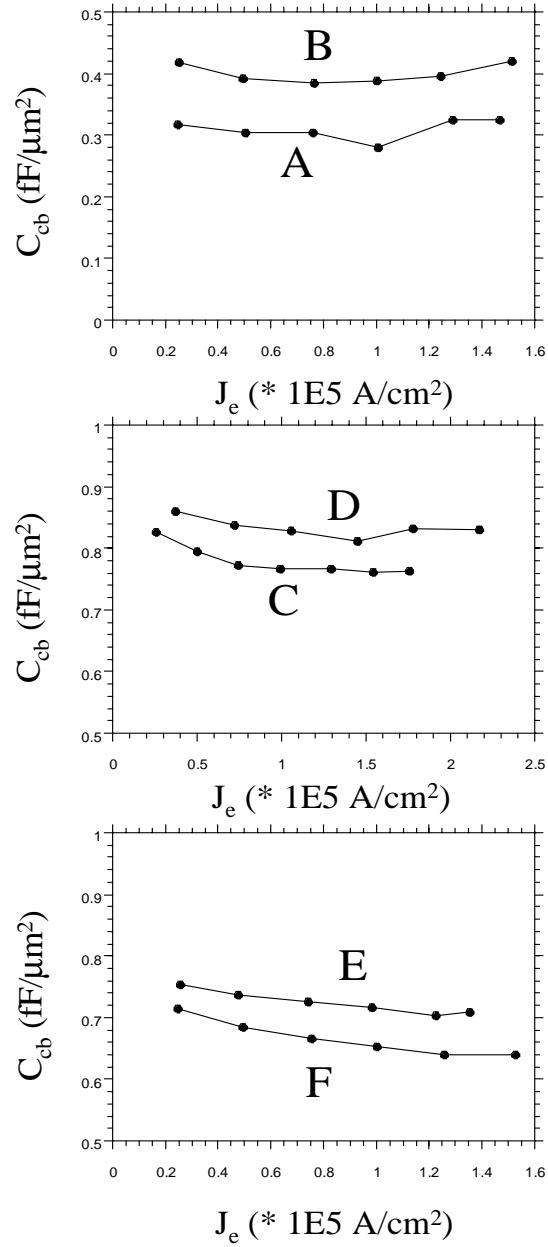
Figure 5.6: Normalized C_{cb} vs. J_e at $V_{ce} = 1.0$ V.

Table 5.2: Comparison of the device yielding the highest f_τ on each wafer at $V_{ce} = 1.0$ V and the optimum I_c . W_e and W_c are the mask dimensions. J_e is the current density calculated from emitter current divided by the emitter mask dimension.

	A	B	C	D	E	F
T_b (Å)	400	400	400	300	500	400
T_c (Å)	3000	3000	2000	2000	2000	2000
Pulse Dope ($\times 10^{11}/cm^2$)	5	0	0	0	3.5	3.5
R_{ex} ($\Omega \times \mu m^2$)	56.8	44.4	48.0	38.6	49.6	42.0
C_{je} (fF/ μm^2)	4.6	5.5	5.0	4.2	2.1	4.8
C_{la} (fF)	5.1	7.0	4.0	7.0	4.4	7.0
$W_e \times L_e$ ($\mu m \times \mu m$)	1×8	1×12	1×12	0.7×12	1×8	1×12
$W_c \times L_c$ ($\mu m \times \mu m$)	2×12	1.5×15	1.5×15	1.1×15	2×12	1.5×15
J_e (mA/ μm^2)	1.25	1.25	1.75	2.14	1.38	1.5
$\tau_b + \tau_c$ (ps)	0.46	0.55	0.46	0.37	0.38	0.46
$r_e C_{je}$ (ps)	0.11	0.13	0.087	0.079	0.056	0.096
$r_e C_{cb}$ (ps)	0.022	0.018	0.025	0.022	0.045	0.025
$r_e C_{la}$ (ps)	0.016	0.014	0.006	0.011	0.012	0.012
$R_{ex} C_{cb}$ (ps)	0.050	0.033	0.068	0.063	0.11	0.053
$R_{ex} C_{la}$ (ps)	0.036	0.026	0.016	0.032	0.027	0.025
f_τ (GHz)	229	197	228	275	250	235

5.2(c) was used which led to increased parasitic layout capacitance. For some reason the R_{ex} on this wafer was substantially lower than all of the others. The C_{je} indicates a level of undercut somewhat more than the other cold-etched samples which would be expected to give a slightly higher R_{ex} than the others. The baseline emitter layer structure and cap was used. Whatever the reason for the low R_{ex} , the combination of low transit time, relatively low charging time, and the low RC time constants from the low R_{ex} combined to produce a device with record f_τ of 275 GHz.

Sample E yielded a very high f_τ , especially considering the 500 Å base. The transit time is quite low, rivaled only by sample D. At present there is no plausible explanation for the low transit time; sample F had the same collector layer structure and a 400 Å base resulting in a higher transit time. Sample F also had an InAs emitter cap layer which was expected to reduce the R_{ex} , but this did not happen. Sample E benefited from the lighter doping in the base-emitter grade, as the C_{je} can be seen from table 5.2 to be a fraction of what it is in any other sample. The R_{ex} of sample E is somewhat high just as that of the other sample etched under room temperature condition, sample A.

5.5 Pulse Doping in the Collector

The pulse doping in the collector continues to be an issue. It seems to be desirable in a 3000 Å thick collector due to the observed improvement in the collector transit time. In highly scaled, high f_{max} devices (reported by Michelle Lee at UCSB [28]) it may still be necessary to include the pulse doping in order to raise the Kirk Effect threshold current density. Samples A and B exhibited breakdown before they reached the current densities ($2 - 3 \times 10^5$ A/cm²) at which Lee's devices achieve the highest f_{max} . At current densities of this magnitude, it is entirely possible that significant base pushout may occur if not for the presence of the pulse doping. The pulse doping certainly does not seem to have any disadvantage for the case of the 3000 Å collector.

The results from the presence or lack of pulse doping in 2000 Å thick collector samples (C, D, E, F) can be interpreted the same way. Kirk Effect is stronger in samples C and D without the pulse doping, as samples E and F show monotonically decreasing C_{cb} at V_{ce} of 1.0 V while C and D do not. Sample D achieved the highest f_τ of all samples due mostly to its

(anomalously) low R_{ex} and the low transit time due to the 300 Å base. As the devices are scaled further and the achievable current densities increase, the pulse doping may be necessary even in these thinner collector devices to forestall the onset of Kirk Effect. If future designs are to incorporate even thinner collector semiconductor regions, the issue of whether or not to pulse dope should be determined experimentally in the case of the device geometries of interest.

5.6 Collector-Base Capacitance

For samples B, D, and F, fig. 5.7 plots C_{cb} as a function of bias current density for various V_{ce} . In general for V_{ce} less than 1.0 V, the devices exhibit an increase in C_{cb} as the collector current increases; manifestation of Kirk Effect. It is expected to occur at these low V_{ce} values due to the reduced electric field present at the base - collector junction. Above V_{ce} of 1.0 V the C_{cb} actually decreases as more collector current is injected, due to space - charge screening effects in the collector. This space - charge screening effect is described in the paper by Camnitz and Moll [25]; the equation derived there for the intrinsic collector - base capacitance is

$$\frac{C_{ci}}{A_e} = \frac{\epsilon}{W_c} - \frac{\kappa_1 J_c}{2} \left(1 - \frac{\kappa_1 J_c W_c}{6\epsilon}\right), \quad (5.2)$$

the factor κ_1 is a coefficient accounting for the negative differential mobility evident in the velocity - field curve. The termination of some electric field lines on the traveling electrons reduces the interaction between base and collector. The result is a real reduction in C_{cb} as current is increased, this effect is also known as "capacitance cancellation", as the increase in space charge in the collector reduces the collector - base capacitance. This is probably the dominant effect resulting in the extremely high values of f_{max} reported in highly scaled transferred-substrate HBTs. In devices fabricated with deep submicron collector dimensions, the reduction in C_{cb} accounts for a large fraction of the total C_{cb} resulting in an extreme dependence of f_{max} on bias current [28].

At some V_{ce} the Kirk effect and the space - charge screening effect effectively cancel each other out resulting in a relatively constant C_{cb} curve as a function of bias current. For the 2000 Å collector samples with the pulse doping (samples E and F) the space - charge screening effect dominates at

V_{ce} of 1.0 V as seen in fig. 5.6. Samples C and D, 2000 Å collector with no pulse doping, show a stronger Kirk Effect at 1.0 V V_{ce} as the C_{cb} falls with increasing current but then levels off. Samples A and B with the 3000 Å collector show cancellation of the two effects resulting in little dependence of C_{cb} on bias current.

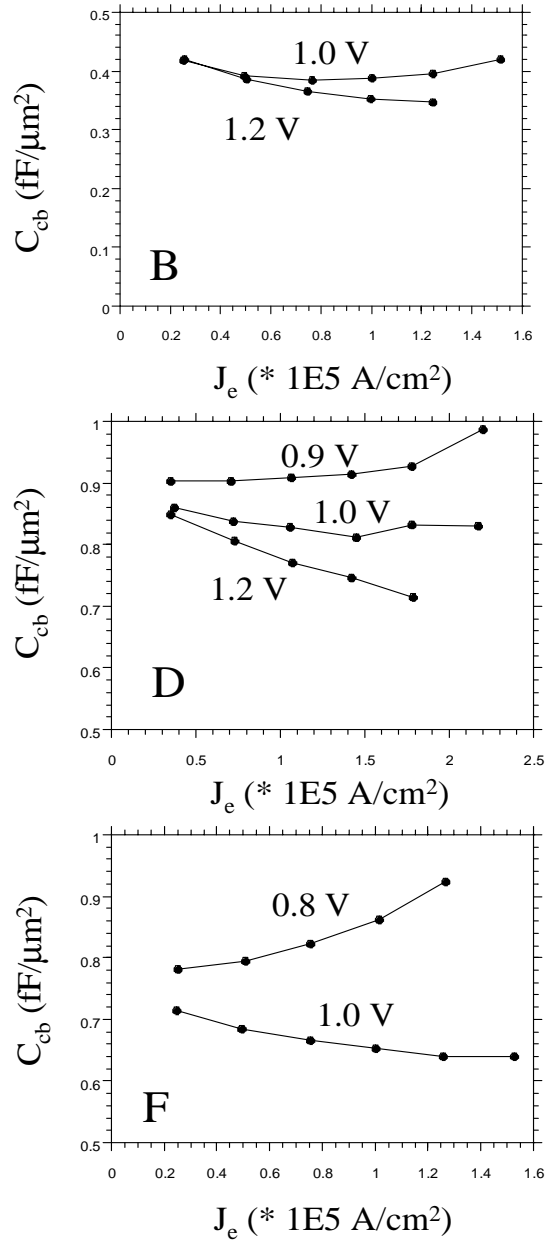


Figure 5.7: Normalized C_{cb} vs. J_e at labeled V_{ce} for samples B, D, F.

Chapter 6

Resistive Feedback Amplifiers

It is useful to construct amplifiers that can work in the standard 50- Ω environment. Probably the simplest amplifier would be the common-emitter configuration of a transistor, where the input drives the base, and the output is taken at the collector. A network analyzer, which tests the device with 50 Ω source and load impedances, will show that there is significant power loss when the device is driven in this way due to the impedance mismatch at input and output. This is due to the very non-50- Ω nature of the input and output of the intrinsic device. A common-emitter device is very inefficient in a 50- Ω environment.

One way to reduce the reflections on input and output is to shunt the input and output with 50- Ω resistors to ground. This is an effective technique from the standpoint of impedance matching, but has at least one serious drawback. A large fraction of the signal power is now dissipated in these resistors instead of being provided to the input and delivered from the output. So while this may protect the voltage source from the harmful effects of standing waves and ripples in frequency response due to standing waves or long 50 Ω interconnects, it is still a very inefficient method of impedance matching in terms of available gain.

Resistive feedback can be used from output to input in order to provide 50- Ω impedance at the input and output ports. The value of resistance to use is set by the gain, with $R_f = Z_o(1 - A_v)$, where A_v is the voltage gain of the amplifier. The resistive feedback amplifier (RFA) can be implemented in many of the standard single or multi-transistor gain stages. Deviation from 50 Ω will occur at high frequencies as the parasitic capacitors in the device or

circuit will provide a shunt path for the high-frequency currents. The high f_τ and f_{max} demonstrated by transferred-substrate HBTs facilitates the construction of resistive feedback amplifiers that offer superior performance combined with the ease of design and testing provided by these small circuits designed for a 50- Ω environment.

An alternative to resistive feedback is to use reactive tuning in a reactive tuned amplifier to provide a 50- Ω input and output impedance. The limitation of reactive tuning is summed up by the word "tuning"; the 50 Ω -impedance is available only over a narrow range of frequency. If tuned properly, the maximum available gain of the stage can be realized, the gain is not limited by f_τ . But many applications require broadband amplification, some require gain even at DC, and so this reactive tuning is not always feasible. Furthermore, the design of tuned amplifiers requires precise knowledge of both the device parasitics and the transmission line parameters; otherwise tuning will be adversely affected.

The design of RFAs was undertaken for many reasons. They are useful as building blocks in larger systems, as well as being useful amplifiers in their own right. RFAs are a good indicator of the current status of the process. They are the simplest circuits that require all of the passive elements of this IC process: resistors, MIM bypass capacitors, microstrip inductors and microstrip transmission lines. As opposed to tuned amplifiers, the gain of the RFAs fabricated in this work has demonstrated remarkable insensitivity to device parasitics and transmission line parameters. (The return losses, however, have not). RFAs designed in this work have demonstrated record gain-bandwidth products and with far lower power dissipation than competing technologies that must use distributed amplification in order to achieve the same bandwidth. Most distributed amplifiers do not provide gain at DC and at low frequencies due to the use of blocking capacitors in the signal path.

The RFAs in this work have been designed for ease of testing. This lab is currently equipped with a 0-50 GHz and a W-band 75-110 GHz test set. Amplifiers have been designed to have a 3-dB bandwidth that lies within one of these two ranges, so that when possible it can be measured and not extrapolated. This has not always been possible, as the bandwidth achieved has not always been what was designed for.

