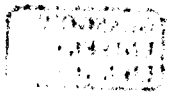
Development of Advanced Technologies for the Fabrication of III-V High Electron Mobility Transistors

Euan James Boyd

Submitted for the degree of
Doctor of Philosophy

Department of Electronics and Electrical Engineering
University of Glasgow
January 2004



Abstract

Over the past 5 years there has been an increase in the number of applications that require devices that operate in the millimetre range (30-300GHz). This demand has driven research into devices that will operate at frequencies above 100GHz. This performance has been achieved using two main technologies, the Heterojunction Bipolar Transistor (HBT) and the High Electron Mobility Transistor (HEMT). At present it is a HEMT device that holds the record for the highest operating frequency of any transistor. It is this technology that this project concentrates on.

In order to fabricate devices that operate at these frequencies two methods are commonly employed. The first is to vary the material of the device, in particular, increasing the indium content of the channel. The second method is to reduce the physical dimensions of the transistors, including reducing the gate length of the device therefore reducing transit time and gate capacitance. Reducing the separation of the source-drain ohmic contacts or employing a self-aligned ohmic strategy reduces the associated parasitic resistances. This project will concentrate on the scaling of the gate length in addition to the reduction of parasitic resistances with the use of self-aligned ohmic contacts.

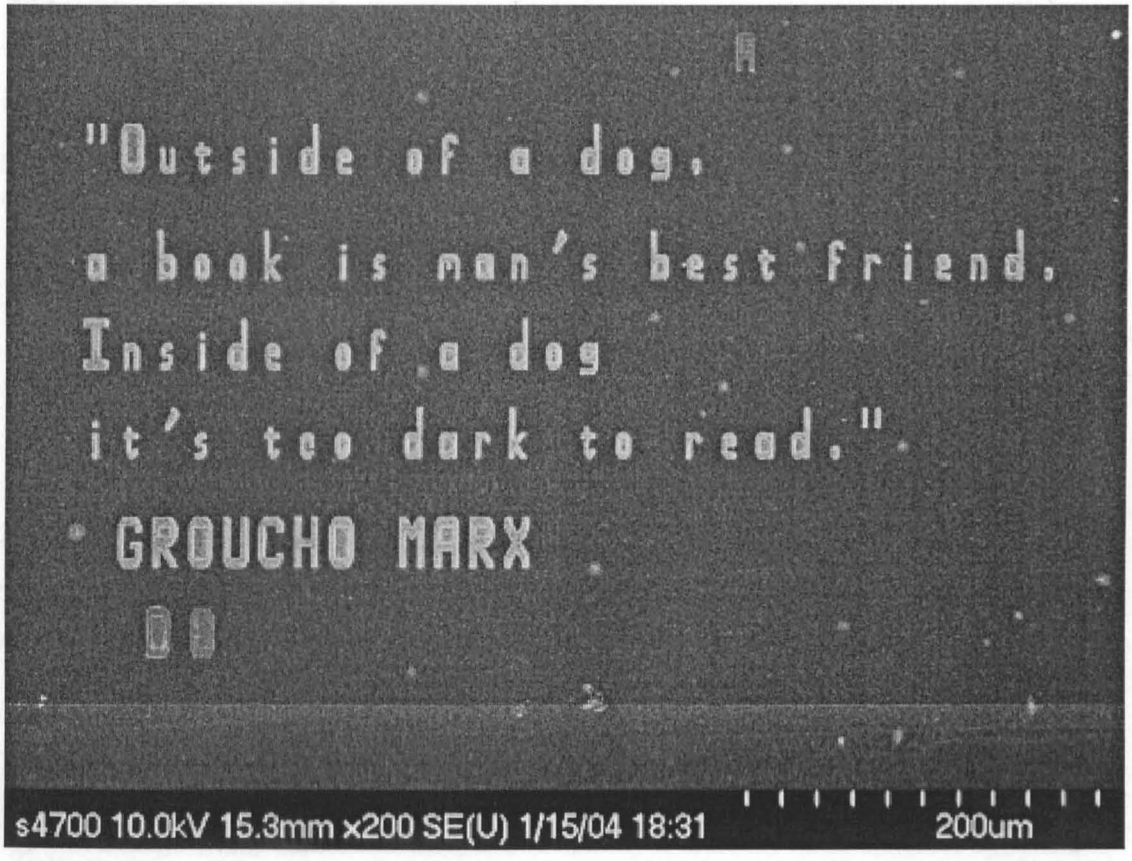
This work includes the realisation of the first self-aligned 120nm T-Gate GaAs pHEMT fabricated at the University of Glasgow. These devices required the development of two key technologies, the non-annealed ohmic contact and the succinic acid based selective wet etch. The self-aligned devices showed good RF performance with a f_t of 150 GHz and a f_{max} of 180 GHz which compares favourable with results of 120nm GaAs pHEMTs previously fabricated at Glasgow.