A variety of resistive feedback amplifier configurations have been considered for this work, shown in fig. 6.1. The common emitter is simple,

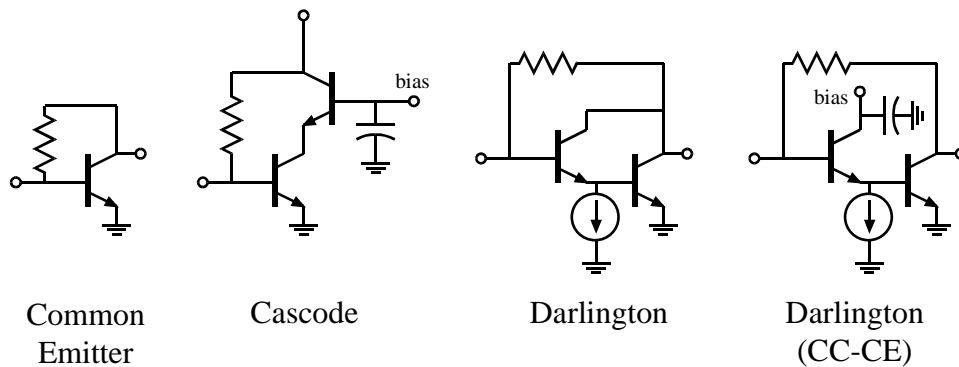


Figure 6.1: Resistive feedback amplifier configurations

but multiple transistor configurations can achieve greater bandwidth. The cascode configuration increases bandwidth due to the low voltage gain of the input transistor. The voltage gain of the input transistor depends on the relative sizes of the transistors, but most likely will be near unity. The Miller Effect (in which the collector - base capacitance is effectively multiplied by the gain between input and output of the relevant transistor) is effectively eliminated. The difficulty with the cascode is in the biasing of the common-base transistor. The base of this transistor should be an AC ground in order to achieve the maximum gain, requiring a large bypass capacitor. Such a large bypass capacitor (order of pF) takes up a large area and makes the layout awkward. A radial stub capacitor is required in order to eliminate the risk of resonance and/or oscillations. Unless the two transistors in the cascode are of nearly the same size, they should be AC coupled in order to provide independent bias current to each one. This AC coupling introduces the aforementioned low frequency cutoff which would attenuate low frequency signal information. The freedom to choose different transistor sizes is also desirable, and so the cascode configuration was not pursued.

The Darlington configuration wherein the collectors are connected at the output is used in some applications to increase the effective current gain and as a replacement for a single transistor. The drawbacks of this configuration are a reduced output resistance (R_o becomes the parallel combination of the R_o from each transistor) and the fact that there is a parasitic capacitance from output to input (C_{cb} of the input transistor) that is effectively

Table 6.1: Device specifications for Darlington amplifier.

Device	Emitter dimension	Collector dimension	Bias Current
Q1	$1 \times 8 \mu\text{m}^2$	$2 \times 12 \mu\text{m}^2$	5.8 mA
Q2	$1 \times 16 \mu\text{m}^2$	$2 \times 20 \mu\text{m}^2$	11.6 mA

multiplied due to Miller effect by $(1 - A_v)$.

Finally, the common-collector common-emitter configuration has been examined. The collector of the emitter-follower transistor can be independently biased. No Miller effect is observed at the input to the emitter-follower transistor, as the gain here is less than one. The Miller effect that occurs at the input to the common-emitter transistor, that is, in the internal node, is less of a restriction than in an isolated common-emitter device with 50Ω input impedance. The Miller capacitance, instead of multiplying the 50Ω to form an RC time constant, is multiplied by the resistance looking back into the emitter of the emitter-follower transistor. This resistance is $r_e + R_{bb}/\beta + 50\Omega/\beta$. For a transistor of sufficient size and with sufficient current, the sum of these terms will be substantially less than 50Ω . An increase in the bandwidth results as the usually dominant input pole is pushed higher in frequency.

Simulation of these amplifiers must take into account the parasitics introduced by the interconnect wiring. There is a substantial difference between the simulated performance of the amplifier with and without these layout parasitics. The interconnect capacitance can cause gain peaking and/or a decrease in the 3dB bandwidth of the amplifier. If the interconnects are already as short or narrow as possible, the effects on the circuit performance must either be accepted or the circuit modified in some way to compensate for it.

6.1 First Generation Amplifier Design

A circuit diagram for the Darlington feedback amplifier is shown in fig. 6.2. A chip photo is shown in fig. 6.3. Provision is made for independent biasing of the collector of Q1, requiring a third pad. The area of the chip is

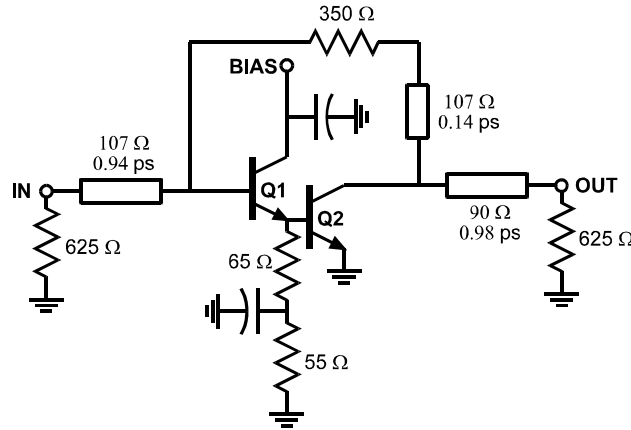


Figure 6.2: Circuit diagram of Darlington Amplifier. Line section characteristic impedance and electrical length as indicated.

then pad limited at $0.52 \text{ mm} \times 0.43 \text{ mm}$, or 0.224 mm^2 . Simulated values of DC gain and 3 dB bandwidth are 15.2 dB and 52 GHz, respectively. The radial stub bypass capacitors have a (simulated) capacitance of 300 fF. Tuning inductors and shunt resistors to ground are used on input and output to reduce reflection loss and increase bandwidth. No thermal via is provided on the degeneration resistors of Q1, as the parasitic capacitance from the via degrades the amplifier bandwidth. As the IC process is still under development, design rules do not yet exist pertaining to the current carrying capacity of resistors with and without heat sinks. The feedback path consists of a 350Ω resistor in series with a short section of high impedance interconnect (fig. 6.2). A thermal via is provided on each device, which on the input device represents a small but not negligible parasitic capacitance from the emitter to ground. This capacitance arises from the overlap of the M1 to the grounded M2 through the SiN dielectric.

Fig. 6.4 and fig. 6.5 show a circuit diagram and chip photo of the so-called "mirror doubler" amplifier, a configuration in which the current-gain cutoff frequency of *the stage* is effectively doubled [30]. The amplifier of fig. 6.4 uses an emitter follower to drive a mirror doubler. A third pad is also used in this design to provide independent collector bias of the emitter follower, resulting in the same chip area of 0.224 mm^2 . DC gain and 3 dB bandwidth simulate at 7.5 dB and 94 GHz. The radial stub bypass

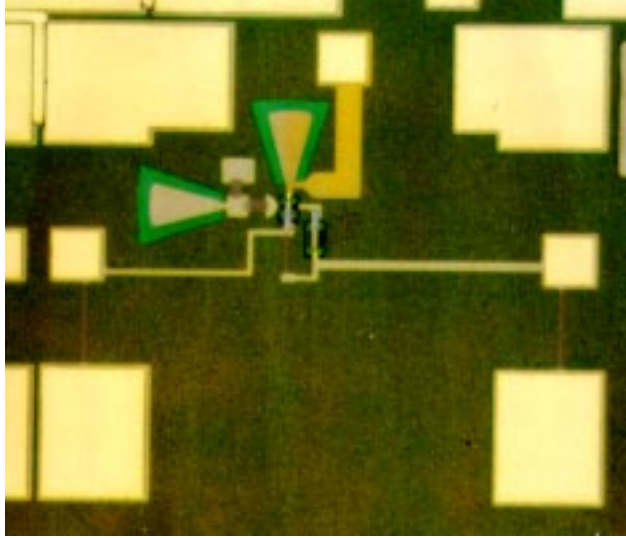


Figure 6.3: Chip photo of Darlington Amplifier.

capacitor is designed to have 150 fF capacitance. Tuning inductors are also used in this design, as well as a shunt resistor to ground on the output. The diode Q4 is for DC level shifting, to maintain the V_{ce} of Q5 near the region of peak RF performance. The feedback path consists of a 150 Ω resistor in series with a small inductance. Thermal vias are provided on each device except for the diode Q4. The diode Q4 is three times larger than Q2, Q3, and Q5 to minimize parasitic resistance at this node. The diode is biased at a low 2.3×10^4 A/cm² current density, reducing the need for a thermal via. A thermal via would present an imposing parasitic capacitance to ground for such a large device, and so it has been eliminated.

Power dissipation in the emitter degeneration resistor of Q1 is a difficulty. There is no thermal via on this resistor. This resistor demands the highest current carrying capacity of any resistor on the two amplifiers at 0.5 mA/ μ m under designed biasing conditions. The resistor fails if Q1 is biased at greater than about 15 mA emitter current. Depending upon their length, resistors without vias will fail at current densities in the range of 0.2 to 0.5 mA per micron of width. In the present design, addition of a thermal via to the resistor would prevent thermal failure, but the via capacitance would significantly degrade circuit bandwidth. Without the via, a resistor

Table 6.2: Device specifications for mirror doubler amplifier.

Device	Emitter dimension (on Mask)	Collector dimension (on Mask)	Bias Current
Q1	$1 \times 16 \mu\text{m}^2$	$2 \times 20 \mu\text{m}^2$	11.6 mA
Q2	$1 \times 8 \mu\text{m}^2$	$2 \times 12 \mu\text{m}^2$	5.8 mA
Q3	$1 \times 8 \mu\text{m}^2$	$2 \times 12 \mu\text{m}^2$	5.8 mA
Q4	$1 \times 25 \mu\text{m}^2$	$2 \times 29 \mu\text{m}^2$	5.8 mA
Q5	$1 \times 8 \mu\text{m}^2$	$2 \times 12 \mu\text{m}^2$	5.8 mA

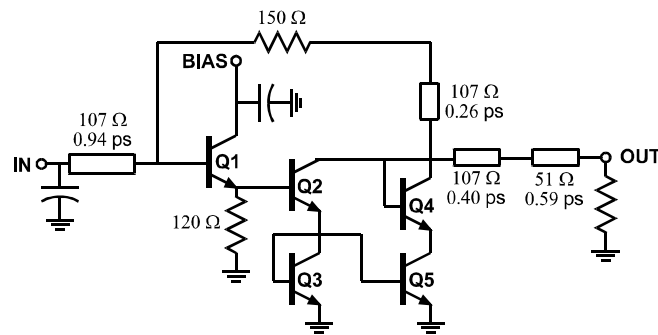


Figure 6.4: Circuit diagram of Mirror Doubler.

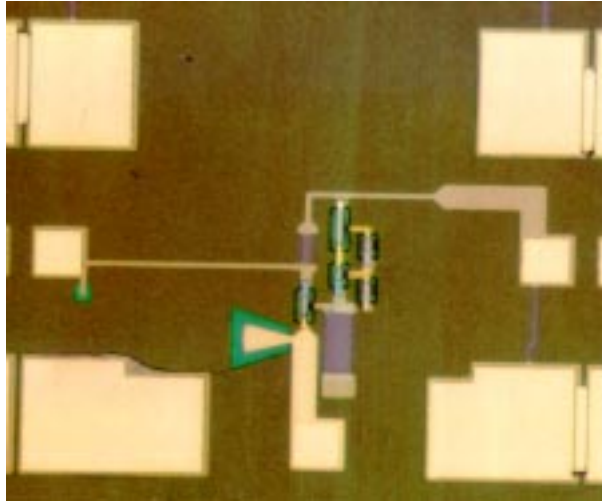


Figure 6.5: Chip photo of Mirror Doubler Amplifier.

chosen for a safe current density would be at least 60 microns wide and 144 microns long, dominating the IC area.

6.2 First Generation Amplifier Results

The amplifiers were measured on wafer from 0.5-50 GHz and from 75-110 GHz. S-parameters vs. frequency are plotted in fig. 6.6 and fig. 6.7 for the darlington amplifier and the mirror doubler amplifier, respectively. The DC gain of the Darlington amplifier is close to simulation at 15.6 dB, with a 3 dB bandwidth that is greater than 50 GHz. The gain peaks at 16.0 dB, and falls to 14.9 dB at 50 GHz. At 75 GHz, the gain has fallen to 8.8 dB, preventing precise determination of the 3 dB bandwidth. An estimate of the 3 dB bandwidth can be made by connecting the 50 GHz and the 75 GHz measurements with a line. This gives a 3 dB bandwidth of 60 GHz, for a gain-bandwidth product of 360 GHz. The power dissipation is 36.3 mW.

The DC gain of the mirror doubler amplifier is 6.8 dB. The gain steadily rises with frequency until at least 50 GHz where it reaches 9.5 dB. The frequency at which the gain has dropped to 3 dB below the DC gain is 85 GHz, giving a gain-bandwidth product of 180 GHz. The power dissipation

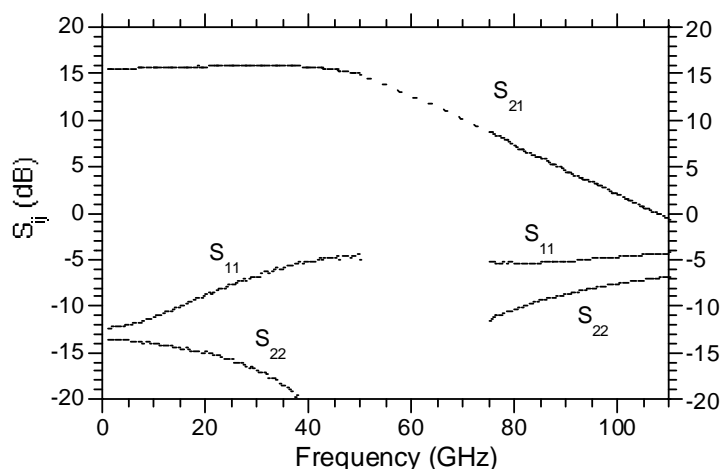


Figure 6.6: Measured S parameters of the Darlington Amplifier.

is 24.2 mW.

6.3 Second Generation Amplifier Design and Results

In second generation designs of the Darlington amplifier, the resistive biasing of the emitter-follower transistor is replaced with a current mirror bias source, avoiding difficulties with dissipation. Two new specifications were attempted. Since the 15.6 dB gain amplifier achieved greater than 50 GHz bandwidth and therefore the 3dB bandwidth could not be precisely measured, an amplifier with more than 18 dB of gain was designed for about 50 GHz bandwidth. The mirror-doubler amplifier exhibited gain peaking of almost 3dB. A high bandwidth amplifier was desired with flatter gain-frequency characteristics.