The investigation of gate length scaling to device performance included the development of two lithographic process capable of producing HEMT with a gate length of 50nm and 30nm

respectively in addition to a method of sample preparation that allows these devices to be analysed using TEM techniques.

This work has lead to the realisation of 50nm T-gate metamorphic HEMTs using a PMMA/copolymer resist stack, these devices displayed an excellent yield, with over 95% of devices working. The uniformity of the gate process was also high with a threshold voltage of -0.445V with a standard deviation of 0.005V. The devices demonstrated an f_t of 330GHz and a f_{max} of 260GHz making these devices some of the fastest transistors that have ever been fabricated on a GaAs substrate.

The second lithography process was developed to realise T-gates with a gate length of less than 50nm. This processed used a two stage "bi-lithography" process to minimise the effect of forward scattering through the resist. The gate footprint was transferred into a SiO_2 gate by a dry etch process. This lithography process was integrated into a full process flow for lattice matched InP HEMTs Using this process, HEMTs were fabricated with a T-gate of 25nm. This is the smallest T-gate device that has been fabricated at the University of Glasgow and is comparable with the smallest HEMT devices in the world.


"Outside of a dog,
a book is man's best friend.
Inside of a dog
it's too dark to read."
GROUCHO MARX
s4700 10.0kV 15.3mm x200 SE(U) 1/15/04 18:31
200um

Publications

Iain Thayne, Xin Cao, David Moran, Euan Boyd, Khaled Elgaid, Helen McLelland, Martin Holland, Stephen Thoms, Colin Stanley, "Very High Performance 50 nm T-gate III-V HEMTs Enabled by Robust Nanofabrication Technologies", To be presented at IEEE NANO 2004, Munich, August 2004.

D.A.J. Moran, E. Boyd, F. McEwan, H. McLelland, C. Stanley, I. Thayne, "Sub 100nm T-gate uniformity in InP HEMT technology", To be presented at GaAs Mantech, Florida, May 2004

Iain Thayne, Euan Boyd, Xin Cao, Khaled Elgaid, Martin Holland, Helen McLelland, Fiona McEwan, Douglas Macintyre, David Moran, Colin Stanley, Stephen Thoms, "Advanced III-V HEMT MMIC Technologies for Millimeter-Wave Applications" European Microwave Week, Munich, Germany, October 2003

X. Cao, E. Boyd, H. McLelland, S. Thoms, C. Stanley, I. Thayne, "mm-wave performance of 50nm T-Gate AlGaAs/InGaAs pseudomorphic high electron mobility transistors with f_T of 200 GHz", European Microwave Week, Munich, Germany, October 2003

X. Cao, S. Thoms, D. Macintyre, H. McLelland, E. Boyd, K. Elgaid, R. Hill, C. Stanley, I. Thayne "Fabrication and performance of 50 nm T-gate for high electron mobility transistors, Microelectronic and Nanoelectronic Engineering 2003, Cambridge, September 2003