A generic circuit diagram of the revised amplifier designs is shown in fig. 6.8. The addition of the current mirror does require somewhat more power dissipation, but with improved bias stability. Degeneration resistors on the current mirror transistors have been provided with heat sinks, as parasitic

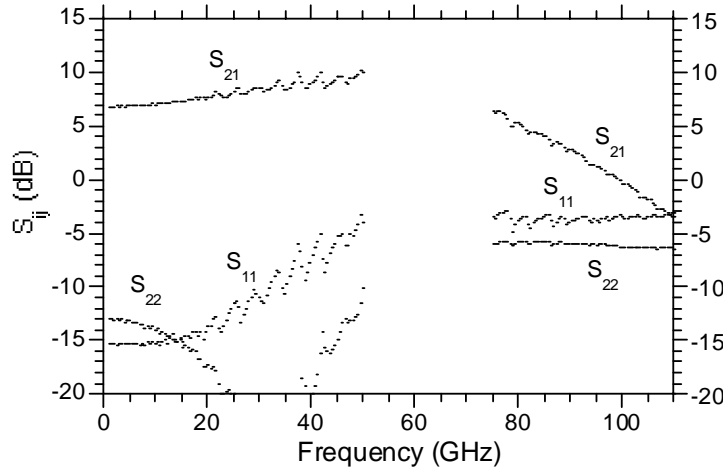


Figure 6.7: Measured S parameters of the Mirror Doubler Amplifier.

capacitance to ground at this node does not affect the circuit performance. The feedback resistor and the shunt resistors to ground on the input and output do not have heat sinks, and have been sized to conform with more conservative design rules. The bypass capacitor on the collector of Q1 is a radial stub estimated at 300 fF. Larger values of capacitance here would provide a lower impedance path to ground for AC signals, but simulations showed that this only increased the gain peaking and return losses of the circuit. The size of this chip is still limited by the area of the pads; the pads are designed to accommodate the 150 μm pitch microwave probes.

A chip photo of the high-gain amplifier is shown in fig. 6.10; the wider bandwidth version would be nearly indistinguishable at this magnification and so is not shown. A slightly more complicated layout is evident due to the current mirrors, and there is a point at which a M2 biasing line crosses over the M1 output inductor. This is necessary due to the topology of the circuit and the number of biasing pads. The parasitic capacitance of this crossover has been made as small as possible by having the M2 biasing line taper down to 4 μm wide at the point of the crossover, resulting in a $4 \times 4 \mu\text{m}^2$ capacitor from M1 to M2 through silicon nitride. The capacitance is a nearly insignificant 4 fF, however the sum of many of these crossover

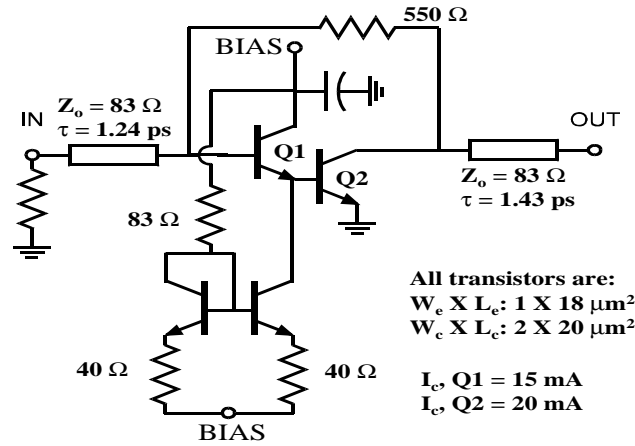


Figure 6.8: Circuit diagram of high - gain, second generation amplifier.

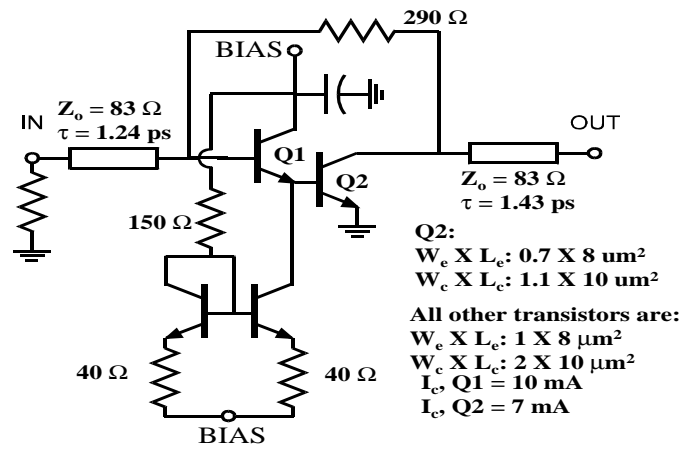


Figure 6.9: Circuit diagram of wide - bandwidth, second generation amplifier.

capacitances in a large chip could significantly affect performance.



Figure 6.10: Chip photo of high-gain Darlington amplifier.

S-parameters vs. frequency are plotted in fig. 6.11 and fig. 6.12 for the high-gain amplifier and the high-bandwidth amplifier, respectively. The DC gain of the high-gain amplifier is again close to simulation at 18.1 dB, with a 3 dB bandwidth that is greater than 50 GHz. The gain peaks at 18.5 dB, and falls to 16.0 dB at 50 GHz. W-band measurements were not performed for this amplifier, as the 3dB bandwidth appears to be only slightly above 50 GHz. An estimate of the 3 dB bandwidth can be made by connecting the 50 GHz and the 75 GHz measurements with a line. The gain-bandwidth product for this Darlington amplifier is over 400 GHz, and the power dissipation is 100 mW. The high-bandwidth amplifier has a DC gain of 6.3 dB, somewhat lower than the 8.5 dB expected from the simulation. The gain dips down to 5.7 dB at 20 GHz, but rises back up to 6.3 dB at 50 GHz. W-band measurements indicate a 3dB bandwidth of 81 GHz, significantly lower than expected from simulation.

A possible explanation for the lower gain and bandwidth than was expected from simulation is due to the use of a $0.7 \mu\text{m}$ emitter stripe in transistor Q2. For the simulation, the device model only requires specification of the emitter area, it does not take into account the emitter stripe width. Recall from section 4.6 that the emitter resistance is not inversely

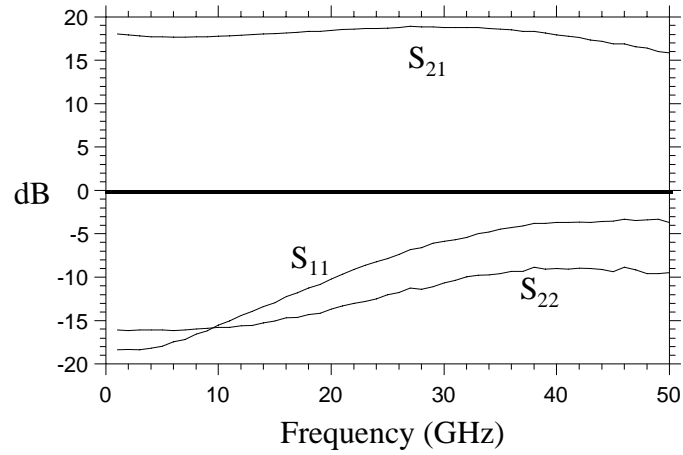


Figure 6.11: Measured S parameters of the high-gain Darlington Amplifier.

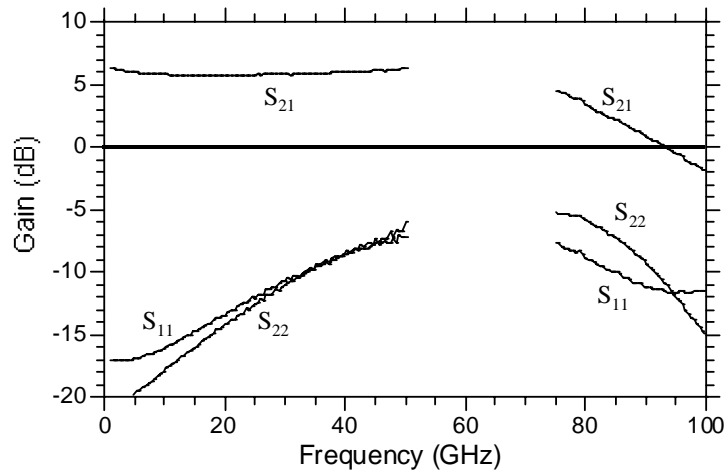


Figure 6.12: Measured S parameters of the high-bandwidth Darlington Amplifier.

proportional to the emitter area, but increases much faster than expected when the emitter stripwidth is reduced. The degeneration resistance on Q2 in this circuit was therefore substantially higher than expected.

6.4 Discussion

The gain-bandwidth product of these resistive feedback amplifiers far exceeds what is currently considered state-of-the-art. The return losses have fallen far short of expectation, however. Amplifiers reported in the literature are usually designed to have both S_{11} and S_{22} below -10dB over the entire circuit bandwidth. This is an arbitrary cutoff for what is considered to be an acceptable amount of reflection at the input and output ports of the amplifier. Put another way, an amplifier with an S_{11} or S_{22} above -10 dB is said to have an unacceptable amount of "return loss", due to the reflection of incident power.

The S_{11} and S_{22} show the same characteristic shape. At DC, these s-parameters start very low, below -20 dB. This indicates a good match to 50 Ω at these low frequencies, simply confirmation that the feedback resistor is near the designed value. Both S_{11} and S_{22} rise as the frequency increases, and at some frequency near 30 GHz, they both peak at values well above -10 dB, then drop off again as the frequency increases further.

The combination of device parasitics, the attempt at tuning networks, and the pad capacitance cause the input and output impedances to deviate from 50 Ω at higher frequencies. Variations in layer structure, device performance, and processing can all cause an increase in the return loss. The tuning networks have been designed in conjunction with a certain device model, and if the device parasitics differ from the simulation, the tuning network will be ineffective. One of the important variations in processing that may easily occur is in the etching of the BCB. Overetching the BCB will cause the BCB to be too thin, increasing the parasitic capacitance of the microstrip wiring. This results in lower impedance lines than expected, which will also throw off the tuning networks.

The tuning networks are not tuning networks in the sense of narrowband matching. For the most part (except for the case of the first generation Darlington amplifier) the tuning networks are simply a series inductor designed to tune out some of the input or output capacitance. The first generation Darlington amplifier has this series inductor, but also a shunt capacitor

which brings the reactive tuning closer to 50Ω at the highest extreme of the 3dB bandwidth.

Chapter 7

Fiber-Optic Receiver Design

A few building blocks of fiber-optic ICs have been simulated and fabricated. Testing and revisions to these components were not pursued as attempts to increase transistor f_τ were in progress. In order to obtain the device parameters used in modeling these circuits, a conservative layer structure was used which was not interesting from the point of view of improving f_τ . As described below, these chips would be useful for high bit-rate fiber-optic receivers. The InP based materials system used in this work is ideal for photodetection at 1.55 μm wavelength. The 1.55 μm wavelength is nearly ideal for long-haul fiber-optic communication, as absorption loss exhibits a local minimum at this wavelength.

A block diagram of an example fiber-optic receiver is shown in fig. 7.1. The data is converted to electrical form by the photodiode. The photodiode should have as high a quantum efficiency as possible. That is, the fewer the number of incident photons required to produce an electron-hole pair, the better the sensitivity of the receiver to weak incoming signals. The photodiode must also have a flat response over the required bandwidth, and not exhibit gain peaks or nulls. Coupling an optical fiber to the photodiode should be simple and reproducible, incident optical power should be completely delivered to the photodiode.

Following the photodiode is a preamplification stage to boost the signal. The preamplifier has many requirements. The noise figure of the receiver depends to a large extent on the gain of the preamplifier. The preamplifier should have as high a gain as required to minimize noise, but the gain should not be so high as to saturate the following stages. The gain vs. frequency

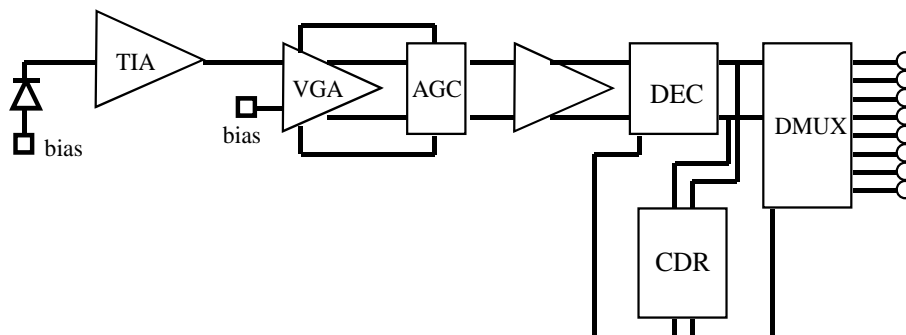


Figure 7.1: Fiber-optic receiver block diagram showing transimpedance amplifier (TIA), variable gain amplifier (VGA), automatic gain control (AGC), decision circuit (DEC), clock and data recovery circuit (CDR), and demultiplexer (DMUX).

characteristic of the preamplifier should be flat. Gain ripple or gain peaking with frequency must be minimized in order to prevent further distortion of the incoming pulses. Dispersion of the pulses after traveling through long distances of optical fiber can be severe, and the bit-rate-distance product, used as a figure of merit for long-haul communications, should be limited by the optical fiber and not by the receiver electronics. The photodiode at the input complicates design of the preamplifier. The Darlington feedback amplifiers described in chapter 6 are designed to operate with $50\text{-}\Omega$ source and load resistance. In this case, the source impedance is that of the photodiode, which is comprised of a very small capacitance and a series resistance, resulting in a source impedance far from $50\ \Omega$.

After the preamplification stage, the signal reaches the automatic gain control stage. This is an amplifier with variable gain within a feedback loop that adjusts the gain to maintain a constant output power. There are many factors that may affect the power of the received signal, and it is the job of the AGC stage to prevent the following stages from experiencing any variation in signal power. The output power of the laser diode may drift due to a change in temperature or simply due to aging of the laser. Fiber losses will vary due to use of connectors and splices; the receiver electronics should be able to accommodate a range of signal power. The receiver may be used in a local area network (LAN) within a building, in which case

the signal has not traveled through a significant distance of fiber when it is received. Ideally the same receiver could be used after a long distance of fiber or for when the signal power has been diluted by the optical fiber branching out to provide the signal to many listeners.

The demands on the automatic gain control stage are many. It must have a gain high enough to amplify the weakest incoming signal to the level required by the input of the decision circuit. It must be able to reduce that gain such that the strongest expected input signal does not result in a signal which saturates the input of the decision circuit. In addition to these constraints on the dynamic range of the AGC stage, the bandwidth and the gain flatness constraints are the same as those of the preamplifier, in addition they should be maintained over the whole available range of gains.

The decision circuit that follows the AGC stage can be as simple as a flip-flop. This stage is labeled A/D in the block diagram, as its function is to convert the analog and dispersed incoming signal to digital form. Full logic levels are restored to the analog like signal at this point, and the slow and noisy transitions are cleaned up and made more abrupt. This circuit must be clocked with the clock rate of the incoming data. The clock signal for the decision circuit and the demultiplexer is provided by the clock recovery circuit, which extracts the clock signal from the incoming data.

Finally, the demultiplexer separates the incoming data into its component channels and routes each channel along a separate interconnect. Clocking for the demultiplexer circuit is also provided by the clock recovery circuit.

The fiber-optic receiver components fabricated in this work have been designed to have bandwidth up to 40 GHz, much more bandwidth than is required for a 40 Gb/sec system. The minimum input power this system need be able to handle is the often quoted level of -9 dBm. The voltage swing required at the input of the decision circuit is 600 mV peak-peak. The gain of the photodiode and amplifiers must be sufficient to convert -9 dBm of input optical power to provide a voltage swing of 600 mV. The dynamic range of the AGC stage determines the maximum input optical power that can be handled; here the ability of the AGC stage to reduce the gain is necessary in order to provide only 600 mV peak-to-peak at the input of the decision circuit.

The designs of the component subcircuits are now discussed in greater detail.

7.1 Photodetector

A high-speed p-i-n photodetector is rather simple to implement into the transferred-substrate technology, using the collector layer as the infrared absorber. An n+ subcollector must be part of the layer structure. The quantum efficiency of this photodiode will be low, as the absorbing collector region is only 2000 Å thick. It would be possible, however, to improve this quantum efficiency by applying some reflection reducing dielectric coatings to the surface, as will be described below.

Processing the photodiode begins with the base contact etch, where all of the emitter semiconductor is etched away in the region of the photodiode. The base contact metal forms the back, p-contact. A separate lithography dedicated to the photodiode is used to etch away the base material in a self-aligned manner around the p-contact metal. The isolation etch is used to define the total area of the photodiode, this defines the perimeter of the circle which bounds the n-contact metal. Polyimide is used to planarize the sidewalls exposed during the isolation etch, and the polyimide etch mask opens windows allowing contact to be made to the p-contact. The metal 1 layer is used to increase the thickness of the p-contact metal, as the base metal is quite thin and therefore somewhat resistive. Silicon nitride and BCB are deposited over the top of this structure so far and serve no useful purpose. The substrate removal etch exposes the n+ subcollector layer, to which the ohmic ring contact is deposited as collector metal. The photodetector should be protected with photoresist during the collector recess etch, as the n+ layer is needed for electron conduction.

A cross section of the completed photodiode is shown in fig. 7.2. The quantum efficiency is enhanced if there is substantial reflection from the base metal forcing the unabsorbed light to make one or more additional passes through the InGaAs material. For the best reflectivity, this metal would be gold, but this is not possible as the base contact metal that must undergo heat treatment in order to reduce the contact resistance. The titanium is still somewhat reflective to 1.55 μm radiation, although somewhat less than gold. From the cross-section it is evident that e-beam evaporation of dielectric antireflection coatings are possible. Unless the same material is useful for antireflection coating and for passivation of the exposed collector sidewall, e-beam evaporation and subsequent liftoff must be used. In this way, something of a resonant cavity can be constructed to augment the low

quantum efficiency that results from such a thin absorbing region.

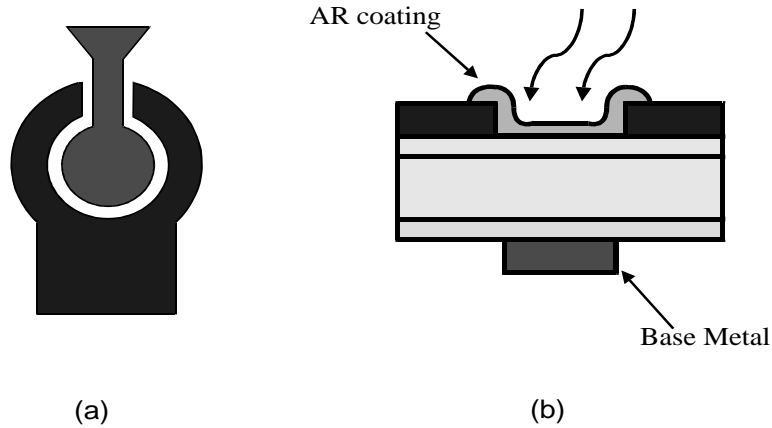


Figure 7.2: Photodiode: (a) Top View (b) Cross section

Unless the p and n regions of the photodiode are of widegap and hence non-absorbing material, some of the light is absorbed in these regions. This is undesirable as the electron and hole created when the photon is absorbed must diffuse through the heavily doped material, an inherently slow process. This is not expected to limit the high-speed response in this photodiode, however. The worst-case scenarios for diffusion are for electrons generated in the base and holes generated in the subcollector, as these carriers would have to diffuse out of the regions in which they were generated before they began their drift across the full extent of the depletion region. In a standard p-i-n homojunction photodiode, the result is a long, slowly decaying "tail" of photocurrent that persists far longer than the transit time of carriers through the depletion region. For this photodiode, this is not expected to occur. The bandgap in the base is graded such that the absorption coefficient decreases from the depletion region to the base contact. The base is also very thin, so little absorption is expected to occur in the base. Of more major concern is the generation of holes in the subcollector. The hole velocity is roughly 5 times slower than the electron velocity, and the hole transit time will dominate. Holes generated in the subcollector also do

not have the benefit of a built-in electric field.

It is the heavy doping in the subcollector that will solve this problem. The Burstein-Moss effect has been experimentally observed in heavily n-doped InGaAs [31]. This effect is the result of the large curvature of the conduction band in InGaAs. The result of this curvature is a low effective mass, more importantly here is the relative scarcity of available electron states near the bandedge due to the high curvature. As the n-doping increases, the energy at which significant numbers of electron states are available increases rapidly. The result is a much wider effective bandgap for the absorption of light with substantial n-doping. For $1 \times 10^{19}/\text{cm}^3$ n-doped InGaAs it has been shown that the effective bandgap increases enough that the absorption coefficient of $1.55 \mu\text{m}$ radiation falls by four orders of magnitude.

The 3dB bandwidth of the photodiode can be estimated. The diameter of the photodiode active region is assumed to be $10 \mu\text{m}$, small for high speed but still of reasonable diameter to couple a fiber into. The p-contact resistivity is 0.7Ω , based on TLM measurements of base contact resistance. The p-series resistance is 0.013Ω , and so will be neglected. The n-contact resistance is 1Ω , and the n-series resistance is 10Ω . The parasitic resistance of the photodiode is about 12Ω . The capacitance can be estimated by $C = \epsilon \frac{A}{d}$ to be 50 fF. The pole frequency corresponding to this RC product is 265 GHz. With the pessimistic assumption that the hole current completely dominates, the transit time bandwidth is ~ 100 GHz assuming a hole velocity of 5×10^6 cm/s in the 2000 \AA collector drift region. The 3dB bandwidth of the photodiode is therefore nearly 100 GHz, and so is not a limiting factor in the bandwidth of this receiver.

Assuming a -9 dBm input signal and a rather pessimistic 10 % quantum efficiency, the photocurrent generated can be calculated with $I = q\eta P/h\nu$ to be $15 \mu\text{A}$. This input current and the drive requirements for the flip-flop determine the maximum transimpedance gain required from the amplification stages. Similarly, the maximum expected input power can be converted to a photocurrent and the minimum transimpedance gain from the amplifiers can be determined.

7.2 Transimpedance Amplifier

The transimpedance amplifier (TIA) was basically designed to achieve as much gain as possible with a 40 GHz 3dB bandwidth. Furthermore, the circuit topology used was that of the Darlington feedback amplifier, as this circuit had already been designed and characterized in this work. This was a naive approach to the TIA design, as the possibility that there would be too much gain was not considered. If the TIA has too much gain, the variable gain amplifier may not have sufficiently low gain (or high attenuation) so as not to saturate the input of the decision circuit. Further complications to the TIA design are considerations of receiver sensitivity (noise performance), the high impedance load at the input which in no way approximates 50Ω , and compatibility of the bias conditions at the output of the TIA which is also the input of the variable gain amplifier.

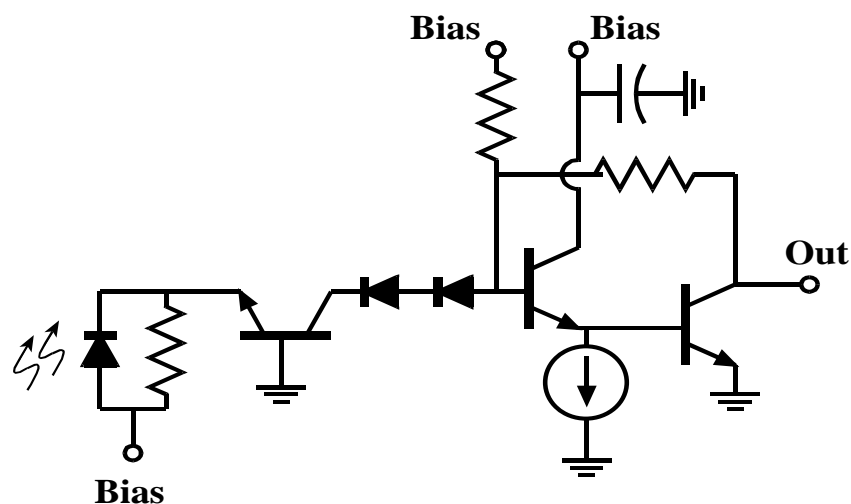


Figure 7.3: Circuit Diagram of the Transimpedance Amplifier

The circuit diagram of the TIA is shown in fig. 7.3. The Darlington feedback amplifier can be recognized in the output of the TIA. A common-base stage is used to buffer the photodiode from the feedback amplifier due to the low pole frequency resulting from the photodiode capacitance and the high input impedance of the feedback amplifier. The common base

stage itself presented some design tradeoffs. A large transistor is desirable in order to realize a high gain, beneficial from the standpoint of noise in the receiver. But a large transistor would require large currents to flow through small resistors in order to provide bias current to the emitter and collector nodes of the transistor. These resistors would introduce a large component of thermal noise into the front end. A smaller transistor not only provides less gain, but also has increased resistive parasitics that decrease the bandwidth when coupled with the photodiode capacitance.

The common-base transistor in use is sized such that both the thermal resistance and the shot noise contributions to the noise are comparable. The feedback amplifier block that follows the common base stage has been designed to give the largest gain possible over 40 GHz bandwidth, but with less attention to noise performance. Here it is important to reduce the output return loss, S_{22} , in order to provide a better impedance match to the 50 Ω input to the variable gain amplifier.

A chip photomicrograph is shown in fig. 7.5 for a nearly completed TIA. The photodetector is visible adjacent to the signal pad on the left, followed by a 250 μm long tuning inductor (80 pH design). To the right is the common base configured transistor, a level shifting diode, then the Darlington feedback amplifier output.

The S_{22} and transimpedance gain are plotted as functions of frequency in fig. 7.5. The transimpedance gain at DC is 54.3 db- Ω , which is about 518 Ω . For a 15 μA input current, this corresponds to an 8mV output voltage amplitude. The gain peaking is only 0.3 dB, and the 3dB- Ω bandwidth is 43 GHz. Static power dissipation is 50 mW, an extremely low value for an amplifier with this gain and bandwidth.

7.3 Variable Gain Amplifier

The variable gain amplifier (VGA) consists of a gain-control stage and four amplification stages. The four amplification stages in the VGA are simply differential amplifiers that increase the overall gain of the circuit. The VGA has been designed to allow control of the gain by varying the DC voltage on one of the bias pads. The automatic gain control (AGC) feature has not been designed.

Three methods of implementing the gain control are standard [32]. These have been classified as variable feedback, variable transconductance,

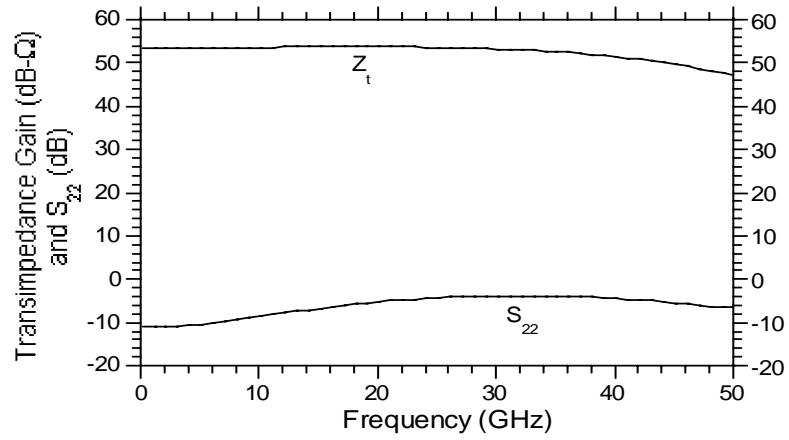


Figure 7.4: Simulated results of transimpedance amplifier

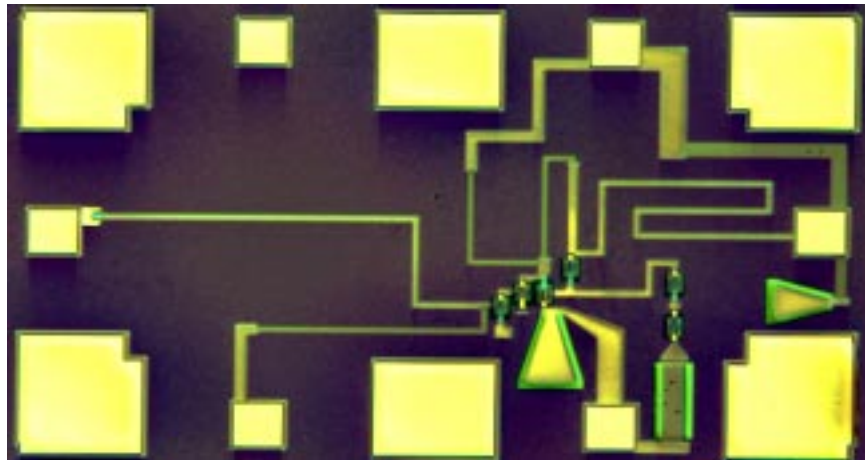


Figure 7.5: Chip photograph of transimpedance amplifier

and current steering. The variable feedback method is one in which a feedback resistance is varied in order to vary the gain. This feedback resistance may be a series resistance in the emitter lead or feedback between the collector and base. This method is often used in hybrid circuits in which a FET or p-i-n diode can be used as the variable resistor. Reports from the literature indicate that it is difficult to achieve more than 15 dB of gain variation using this method and still maintain gain flatness over high bandwidth. Also, a monolithic implementation presents difficulties.

The variable transconductance method of achieving gain control involves modulation of the bias current of the transistors. This modulates the transconductance of the transistors and therefore modulates the gain of the stage. The first attempt at designing the VGA for this receiver was done in this method. Differential amplifiers were designed and simulated with provision for modulation of the base voltage on the current mirror transistors. A two-stage differential amplifier, with preamplifier and variable gain stage, resulted in a maximum gain of 17 dB with 50 GHz gain-bandwidth and gain peaking of about 1 dB. The gain was tunable down to about 12 dB with the same bandwidth and gain peaking. As the gain was further reduced, the bandwidth started to decrease.