D.A.J. Moran, E. Boyd, K. Elgaid, F. McEwan, H. McLelland, C.R. Stanley, I.G. Thayne, "Self-aligned T-gate InP HEMT realisation through double delta doping and a non-annealed ohmic process" Microelectronic and Nanoelectronic Engineering 2003, Cambridge, September 2003

D.A.J. Moran, K. Kalna, E. Boyd, F. McEwan, H. McLelland, L.L. Zhuang, C. Stanley, A. Asenov, I. Thayne, "Self-aligned 0.12 μm T-gate In.53Ga.47As/In.52Al.48As HEMT technology utilising a non-annealed ohmic contact strategy ESSDERC, Estoril, September 2003

Y. Chen, D.S.Macintyre, X. Cao, E. Boyd, D. Moran, H.McLelland, I. Thayne, S. Thoms, "The Fabrication of 50 nm T-Gates Using a PMMA/LOR/UVIII Resist Stack" EIPBN 2003, Tampa, USA, May 2003

Y. Chen, D. S. Macintyre, E. Boyd, D. Moran, I. Thayne and S. Thoms, "High electron mobility

transistors fabricated by nanoimprint lithography", Microelectronic Engineering, 67-68, pp189-195, 2003.

D. Moran, E. Boyd, H. McLelland, K. Elgaid. Y. Chen, D. S. Macintyre, S. Thoms, C. R. Stanley, I. G. Thayne "Novel technologies for the realisation of GaAs pHEMTs with 120 nm self-aligned and nanoimprinted T-gates", Microelectronic Engineering, 67-68, pp769-774, 2003

D.L. Edgar, Y. Chen, F. McEwan, H. McLelland, E. Boyd, D. Moran, S. Thoms, D.S. Macintyre, K. Elgaid, X. Cao, C.R. Stanley, I.G Thayne, "millimeter-wave performance of InAlAs/InGaAs HEMTs using a UVIII/PMMA bilayer for 70 nm T-gate fabrication", European Microwave Conference 2002, Milan, Italy, October 2002

E. Boyd, D. Moran, H.McLelland, K. Elgaid, Y. Chen, D. Macintyre, S. Thoms, C.R. Stanley, I.G. Thayne, "120 nm gate length e-beam and nanoimprint T-gate GaAs pHEMTs utilizing non-annealed ohmic contacts", 2002 International Symposium on Compound Semiconductors, Lausanne, Switzerland, October 2002

Y. Chen, D. S. Macintyre, E. Boyd, D. Moran, I. Thayne and S. Thoms, "High electron mobility transistors fabricated by nanoimprint lithography", Microelectronic and Nanoelectronic Engineering 2002, Lugano, Switzerland, October 2002.

D. Moran, E. Boyd, H. McLelland, K. Elgaid. Y. Chen, D. S. Macintyre, S. Thoms, C. R. Stanley, I. G. Thayne "Novel technologies for the realisation of GaAs pHEMTs with 120 nm self-aligned and nanoimprinted T-gates", Microelectronic and Nanoelectronic Engineering 2002, Lugano, Switzerland, October 2002.

Y. Chen, D. Macintyre, E. Boyd, D. Moran, I. Thayne, S. Thoms, "Fabrication of high electron mobility transistors with T-gates by nanoimprint lithography", J. Vac. Sci. Technol. B, 20(6), pp2887-2890, 2002

Acknowledgements

I would like to thank all the people that have helped me over the course of this work, with their technical expertise, there knowledge and their enthusiasm without which this project would have been impossible within three years.

Firstly I would like to thank my fellow PhD student, friend and flatmate, Dave Moran for his work on the self-aligned GaAs pHEMT described in Chapter 5 as well as the countless discussions of our new ideas, particularly after a few beers. Thanks are also due to the all the people that have helped me over the course of my PhD in various areas that this work has touched upon. Khaled Elgaid for his help with the RF measurement and calibration, Hai-ping Zhou for her assistance in developing the dry-etch process for the sub-50nm gate process, Billy Smith and Sam McFadzean for their help in the work carried the preparation and imaging the TEM samples. Particular thanks go to Yi-Fang Chen for his great enthusiasm and expertise in the fabrication of T-gates.