The ability to vary the gain over a 5-dB range (less than 2) is inadequate to provide enough dynamic range for a useful receiver. A possible fix is to cascade gain control stages to increase the dynamic range. The only complication here is the distribution of the gain variation control to multiple current mirror transistors. Precise matching in the current mirror transistors is necessary in order to provide equal drive to each gain stage from one control voltage source. Due to normal process variations and the fact that these current mirror transistors will be separated from each other by hundreds of microns, achieving dynamic range in this way is not viable.

The current steering approach uses the control voltage to provide or remove signal current from the load resistors. A detail of the gain control stage is shown in fig. 7.6. The Gilbert cell design can be recognized in stage A, the bases of the inner transistors of the differential pair are wired together and connected to the control voltage. The bases of the outer transistors are wired together and connected to a reference voltage. The gain is varied by variation of the control voltage on the inner transistors, as these transistors turn on, bias current is "steered" from the outer transistors to these inner transistors, reducing the gain. In this way, large variations in the gain are

possible. The stage labeled "B" is there to maintain a constant bias voltage on the output of the gain control stage. As the current through the load resistors in stage A is modulated, the bias voltages at the outputs of stage A change. Stage B is essentially the inverse of stage A, as the outer transistors in stage A are turned off, the outer transistors in stage B are turned on, and these transistors sink current from the same load resistor. In this way, the bias voltage at the outputs of stage A is held constant and does not adversely affect the performance of the subsequent gain stage.

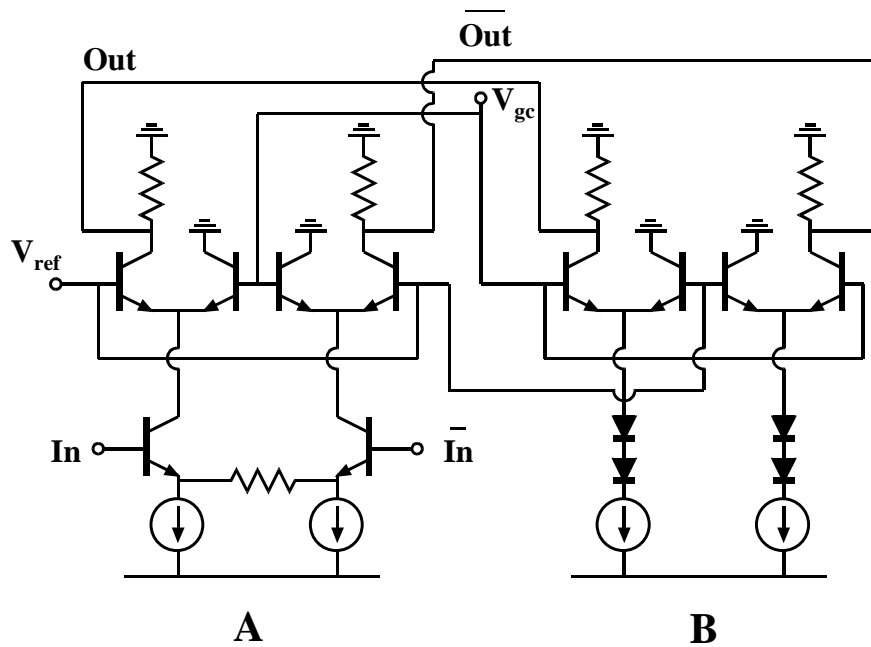


Figure 7.6: Gain control stage with in, in, out, out, gain control voltage (V_{gc}) and reference voltage (V_{ref}) labeled.

The simulated results at the maximum and minimum useful gains of the VGA are shown in fig. 7.7 for four different control voltages. This is for the VGA driven single-ended; the other input is connected to a bias voltage. The gain is variable from +36 dB to -4 dB, a dynamic range of 40 dB, or 100:1. The return losses S_{11} and S_{22} are below -10 dB over the entire frequency range and over the whole range of gain. The gain is also

very flat; implying this is a useful building block for the receiver.

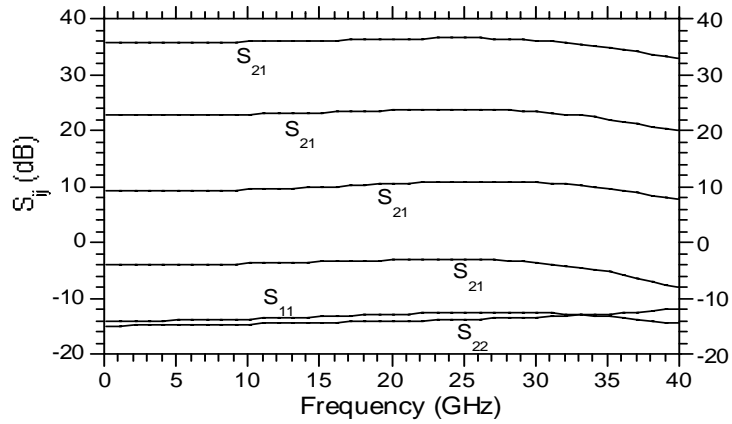


Figure 7.7: Simulated results of variable gain amplifier

A very important point not considered in this design relates to the fact that the output stage of this chip should be able to deliver a 600mV peak to peak signal. The VGA has been simulated with the hybrid-pi model of the transistor, valid only for small-signal operation. A solution is to use output buffer stages that provide gain as well as tolerance of larger voltage excursions.

A chip photo of the VGA is shown in fig. 7.8. The transistor count is 85, and power dissipation is estimated to be 880 mW into a -4.0 V supply voltage.

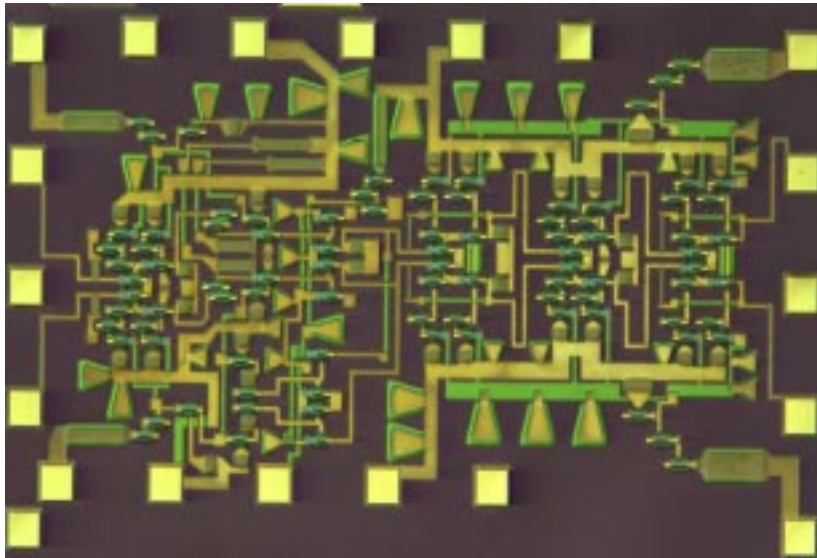


Figure 7.8: Chip photograph of variable gain amplifier

Chapter 8

Conclusion

8.1 Achievements

In this work, transferred-substrate InAlAs/InGaAs heterojunction bipolar transistors have been grown, fabricated, tested and characterized. Analog amplifiers have been built in order to realize some of the potential promised by the RF gains. The MBE growth of these transistors has been refined and standardized in the hopes of achieving reproducibility in the layer structure. Bandgap engineering has been undertaken in order to improve the DC and RF characteristics of the devices. The transferred-substrate process allows lithographic definition of emitter and collector, as well the process provides NiCr resistors, MIM capacitors, microstrip wiring and thermal heatsinking.

Discrete devices have been fabricated with record f_τ of 275 GHz and f_{max} of 264 GHz. The highest f_τ demonstrated in transferred-substrate HBTs prior to this work was 140 GHz. The improvement in f_τ has been achieved through thinning the semiconductor layers, attention to layout parasitics, the use of base bandgap grading, and improvement in the Kirk Effect threshold current density.

Single-stage amplifiers have been fabricated in this work that demonstrate record performance. A gain-bandwidth product of over 400 GHz has been obtained with a Darlington feedback amplifier, with a DC gain of 18.1 dB, a 3dB bandwidth over 50 GHz, and very flat gain-frequency characteristics. This is a record gain-bandwidth product for a single-stage HBT amplifier, either distributed or lumped. Another resistive feedback amplifier in the mirror-doubler configuration demonstrates record 3dB bandwidth for

an HBT amplifier of 85 GHz. The power consumption of these amplifiers is in the tens of milliwatts.

For the technology development aspect of this work, a gaseous carbon doping source has been fabricated, installed, and characterized in collaboration with Ryan Naone and Eric Hegblom at UCSB and under the supervision of the UCSB MBE lab supervisor, John English. The carbon doping source provides the ability to dope GaAs and $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ well into the $1 \times 10^{20} / \text{cm}^3$ range, and alleviates the fear of p-type dopant diffusion. The surface morphology of calibration samples grown even at these high doping levels is excellent, and the hole mobility for a given doping level is as good or better than that provided by the beryllium dopant.

The MBE growth process has been refined and standardized in an effort to create a reliable and reproducible supply of material for the group. A baseline layer structure has been developed that can yield f_τ over 200 GHz routinely and in highly scaled devices has yielded values of f_{max} over 1 THz. Another layer structure has been developed utilizing an Ohmic subcollector that finds application in high-speed, low power logic such as current mode logic (CML). The f_τ and f_{max} obtained with this layer structure and conservative device geometries are both normally above 200 GHz. The real advantage is the improved performance at low V_{ce} due to the thinner and Ohmic collector used in this layer structure.

Near the beginning of this work, Professor Rodwell's group was just beginning to undertake fabrication of demonstration ICs (such as resistive feedback amplifiers). The fabrication of simple ICs was done as part of an evolving effort to transform the transferred-substrate process from one that could demonstrate scalable discrete HBTs into one that proposed a novel solution to wiring and thermal issues in large scale ICs. The intention was to use the high f_{max} of the transferred-substrate HBT to fabricate distributed or tuned circuits which can benefit from a device with a high $f_{max}:f_\tau$ ratio.

The high f_τ now available from the transferred-substrate HBTs has opened up a number of other applications for this process. Probably the most important consequence of achieving high f_τ has been the ability to fabricate fast digital ICs. This in turn has generated demand for higher levels of integration and density on wafer.

But there have been many important achievements: HBTs with the highest f_{max} ; highest f_τ ; the widest bandwidth and highest gain bandwidth HBT analog amplifiers ; and the highest frequency digital logic benchmark,

the static frequency divider.

8.2 Future Work

Upon reflection of how to discuss future work, it has been interesting to review the dissertations of past group members who have participated in this transferred-substrate HBT project. Many of the goals mentioned in those dissertations have been achieved, this is especially inspiring considering how far-fetched those ideas sounded at the time!

The highest f_τ achieved to date is the 275 GHz result presented in this work. The f_{max} at this bias is 244 GHz. It will be useful to examine how to, for example, double both of these figures of merit.

Starting with the equation for forward delay,

$$\tau_b + \tau_c + \frac{kT}{qI_e}(C_{je} + C_{cb}) + R_{ex}C_{cb} + (r_e + R_{ex})C_{la}, \quad (8.1)$$

it is assumed that the f_τ will be doubled by cutting each of the above terms in half. To decrease τ_b by 2, the base thickness should be cut to 0.7 of its original value. To decrease τ_c by 2, the collector thickness should be reduced by 2. The term involving charging the parasitic capacitors will be rewritten for clarity as

$$\frac{kT}{qI_e}(C_{je} + C_{cb}) = \frac{kT}{qJ_e} \left(\frac{\epsilon}{T_e} + \epsilon \frac{W_c L_c T_e}{W_e L_e T_c} \right). \quad (8.2)$$

Here it can be seen that since T_c has gone down by a factor of 2, if C_{cb} is much greater than C_{je} , the current density J_e must go up by a factor of *four*. If instead C_{je} is much greater than C_{cb} the J_e must only increase by a factor of two.

Next, $R_{ex}C_{cb}$ must be decreased by 2:1. Rewriting $R_{ex}C_{cb}$ as

$$R_{ex}C_{cb} = \epsilon \frac{A_c}{A_e} \rho_e, \quad (8.3)$$

it is evident that ρ_e must be reduced by a factor of 4:1.

Finally, at least as far as f_τ is concerned, the term $R_{ex}C_{la}$ must be reduced by 2:1, which can be done with proper attention to the layout.

Now, for f_{max} , the product $R_{bb}C_{cb}$ must be decreased 2:1. Now T_c has been shrunk by 2:1, and T_b has been reduced by 1.4:1. The reduction in T_b

has only a small effect on R_{bb} , as R_{bb} is constituted mainly of Ohmic contact resistance. Neglecting this, the term $R_{bb}C_{cb}$ must be reduced by a factor of 2:1, but C_{cb} has increased by a factor of two from thinning the collector. If we maintain a constant W_c/W_e ratio of 1.1/0.7 as was the case for this device, W_c and W_e must together shrink by a factor of at least four.

The $f_{max} : f_\tau$ ratio of highly scaled transferred-substrate HBTs is currently far too high. The 1.1 THz f_{max} device demonstrated by Q. Lee has an f_τ of about 200 GHz [28]. High-speed digital logic requires high f_τ . The digital circuitry that has been demonstrated to date is likely limited by f_τ and by other factors such as wiring capacitance, transmission line effects, and underdamped transient responses. Further improvement in the f_{max} of transferred-substrate HBTs would be an academic exercise at this point, instead, efforts should be made to improve the f_τ (but not completely at the expense of f_{max}) and careful attention to circuit design and wire routing during layout.

The first step towards increasing f_τ further or even guaranteeing reproducibility of the current peak values is to reduce R_{ex} . The time constants involving R_{ex} account for 16 % of the forward delay in the 275 GHz result. It is the R_{ex} that prevents high f_τ from being achieved on highly scaled devices, not only in multiplying the collector base capacitance, but those parasitics capacitances introduced by the layout. The transferred-substrate HBT is not yet a scalable device, f_{max} indeed scales, but R_{ex} prevents the scaling of f_τ with device area. So an effort should be made to investigate the cause and nature of the R_{ex} problem at a deeper level than was carried out in this work through the rigorous use of test structures and extensive SEM microscopy of the emitter semiconductor and metallization after exposure to various steps in the process.

The importance of this R_{ex} problem can not be overstated. If this problem can be solved or at least reduced, then there really will no longer be any reason not to build ICs of deep submicron devices. Circuit bandwidths will increase dramatically beyond their current record levels, and transferred-substrate HBTs will be scalable in the complete sense of the word. To do this may require investment in new lithographic techniques, either an optical stepper that can produce emitter stripes narrower than 0.5 μm , or E-beam lithography may need to be employed.

The long sought-after carbon source is here and it works! Carbon doping and more heavily doped bases can now become a reality. A new parameter

space will need to be explored regarding the effects of increased doping on τ_b and R_{bb} . With a lower R_{bb} , somewhat higher C_{cb} can be afforded with the same resultant f_{max} . This will allow more flexibility in the design of tuned and distributed amplifiers that may benefit more from a reduction in R_{bb} than in C_{cb} . The lower $R_{bb} : C_{cb}$ ratio can also provide an increased Kirk threshold for the same f_{max} .

The transferred-substrate HBTs here reported obtain high f_τ in part because of thin base and collector layers. because the collector layer is InGaAs, breakdown voltage is very low and the devices are unsuitable for microwave power amplifiers. A solid phosphorous source has recently been installed in the MBE system at UCSB. This will allow fabrication of InP-InGaAs-InP double-heterostructure HBTs. The InP collector will provide much higher breakdown voltage than obtained in InGaAs and will lead to realization of microwave power devices.

Watts of power at tens or even hundreds of GHz may become a reality with a properly designed phosphorous collector. The use of InP in the collector will also facilitate the outflow of heat from the collector semiconductor to the collector contact metal, due to the much higher thermal conductivity of InP than InGaAs. As a fringe benefit, material quality should improve as the native oxide on the InP substrate can be desorbed under a phosphorous beam and nucleation of the buffer layer will be with a lattice matched, binary compound. The fact that a substantial fraction of the transistor epi will be grown with the lattice matched binary will also improve the tolerance for lattice matching of the ternary InGaAs and InAlAs alloys. If substantial progress is made toward the reduction of R_{ex} , ultra high f_τ will be possible due to the reduction in the R_{ex} and the scaling of the device that will naturally follow this and also that the collector can be made much thinner to reduce the transit time and a reasonable BV_{ceo} will be maintained.

More careful attention must be paid to the layout parasitics discussed in chapter 5, especially for the scaled devices. The capacitance introduced by the M1 and M2 crossovers over the base metal can be reduced significantly resulting in small gains in f_τ for relatively large devices, and larger gains in f_τ for the smaller devices.

A new generation of resistive feedback amplifiers can be fabricated with more attention paid to reducing the return losses. An idea for reducing S_{11} is to introduce another transistor to precede what is normally the input

transistor. This transistor would basically be a low gain, high bandwidth stage with good return loss characteristics. In this way, the S_{11} of the overall circuit could be greatly improved, the gain may or may not be slightly improved, and ideally the bandwidth will be nearly unaffected. Introduction of this new input stage will allow for more freedom in tuning the output to reduce return loss, if the new input stage has any gain, a smaller output resistor can be used to shunt the output and reduce return loss. Introduction of this transistor would prevent this amplifier from being a single stage amplifier, but the improvement in return loss will make it useful in more applications as is.

Appendix A

Process Flow

A.1 Focus check with 2" Si wafer

A Solvent Cleaning

1. Check the resistivity of the D.I. water. It should be > 17 M Ω .
2. Set spinner for 4 krpm for 2 minutes.
3. Get 2" Si wafer from drybox (in old toluene petri dish)
4. Start wafer spinning, then squirt with acetone.
5. Don't let it dry, squirt with ISO, then ACE and ISO again.
6. Put on 200°C hot plate for 5 minutes.

B Photoresist Application and Exposure

1. Cool down after dehydration, 5 min
2. Wafer on spinner chuck with vacuum, blow with N₂
3. Apply SPR 950-0.8 with syringe and filter to cover wafer
4. Spin at 2.5 krpm for 30 sec
5. Soft Bake, 90°C, 1 min. on hot plate
6. Apply CEM with syringe and filter to cover wafer
7. Spin at 4 krpm for 30 sec
8. Use "Smartset Array" mask plate, run job as follows: "FOCUS SFOC2\FOC", note down the best focus number returned by computer.

Align emitters perpendicular to the major flat of the wafer

A.2 Emitter Contacts (Mask Layer 1)

A Solvent Cleaning

1. Check the resistivity of the D.I. water. It should be $> 17 \text{ M}\Omega$.
2. Cold ACE 5 min.
3. Hot METH 5 min.
4. Hot ISO 5 min.
5. Running DI 5 min.
6. Blow dry with N₂
7. Dehydration bake, 120°C, 30 min in petri dish without cover

B Photoresist Application and Exposure

1. Cool down after dehydration, 5 min
2. Wafer on spinner chuck with vacuum, blow with N₂
3. Apply SPR 950-0.8 with syringe and filter to cover wafer
4. Spin at 2.5 krpm for 30 sec
5. Soft Bake, 90°C, 1 min. on hot plate
6. Apply CEM with syringe and filter to cover wafer
7. Spin at 4 krpm for 30 sec
8. Expose for 2.4 seconds, set focus to Si wafer focus + 40 - system focus.

C Development

1. Post Bake, 100°C, 2 min, 10 sec
2. Rinse in running D.I., 30 sec.
3. Develop in MF-701 for 2 min, 20 sec
4. Rinse in running D.I., 3 min

D Oxygen Plasma Descum of Photoresist

1. 300mT of O₂
2. power = 100W at low frequency.
3. run for 15 seconds

E Evaporation

1. Mix a dilute solution of HCl : H₂O :: 1 : 10

2. Dip in dilute HCl for 15 sec.
3. Rinse in DI for 3 min.
4. Blow dry with N₂
5. Place wafer on E-Beam mount. Level it.
6. Put in a new crystal. A new crystal should always be used for emitters.
7. Pump down to below 1×10^{-6} torr
8. Deposit material:

Material	Thickness(\AA)	Dep. Rate ($\text{\AA}/\text{sec}$)	Approx. Vernier
Ti	200	1-2	1.65
Pt	500	1	1.85
Au	8000	15	1.5
Si	500	2-3	1.6

(use the small crucible for Au - be sure to load the right amount of charge. If charge is less Au might run out before the desired thickness is reached and if it is more then it might spill out of the crucible while melting)

F Liftoff -->> DO NOT LET ACE DRY ON WAFER! <<--

1. Soak wafer in beaker of ACE until metal comes loose.
2. If the liftoff is stubborn, leave the wafer soaking in ACE overnight. Because ACE evaporates quickly, seal the top of the beaker with foil.
3. ACE squirt bottle. Do not squirt except at edges of wafer and only with wafer immersed in acetone. Do not squirt hard.
4. Rinse with METH then ISO (squirt bottle)
5. Rinse in running DI water for 3 min.
6. Blow dry with N₂.
7. Check under microscope, then Dektak thickness of metal.

A.3 Base Contact etch (no mask required)

A Oxygen Plasma Descum

1. 300mT of O₂
2. power = 100W at low frequency.
3. run for 15 seconds

B Surface Prep

1. Mix a dilute solution of NH₄OH : H₂O :: 1 : 10
2. Dip in dilute NH₄OH for 10 sec.
3. Rinse in DI for 3 min.
4. Blow dry with N₂.

C Dry Etch

1. Clean both chucks and shield with ISO. Turn laser on now.
2. O₂ plasma clean (20sccm @ 125 mT, 500V bias for 20 min)
3. Precoat chamber (M/H/A 4/20/10 sccm @ 125 mT, 300V bias for 20 min)
4. Load wafer and align laser monitor, then pump down to low E-6 torr.
5. Etch (M/H/A 4/20/10 sccm @ 75 mT, 500V bias) until slope changes.
6. Increase pressure to 110mT. Etch until desired stop point reached.
7. Remove polymer (O₂ 50 sccm @ 125 mT, 200V bias for a quarter of etch time).
8. Vent chamber and remove sample.

D Si removal

1. 300mT of CF₄
2. power = 100W at low frequency.
3. run for 1 minute per 500A Si.
4. Check that Si has been removed with optical microscope.
5. Remove samples, clean chamber for 10 minutes with O₂ (300W, 300mT)
6. Replace samples and do standard descum (O₂ @ 100W, 300mT for 15 sec)

E Selective etch

1. Mix a dilute solution of $\text{NH}_4\text{OH} : \text{H}_2\text{O} :: 1 : 10$
2. Dip in dilute NH_4OH for 10 sec.
3. Rinse in DI for 3 min.
4. Blow dry with N_2 .
5. Mix Solution B - $\text{HCL}:\text{H}_2\text{O} :: 4:1$ in HCL beaker (plastic) with Soln B.
6. Mix Solution A - $\text{HBr}:\text{glacial acetic} :: 1:1$ in Soln A beaker with stirrer bar. Mix well.
7. Mix Solution A & B 1:1 by adding Soln B then Soln A to A+B beaker. Mix well
8. Put the A+B beaker in the ice, wait till 10°C .
9. NO stirring, etch for 22 seconds (in water after 22 seconds).
10. Rinse in DI for 3 min.
11. Blow dry with N_2 .

F Nonselective etch

1. Mix etchant as follows: 55ml of 1M citric acid in 220ml DI. Mix well. Add 5ml peroxide. Mix well. Add 1ml phosphoric acid. Mix well.
2. Stirring at 200 rpm etch for 25 seconds by suspending the wafer in a basket.
3. Rinse in DI for 3 min.
4. Blow dry with N_2 .

A.4 Base Contact (Mask Layer 2)

A Dehydration bake, 120°C , 30 min. in petri dish without cover

B Photoresist Application and Exposure

1. Cool down after dehydration, 5 min.
2. Wafer on spinner chuck with vacuum, blow with N_2
3. Apply EIR 5214 with syringe and filter to cover wafer
4. Spin at 6 krpm for 30 sec.
5. Hot Plate Bake, 95°C , 1 min.
6. Expose for 0.32 sec., focus of 24 - system focus

7. Hot Plate Bake, 105°C, 1min.
8. Flood Expose at 7.5 mW/cm² for 30 sec.

C Development

1. AZ 400K : DI :: 1:5.5.
2. Develop for 40 sec.
3. Rinse in running DI water for 3 min.
4. Blow dry with N₂.
5. Observe under microscope. If there is a suspicion that some photoresist is remaining may go back into developer for 5 - 10 secs longer.

D Oxygen Plasma Descum of Photoresist

1. 300mT of O₂
2. power = 100W at low frequency.
3. run for 15 seconds

E Evaporation

1. Mix a dilute solution of HCl : H₂O :: 1 : 10
2. Dip in dilute HCl for 15 sec.
3. Rinse in DI for 3 min.
4. Blow dry with N₂
5. Place wafer on E-Beam mount. Level it.
6. Put in a new crystal. A new crystal should always be used for emitters.
7. Pump down to below 1 x 10⁻⁶ torr
8. Deposit material:

Material	Thickness(Å)	Dep. Rate (Å/sec)	Approx. Vernier
Ti	200	1-2	1.65
Pt	500	1	1.85
Au	800	15	1.5

F Liftoff --> DO NOT LET ACE DRY ON WAFER! <<--

1. Soak wafer in beaker of ACE until metal comes loose.

2. If the liftoff is stubborn, leave the wafer soaking in ACE overnight. Because ACE evaporates quickly, seal the top of the beaker with foil.
3. ACE squirt bottle. Do not squirt except at edges of wafer and only with wafer immersed in acetone. Do not squirt hard.
4. Rinse with METH then ISO (squirt bottle)
5. Rinse in running DI water for 3 min.
6. Blow dry with N₂.
7. Check under microscope, then Dektak thickness of metal.

G RTA for 60 seconds at 300°C. Run a test program first.

A.5 Isolation (Mask Layer 3)

A Solvent Cleaning

1. Check the resistivity of the D.I. water. It should be $\leq 17\text{MW}$.
2. Cold ACE 3 min.
3. Hot METH 3 min.
4. Hot ISO 3 min.
5. Running DI 3 min.
6. Blow dry with N₂
7. Dehydration bake, 120°C, 30 min. in petri dish without cover

B Photoresist Application and Exposure

1. Cool down after dehydration, 5 min.
2. Wafer on spinner chuck with vacuum, blow with N₂
3. Apply SPR 518-A with syringe and filter to cover wafer
4. Spin at 4.0 krpm for 30 sec.
5. Hot Plate Bake, 90°C, 1 minute
6. Expose for 0.8 sec, focus of 20 - system focus
7. Hot Plate Bake, 110°C, 1 minute

C Development

1. Develop in full beaker of MF-701 for 1 min, 30 seconds

2. Rinse in running DI water for 3 min.
3. Blow dry with N₂.
4. Observe under microscope. If there is a suspicion that some photoresist is remaining may go back into developer for 5 - 10 secs longer.

D Oxygen Plasma Descum of Photoresist and Hardbake.

1. 300mT of O₂
2. power = 100W at low frequency.
3. run for 20 seconds.
4. Hardbake photoresist at 120C for 20 min.

E Isolation Etch

1. Load Si wafer to be used in RIE #5
2. Run LeeGAN or GaNCLN
3. Set up laser monitor, try to center laser on Si wafer
4. Etch conditions:
 - (a) Cl₂ flow rate = 20.0 sccm
 - (b) chamber pressure = 5.0 mTorr
 - (c) P=150W (this is controlled)
 - (d) Voltage should be 150-175 V
5. Etch until 30 seconds after slope change. (6 min., 30 sec. For 3 kA collector).

F Resist strip

1. Remove photoresist with acetone in liftoff beaker for 3 minutes.
2. Rinse by spraying with METH and ISO.
3. Rinse in running DI water for 3 min.
4. Blow dry with N₂.

G Oxygen Plasma Photoresist Removal (optional, at your discretion)

1. 300mT of O₂
2. power = 100W at low frequency.
3. run for 30 seconds.

4. examine under microscope to ensure photoresist removal.

H Characterization of process so far.

1. Dektak etch depth, measure TLMs.

A.6 Poly Planarize (Mask Layer 4)

A Solvent Cleaning

1. Get Poly out of refrigerator, warm up!!
2. Check the resistivity of the D.I. water. It should be $> 17M\Omega$.
3. Cold ACE 5 min.
4. Hot METH 5 min.
5. Hot ISO 5 min.
6. Running DI 5 min.
7. Blow dry with N₂
8. Dehydration bake, 120°C, 30 min. in petri dish without cover

B Poly Spin & Cure

1. Mix adhesion promoter in designated beaker (1 mL of VM-651 from bottle using plastic dropper, 200 mL of D.I.). Stir then get new dropper.
2. Wafer on spinner chuck with vacuum, blow with N₂.
3. Apply adhesion promoter to cover wafer.
4. Let it sit on the wafer for 20 seconds.
5. Spin at 3.0 krpm for 60 sec.
6. Hot plate bake, 125°C, 3 minutes.
7. Apply DuPont Ployimide to cover wafer with syringe and 1 μ m filter.
8. Let sit on wafer for 20 seconds.
9. Spin at 2.5 krpm for 30 sec (to give 1.8 μ m film).
10. Hard bake polyimide in programmable oven as follows in petri dish without cover, and with the base of the petri dish covered with aluminum foil.
 - (a) hold at 90°C for 60 min.

- (b) ramp at 4°C per min. to 150°C.
- (c) hold at 150°C for 60 min.
- (d) ramp at 4°C per min. to 230°C.
- (e) hold at 230°C for 60 min.
- (f) ramp at 4°C per min. to 170°C.

C Photoresist Application and Exposure

1. Remove from oven when below 90°C.
2. Wafer on spinner chuck with vacuum, blow with N₂
3. Apply AZ 4330 with syringe and filter to cover wafer
4. Spin at 5.0 krpm for 30 sec.
5. Soft Bake, 90°C, 30 min. in petri dish without cover
6. Hard Bake, 120°C, 30 min. in petri dish without cover

D Polyimide Etchback

1. Load RIE#1 according to instructions
2. Pump down to low E-6.
3. Set up laser monitor
 - (a) Look for diffraction pattern to identify beam. Laser signal should be about 500mV. Use outermost spot of laser beam.
 - (b) Set up chart recorder for 1 hour and ~700mV.
4. Etch conditions:
 - (a) O₂ flow rate 7.0 sccm.
 - (b) chamber pressure = 10 mTorr
 - (c) P=60W (control this watching Heathkit)
 - (d) Voltage should be around 350 V
5. Etch for 18 cycles.

E Check

1. Put in SEM and look at 1um emitter fingers to see if they are clear.
2. If not, etch two minutes in PEII-A at 100W, 300 mT then SEM again.

F Clean Up Remove mung from backside with ISO soaked Q-tip

G Photoresist Application and Exposure

1. Wafer on spinner chuck with vacuum, blow with N₂
2. Apply SPR 518-A with syringe and filter to cover wafer
3. Spin at 3.0 krpm for 30 sec.
4. Hot Plate Bake, 90°C, 1 minute
5. Expose for 1.1 sec, focus of 12 - system focus. Must use global alignment.
6. Hot Plate Bake, 110°C, 1 minute

H Development

1. Develop in MF-701 for 1 min, 30 seconds
2. Rinse in running D.I. for 3 minutes
3. Flood Expose at 7.5 mW/cm² for 2 minutes.
4. Hard Bake, 120°C, 30 min. in petri dish without cover

I Etch Poly Poly etch for 30 minutes, PEII-A, 300 mT O₂, 100 W

J Resist strip

1. Remove resist in MF-701, 2 minutes.
2. Rinse in running D.I. for 3 minutes
3. Blow dry with N₂.
4. Inspect under microscope.

K Reflow Bake Hot plate bake, 250°C, 10 minutes to reflow polyimide and complete the cure.

A.7 NiCr Resistors (Mask Layer 5)

A Solvent Cleaning

1. Check the resistivity of the D.I. water. It should be $> 17\text{M}\Omega$.
2. Cold ACE 5 min.
3. Hot METH 5 min.
4. Hot ISO 5 min.
5. Running DI 5 min.
6. Blow dry with N₂
7. Dehydration bake, 120°C, 30 min. in petri dish without cover

B Photoresist Application and Exposure

1. Cool down after dehydration, 5 min.
2. Wafer on spinner chuck with vacuum, blow with N₂
3. Apply EIR 5214 with syringe and filter to cover wafer
4. Spin at 6 krpm for 30 sec.
5. Hot Plate Bake, 95°C, 1 min.
6. Expose for 0.32 sec, focus of 18 - system focus.
7. Hot Plate Bake, 105°C, 1 min.
8. Flood Expose at 7.5 mW/cm² for 30 sec.

C Development

1. AZ 400K : DI :: 1:5.5.
2. Develop for 40 sec.
3. Rinse in running DI water for 3 min.
4. Blow dry with N₂.
5. Observe under microscope. If there is a suspicion that some photoresist is remaining may go back into developer for 5 - 10 secs longer.

D Oxygen Plasma Descum of Photoresist

1. 300mT of O₂
2. power = 100W at low frequency.
3. run for 15 seconds

E Evaporation

1. Mix a dilute solution of HCl : H₂O :: 1 : 10
2. Dip in dilute HCl for 15 sec.
3. Rinse in DI for 3 min.
4. Blow dry with N₂
5. Place wafer on E-Beam mount. Level it.
6. Put in a new crystal. A new crystal should always be used for emitters.
7. Pump down to below 1 x 10⁻⁶ torr
8. Deposit material:

Material	Thickness(\AA)	Dep. Rate ($\text{\AA}/\text{sec}$)	Approx. Vernier
Si	200	2-3	1.65
NiCr	475	1	1.75

F Liftoff -->> DO NOT LET ACE DRY ON WAFER! <<--

1. Soak wafer in beaker of ACE until metal comes loose.
2. If the liftoff is stubborn, leave the wafer soaking in ACE overnight. Because ACE evaporates quickly, seal the top of the beaker with foil.
3. ACE squirt bottle. Do not squirt except at edges of wafer and only with wafer immersed in acetone. Do not squirt hard.
4. Rinse with METH then ISO (squirt bottle)
5. Rinse in running DI water for 3 min.
6. Blow dry with N2.
7. Check under microscope, then Dektak thickness of metal.

A.8 Pad Metal (Mask Layer 6)

A Solvent Cleaning

1. Check the resistivity of the D.I. water. It should be $> 17\text{M}\Omega$.
2. Cold ACE 5 min.
3. Hot METH 5 min.
4. Hot ISO 5 min.
5. Running DI 5 min.
6. Blow dry with N2
7. Dehydration bake, 120°C , 30 min. in petri dish without cover

B Photoresist Application and Exposure

1. Cool down after dehydration, 5 min.
2. Wafer on spinner chuck with vacuum, blow with N2
3. Apply EIR 5214 with syringe and filter to cover wafer
4. Spin at 4 krpm for 30 sec.

5. Hot Plate Bake, 95°C, 1 min.
6. Expose for 0.42 sec, focus of 18 - system focus.
7. Hot Plate Bake, 105°C, 1 min.
8. Flood Expose at 7.5 mW/cm² for 30 sec.

C Development

1. AZ 400K : DI :: 1:5.5.
2. Develop for 55 sec.
3. Rinse in running DI water for 3 min.
4. Blow dry with N₂.
5. Observe under microscope. If there is a suspicion that some photoresist is remaining may go back into developer for 5 - 10 secs longer.

D Oxygen Plasma Descum of Photoresist

1. 300mT of O₂
2. power = 100W at low frequency.
3. run for 15 seconds

E Evaporation

1. Mix a dilute solution of HCl : H₂O :: 1 : 10
2. Dip in dilute HCl for 15 sec.
3. Rinse in DI for 3 min.
4. Blow dry with N₂
5. Place wafer on E-Beam mount for angle of about 30 and rotation.
6. Put in a new crystal.
7. Pump down to below 1 x 10⁻⁶ torr
8. Deposit material:

Material	Thickness(Å)	Dep. Rate (Å/sec)	Approx. Vernier
Ti	200	2-3	1.65
Au	9500	15	1.5
Ti	100	2-3	1.65

F Liftoff -->> DO NOT LET ACE DRY ON WAFER! <<--

1. Soak wafer in beaker of ACE until metal comes loose.
2. If the liftoff is stubborn, leave the wafer soaking in ACE overnight. Because ACE evaporates quickly, seal the top of the beaker with foil.
3. ACE squirt bottle. Do not squirt except at edges of wafer and only with wafer immersed in acetone. Do not squirt hard.
4. Rinse with METH then ISO (squirt bottle)
5. Rinse in running DI water for 3 min.
6. Blow dry with N₂.
7. Check under microscope, then Dektak thickness of metal.

A.9 SiN Etch (Mask Layer 7)**A Solvent Cleaning**

1. Wipe lower electrode, housing of PECVD with ISO soaked wipes. Careful, HOT!
2. Run 60CLNSiN.
3. Check the resistivity of the D.I. water. It should be $> 17M\Omega$.
4. Cold ACE 3 min.
5. Hot METH 3 min.
6. Hot ISO 3 min.
7. Running DI 3 min.
8. Blow dry with N₂

B SiN Deposition, PR application

1. Load wafer in PECVD, run SiN30. Remember orientation.
2. Unload, rinse with ISO beaker with ISO (of course) for 1 minute.
3. Load wafer in PECVD, rotate 90° from previous orientation, run SiN20.
4. Unload, let wafer cool for 5 minutes.
5. Wafer on spinner chuck with vacuum, blow with N₂
6. Apply SPR 518-A with syringe and filter to cover wafer

7. Spin at 4 krpm for 30 sec.
8. Hot Plate Bake, 90°C, 1 min.
9. Expose for 0.9 sec, focus of 20 - system focus.
10. Hot Plate Bake, 110°C, 1 min.

C Development

1. Develop in full beaker of MF-701 for 1 min. 30 seconds
2. Rinse in running DI water for 3 min.
3. Blow dry with N₂.
4. Observe under microscope. If there is a suspicion that some photoresist is remaining may go back into developer for 5 - 10 secs longer.

D Oxygen Plasma Descum of Photoresist

1. 300mT of O₂
2. power = 100W at low frequency.
3. run for 20 seconds
4. Hardbake photoresist at 120°C for 20 min.

E SF₆, O₂, Ar RIE3

1. Clean system with ISO and clean wipes.
2. O₂ clean of system as follows:
 - (a) O₂ flow rate = 20.0 sccm
 - (b) chamber pressure = 10.0 mTorr
 - (c) Voltage=500 V
 - (d) Minimize reflected power by tuning
3. Load sample, pump down to at least 1 x 10⁻⁵ torr.
4. Etch conditions:
 - (a) SF₆,O₂,Ar flow rate = 5.0, 3.0, 10.0 sccm
 - (b) chamber pressure = 20.0 mTorr
 - (c) Voltage=250 V
 - (d) Minimize reflected power by tuning
 - (e) Etch for 5 minutes
5. Look under microscope. Etch in 1 min. increments if necessary.

F Resist strip

1. Remove photoresist with acetone in liftoff beaker for 3 minutes.
2. Rinse by spraying with METH and ISO.
3. Rinse in running DI water for 3 min.
4. Blow dry with N₂.

G Oxygen Plasma Photoresist Removal (optional, at your discretion)

1. 300mT of O₂
2. power = 100W at low frequency.
3. run for 30 seconds.
4. examine under microscope to ensure photoresist removal.

A.10 Metal 2 (Mask Layer 8)**A Solvent Cleaning**

1. Check the resistivity of the D.I. water. It should be $> 17M\Omega$.
2. Cold ACE 5 min.
3. Hot METH 5 min.
4. Hot ISO 5 min.
5. Running DI 5 min.
6. Blow dry with N₂
7. Dehydration bake, 120°C, 30 min. in petri dish without cover

B Photoresist Application and Exposure

1. Cool down after dehydration, 5 min.
2. Wafer on spinner chuck with vacuum, blow with N₂
3. Apply EIR 5214 with syringe and filter to cover wafer
4. Spin at 4 krpm for 30 sec.
5. Hot Plate Bake, 95°C, 1 min.
6. Expose for 0.42 sec, focus of 18 - system focus.
7. Hot Plate Bake, 105°C, 1 min.
8. Flood Expose at 7.5 mW/cm² for 30 sec.

C Development

1. AZ 400K : DI :: 1:5.5.
2. Develop for 55 sec.
3. Rinse in running DI water for 3 min.
4. Blow dry with N2.
5. Observe under microscope. If there is a suspicion that some photoresist is remaining may go back into developer for 5 - 10 secs longer.

D Oxygen Plasma Descum of Photoresist

1. 300mT of O2
2. power = 100W at low frequency.
3. run for 15 seconds

E Evaporation

1. Mix a dilute solution of HCl : H2O :: 1 : 10
2. Dip in dilute HCl for 15 sec.
3. Rinse in DI for 3 min.
4. Blow dry with N2
5. Place wafer on E-Beam mount for angle of about 30 and rotation.
6. Put in a new crystal.
7. Pump down to below 1×10^{-6} torr
8. Deposit material:

Material	Thickness(\AA)	Dep. Rate ($\text{\AA}/\text{sec}$)	Approx. Vernier
Ti	200	2-3	1.65
Au	9500	15	1.5
Ti	100	2-3	1.65

F Liftoff -->> DO NOT LET ACE DRY ON WAFER! <<--

1. Soak wafer in beaker of ACE until metal comes loose.
2. If the liftoff is stubborn, leave the wafer soaking in ACE overnight. Because ACE evaporates quickly, seal the top of the beaker with foil.

3. ACE squirt bottle. Do not squirt except at edges of wafer and only with wafer immersed in acetone. Do not squirt hard.
4. Rinse with METH then ISO (squirt bottle)
5. Rinse in running DI water for 3 min.
6. Blow dry with N2.
7. Check under microscope, then Dektak thickness of metal.

A.11 BCB (Mask Layer 9)

A Solvent Cleaning

1. Check the resistivity of the D.I. water. It should be $> 17M\Omega$.
2. Cold ACE 3 min.
3. Hot METH 3 min.
4. Hot ISO 3 min.
5. Running DI 3 min.
6. Blow dry with N2
7. Dehydration bake, 120°C , 30 min. in petri dish without cover

B BCB Spin & Cure

1. Turn N2 flow up to maximum (100 scfh) in BlueM oven.
2. Wafer on spinner chuck with vacuum, blow with N2.
3. Apply adhesion promoter AP-8000 with dropper to cover wafer.
4. Spin at 4.0 krpm for 30 sec.
5. Blow dry with N2.
6. Set spinner for 3000 RPM.
7. Apply BCB 57, spin at 3.0 krpm for 30 sec.,.
8. Cure BCB in BlueM oven (Prog 2) in petri dish without cover, and with the base of the petri dish covered with aluminum foil.

C Photoresist Application and Exposure

1. Remove from oven when below 90°C .

2. Cool down after dehydration, 5 min.
3. Wafer on spinner chuck with vacuum, blow with N2
4. Apply EIR 5214 with syringe and filter to cover wafer
5. Spin at 6 krpm for 30 sec.
6. Hot Plate Bake, 95°C, 1 min.
7. Expose for 0.32 sec, focus of 18 - system focus.
8. Hot Plate Bake, 105°C, 1 min.
9. Flood Expose at 7.5 mW/cm² for 30 sec.

D Development

1. AZ 400K : DI :: 1:5.5.
2. Develop for 44 sec.
3. Rinse in running DI water for 3 min.
4. Blow dry with N2.
5. Observe under microscope. If there is a suspicion that some photoresist is remaining may go back into developer for 5 - 10 secs longer.

E Oxygen Plasma Descum of Photoresist

1. 300mT of O₂
2. power = 100W at low frequency.
3. run for 15 seconds

F Evaporation

1. Place wafer on E-Beam mount.
2. Put in a new crystal in the crystal monitor if reading > 10.
3. Pump down to below 1 x 10⁻⁶ torr
4. Deposit material:

Material	Thickness(Å)	Dep. Rate (Å/sec)	Approx. Vernier
Ni	100	1.5	1.5

G Liftoff -->> DO NOT LET ACE DRY ON WAFER! <<--

1. Soak wafer in beaker of ACE until metal comes loose.

2. Prop wafer up in beaker and squirt the whole wafer REAL HARD with acetone. Do this for 10 minutes or so, occasionally dunking the wafer back into the beaker of acetone.
3. ACE squirt bottle (a final rinse).
4. Rinse with METH then ISO (squirt bottle).
5. Rinse in running DI water for 3 min.
6. Blow dry with N₂.
7. Check under microscope.

H BCB etch

1. Clean RIE#3 with ISO and clean wipes.
2. O₂ clean of system as follows:
 - (a) O₂ flow rate = 20.0 sccm
 - (b) chamber pressure = 50.0 mTorr
 - (c) Voltage=500 V
 - (d) Minimize reflected power by tuning
 - (e) Etch for 30 minutes
3. Load sample, pump down to at least 1×10^{-5} torr.
4. Etch conditions:
 - (a) SF₆,O₂ flow rate = 6.0, 10.0 sccm
 - (b) chamber pressure = 50.0 mTorr
 - (c) Voltage=350 V
 - (d) Minimize reflected power by tuning
 - (e) Etch for 11.5 minutes
5. Look under microscope. Small vias may already be clear. Large vias still brownish.
6. Repeat O₂ clean as in (2), for at least 10 minutes.
7. Etch Ni in Ni etchant @ 60C with stirring @ 300rpm for 30 seconds after all Ni has disappeared (total etch time 45 secs - 1 min for 150 A Ni)..
8. Back into RIE#3 for blanket etch.
 - (a) SF₆,O₂ flow rate = 6.0, 10.0 sccm
 - (b) chamber pressure = 50.0 mTorr
 - (c) Voltage=350 V

- (d) Minimize reflected power by tuning
- (e) Etch for 8 minutes

A.12 Flip

A Solvent clean and dehydration bake carrier wafer. Load it in sputter machine with the real wafer. Pump down to 1.5 E-6 torr.

1. Flow 25 sccm Ar, sputter Ti, 10mT, 0.1KW, 5min.
2. Flow 25 sccm Ar, sputter Au, 10mT, 0.2KW, 10 min.
3. Let sources cool for 10 min.
4. Vent and Unload.

B Backside SiN deposition(SIN10). Put the wafer upside down on a clean Si wafer.

C Au plating @ 5mA, 2.5 hours. Make sure no wires can get shorted or contact each other.

D Bonding

1. Get our chuck, clean off with ISO soaked wipes. Switch Pedestal VAC switch on front down to remove or place chuck, then switch it back up.
2. Place InP wafer on upper chuck, GaAs wafer on lower chuck.
3. Hit UP VAC, LO VAC.
4. Set upper and lower temperatures to 220°C.
5. Spread solder on InP wafer.
6. Release LO VAC, get GaAs wafer and place face down on InP wafer.
7. Wet surfaces with In_{0.4}Pb_{0.6} by dragging GaAs wafer around on InP wafer. After some time, check that GaAs and gold plated surfaces are completely covered with In. Then wet surfaces some more.
8. Remove excess InPb with Q-tip, then move to lower chuck.
9. Turn micrometer on side of bonder to ~5.
10. Turn top piece over, hit clamp (watch from side) then riser.

11. Wafer should be separated from top chuck by small gap now. Apply 400 grams or so of pressure by slowly turning micrometer.
12. Set upper and lower chuck temperatures to 100°C.
13. When chucks reach 50°C, hit riser, then clamp, then turn top piece over.
14. Get wafers, clean up flip-chip bonder.

E SiN Removal

1. Remove SiN by immersing in BHF until SiN disappears.
2. 3 minute DI rinse.

F InP Substrate Removal

1. Mix HCl:DI 3:1 in junk beaker.
2. Dip for 25 minutes or so. If there was polyimide on back of wafer, dip for 2-3 minutes, then try to scrape off polyimide. Repeat until polyimide is gone.
3. Mix another solution in InP etch beaker. Finish off the etch in that beaker.
4. Rinse in DI for 3 minutes.

G Clean Up

1. Test some base-emitter diodes.
2. Clean up edges of wafer with blade.
3. Dip for 30 seconds in 3:1 HCl:DI, then 3 minute DI rinse.

A.13 Collector

A No Solvent Clean. Dehydration bake 120°C, 30 minutes.

B Collector Lithography

1. Configure collector job for mirror image, different alignment mark coordinates.
2. Cool down after dehydration, 5 min
3. Wafer on spinner chuck with vacuum, blow with N₂
4. Apply SPR 950-0.8 with syringe and filter to cover wafer
5. Spin at 6.0 krpm for 30 sec

6. Soft Bake, 90°C, 1 min. on hot plate
7. Apply CEM with syringe and filter to cover wafer
8. Spin at 6 krpm for 30 sec
9. Refigure cell size, average step size over 3rd row for X, and over the 4th column for Y.
10. Enter new cell size into job.
11. Focus offset of +54, (change system focus to 20 or so in order to get this)
12. Expo time of 1.66 seconds

C Development

1. Post Bake, 100°C, 2 min, 10 sec
2. Rinse in running D.I., 30 sec.
3. Develop in MF-701 for 1 min, 30 sec
4. Rinse in running D.I., 3 min

D Oxygen Plasma Descum of Photoresist

1. 300mT of O₂
2. power = 100W at low frequency.
3. run for 15 seconds

E Evaporation

1. Mix a dilute solution of HF : H₂O :: 1 : 20
2. Dip in dilute HF for 30 sec.
3. Rinse in DI for 3 min.
4. Blow dry with N₂
5. Place wafer on E-Beam mount. Level it.
6. Put in a new crystal.
7. Pump down to below 1 x 10⁻⁶ torr
8. Deposit material: (use the small crucible for Au - be sure to load the right amount of charge. If charge is less Au might run out before the desired thickness is reached and if it is more then it might spill out of the crucible while melting)

F Liftoff --> DO NOT LET ACE DRY ON WAFER! <<--

1. Soak wafer in beaker of ACE until metal comes loose.

Material	Thickness(\AA)	Dep. Rate ($\text{\AA}/\text{sec}$)	Approx. Vernier
Ti	200	1-2	1.65
Pt	400	1	1.85
Au	4000	15	1.5

2. If the liftoff is stubborn, leave the wafer soaking in ACE overnight. Because ACE evaporates quickly, seal the top of the beaker with foil.
3. ACE squirt bottle. Do not squirt except at edges of wafer and only with wafer immersed in acetone. Do not squirt hard.
4. Rinse with METH then ISO (squirt bottle)
5. Rinse in running DI water for 3 min.
6. Blow dry with N₂.
7. Check under microscope, then Dektak thickness of metal.

Bibliography

- [1] H. Suzuki, K. Watanabe, K. Ishikawa, H. Masuda, K. Ouchi, T. Tanoue, R. Takeyari, “InP/InGaAs HBT ICs for 40 Gbit/s optical transmission systems”, *GaAs IC Symp. Tech. Dig.*, pp. 215-218, 1997.
- [2] Y. Umeda, T. Enoki, T. Otsuji, T. Suemitsu, H. Yokoyama, Y. Ishii, “Ultrahigh-speed IC technologies using InP-based HEMTs for future optical communication systems”, *IEICE Transactions on Electronics*, pp. 409-418, 1999.
- [3] J. Candy, G. Temes, “Oversampling methods for A/D and D/A conversion”, *Oversampling delta-sigma data converters*, pp. 1-25, 1992.
- [4] U. Bhattacharya, M. J. Mondry, G. Hurtz, I. -H. Tan, R. Pullela, M. Reddy, J. Guthrie, M. J. W. Rodwell, J. E. Bowers, “Transferred-substrate Schottky-collector heterojunction bipolar transistors : first results and scaling laws for high f_{max} ”, *IEEE Electron Device Lett.*, vol. 16, pp. 357-359, 1995.
- [5] Q. Lee, S. C. Martin, D. Mensa, R. Pullela, R. P. Smith, B. Agarwal, J. Guthrie, M. Rodwell, “Deep sub-micron transferred-substrate heterojunction bipolar transistors” *Device Research Conference Dig.*, pp. 26-27, 1998.
- [6] T. deLyon, T. Buchan, N. Kirchner, J. Woodall, G. Scilla, F. Cardone, “High carbon doping efficiency of bromomethanes in gas source molecular beam epitaxial growth of GaAs”, *Appl. Phys. Lett.*, vol. 58, no. 5, pp. 517-519, 1991.

- [7] K. Cheng, A. Cho, W. Bonner, "Beryllium doping in $\text{Ga}_{0.47}\text{In}_{0.53}\text{As}$ and $\text{Al}_{0.52}\text{In}_{0.48}\text{As}$ grown by molecular-beam epitaxy", *J. Appl. Phys.*, vol. 52, no. 7, pp. 4672-4675, 1981.
- [8] J. Miller, D. Collins, N. Moll, "Control of Be diffusion in molecular beam epitaxy GaAs", *Appl. Phys. Lett.*, vol. 46, no. 10, pp. 960-962, 1985.
- [9] R. Hamm, M. Panish, R. Nottenburg, Y. Chen, D. Humphrey, "Ultra-high Be doping of $\text{Ga}_{0.47}\text{In}_{0.53}\text{As}$ by low-temperature molecular beam epitaxy", *Appl. Phys. Lett.*, vol. 54, no. 25, pp. 2586-2588, 1989.
- [10] K. Choquette, M. Hong, R. Freund, J. Mannaerts, R. Wetzell, R. Leibenguth, "Vertical-Cavity surface-emitting laser diodes fabricated by in-situ dry etching and molecular beam epitaxial regrowth", *IEEE Photonics technology Letters*, vol. 5, no. 3, pp. 284-287, 1993.
- [11] Y. Amamiya, H. Shimakawi, N. Furuhata, M. Mamada, N. Goto and K. Honjo, "Lateral p^+/p regrown base contacts for AlGaAs/InGaAs HBTs with extremely thin base layers", *Device Research Conf. Tech. Dig.*, pp. 38-39, 1995.
- [12] D. Miller, M. Micovic, D. Lubyshev, W. Cai, W. Hwang, "Iodine and carbon tetrabromide use in solid source molecular beam epitaxy", *J. Vac. Sci. Technol. B*, vol. 16, no. 3, pp. 1361-1366, 1998.
- [13] R. Metzger, M. Hafizi, R. Wilson, W. Stanchina, J. Jensen, L. McCray, "Be diffusion at the emitter-base junction of graded AlInAs / GaInAs heterojunction bipolar transistors", *J. Vac. Sci. Technol. B*, vol. 10, no. 6, pp. 2347-2350, 1992.
- [14] M. Hafizi, R. Metzger, W. Stanchina, "Stability of beryllium-doped compositionally graded and abrupt AlInAs / GaInAs heterojunction bipolar transistors", *Appl. Phys. Lett.*, vol. 63, no. 1, pp. 93-95, 1993.
- [15] M. Higuchi, T. Ishikawa, K. Imanishi, K. Kondo, "Doping characteristics of Si into molecular-beam-epitaxially grown InAlAs layers", *J. Vac. Sci. Technol. B*, vol. 9, no. 6, pp. 2802-2804, 1991.

- [16] T. Fujii, T. Inata, K. Ishii, S. Hiyamizu, "Heavily Si-doped InGaAs lattice-matched to InP grown by MBE", *Electronics Letters*, vol. 22, no. 4, pp. 191-192, 1986.
- [17] R. Malik, N. Chand, J. Nagle, R. W. Ryan, K. Alavi, A. Y. Cho, "Temperature dependence of common-emitter I-V and collector breakdown voltage characteristics in AlGaAs/GaAs and AlInAs/GaInAs HBT's grown by MBE", *IEEE Electron Device Letters*, vol.13, p.557-9, 1992.
- [18] K. Krishnamurthy, R. Vetry, S. Keller, S. Long, M. Rodwell, U. Mishra, "Wide-band GaN Microwave Power Amplifiers", submitted to *IEE Electronics letters*.
- [19] J. Guthrie, D. Mensa, B. Agarwal, Q. Lee, R. Pallela, M. Rodwell, "HBT IC process with copper substrate", *Electronics Letters*, vol.34, IEEE, p.467-8, 1998.
- [20] R. Williams, "Modern GaAs Processing methods", Artech House, pp. 229-238.
- [21] H. Kroemer, "Two integral relations pertaining to the electron transport through a bipolar transistor with a nonuniform energy gap in the base region", *Solid State Electronics*, vol. 28, pp. 1101-1103, 1985.
- [22] Y. Hiraoka, J. Yoshida, "Two-dimensional analysis of the surface recombination effect on current gain of GaAlAs/GaAs HBTs", *IEEE Transactions on Electron Devices*, vol. 35, no. 7, pp. 857-862, 1988.
- [23] U. Bhattacharya, "Transferred-substrate heterojunction bipolar transistors", *Ph.D. Dissertation*, University of California, Santa Barbara, 1996.
- [24] S. Laux, W. Lee, "Collector signal delay in the presence of velocity overshoot", *IEEE Electron Device Letters*, vol. 11, no. 4, pp. 174-176, 1990.
- [25] L. Camnitz, N. Moll, "An analysis of cut-off frequency behavior of microwave heterostructure bipolar transistors", *Compound semiconductor transistors - Physics and technology*, IEEE Press, pp.21-45.
- [26] W. Joyce, R. Dixon, *Applied Physics Letters*, vol. 31, p. 354, 1977.

- [27] T. Ishibashi, Y. Yamauchi, "A possible near-ballistic collection in an AlGaAs/GaAs HBT with a modified collector structure", *IEEE Transactions on Electron Devices*, vol. 35, no. 4, pt. 1, pp. 401-404, 1988.
- [28] Q. Lee, S. C. Martin, D. Mensa, R. P. Smith, J. Guthrie, S. Jaganathan, Y. Betser, T. Mathew, S. Krishnan, L. Samoska, M. Rodwell, "Sub-micron Transferred-Substrate Heterojunction Bipolar Transistors with Greater than 1 THz f_{max} ", *Device Research Conference Dig.*, June 1999.
- [29] D. Mensa, Q. Lee, R. Pullela, B. Agarwal, J. Guthrie, S. Jaganathan, M. Rodwell, "Baseband Amplifiers in Transferred-Substrate HBT Technology", *GaAs IC Symp. Tech. Dig.*, submitted for presentation, 1998.
- [30] B. W. Lai, "Amplification beyond f_t using NON-Inductive Circuit Techniques", *M.S. Dissertation*, University of California, Berkeley, 1982.
- [31] D. Hahn, O. Jaschinski, H. Wehmann, A. Schlachetzki, "Electron concentration dependence of absorption and refraction in n-InGaAs near the band edge", *Journal of Electronic Materials*, vol. 24, no. 10, pp. 1357-1361, 1995.
- [32] K. Kobayashi, K. Ip, A. Oki, D. Umemoto, S. Claxton, M. Pope, J. Wiltz, "GaAs HBT 0.75 - 5 GHz multifunctional microwave - analog variable gain amplifier", *IEEE Journal of Solid - State Circuits*, vol. 29, no. 10, pp. 1257-1261, 1994.