Without the smooth running of the Ultra-Small Structures Lab a.k.a B3 the work that was described in this thesis would not be possible. Thanks to all the technicians in B3 for their hard work and there enthusiasm, in particular the Ultra-fast systems technicians, Fiona McEwan, Susan and Helen McLellend. Special thanks go to Helen with her encyclopaedic knowledge of the Ultra-Fast transistor processes and her infectious enthusiasm. (with the exception of Tuesdays)

Finally thanks my supervisor Iain Thayne for his hands off supervising style, allowing me to peruse the topic of research that really interested me and for always being there for a wee chat and an approving nod as I put forward my latest idea.

Thanks Everyone.

Table of Contents

1.0 Introduction 1

1.1 Layout of Thesis 3

1.2 References 4

2.0 HEMTS 5

2.1 Structure of a HEMT 6

2.2 Theory of HEMTs 7

2.2.1 The Heterojunction 7
2.2.2 Modulation Doped Heterojunctions 8
2.2.3 Modulation of the 2DEG 10
2.2.4 Metal – Semiconductor Interfaces 17

2.3 Behaviour of a HEMT 24

2.3.1 Current in a HEMT 24
2.3.2 Transconductance 27
2.3.3 Output Conductance 28
2.3.4 Breakdown 28
2.3.5 Small Signal Equivalent Circuit 29
2.3.6 Figures of Merit 33
2.3.7 RF Figures of Merit 34

2.4 Different Types of HEMTs 42

2.4.1 Pseudomorphic Growth 45
2.4.2 Lattice Matched Growth 47

2.5 Conclusions 48

2.6 References 49

3.0 Review of Current Research 53

3.1 GaAs pHEMT 54

3.2 Metamorphic on GaAs 55

3.3 Lattice Matched and Pseudomorphic HEMT's on InP 56

3.4 References 59

4.0 Fabrication Technologies 63

4.1 Wafer Growth 64

4.2 Lithography- 65

4.2.1 Photolithography 66
4.2.2 Electron Beam Lithography 67
4.2.3 Electron Beam Writers 68
4.3.2 Electron Beam Resists 69
4.3.1 Metal Deposition and Lift-off 80
4.3.2 Etching 83

4.4 Fabrication of a HEMT 84

4.4.1 Process flow 84

4.5 References 90

5.0 Development of Self-Aligned HEMT. 92

5. 1 Self-Aligned HEMT 93

5.2 Development of Non-annealed Ohmic Contacts 93

5.2.1 Layer Structures for non-annealed ohmic contact. 94

5.2.2 Ohmic Metal Recipes 95

5.2.3 Further Development of Non-Annealed Ohmic contact 100

5.3 Development of a Selective Gate Recess Etch 101

5.3.1 Succinic Acid etching in other material systems 107

5.3.2 Post Etch Treatments 108

5.4 Realisation of Self-Aligned HEMTs 112

5.5 RF Characterisation 116

5.5.1 Calibrating the VNA 117

5.5,2 Small Signal Modelling and De-embedded S-Parameters 120

5.6 RF Measurement of Self-Aligned Devices 121

5.8 References 128

6.0 Development of Sub-100nm T-gate HEMTs 132

6.1 Gate Length Scaling for High Performance HEMTs. 133

6.2 Development of Sub-100nm Gate Process using Co-Polymer 134

6.2.1 Introduction to 120nm co-polymer process 135

6.2.2 Modifications to 120nm Process 136

6.2.3 Dose Strategy 138

6.2.4 Dose Testing 139

6.2.5 Continuity Tests within a Source-Drain Gap 142

6.3 Fabrication of 70nm HEMTs on wafer A1602 144

6.4 Central Alignment Marker 150

6.5 Development of 50nm Co-polymer Gate Process 152

6.6 Fabrication of 50nm T-gate HEMTs 155

6.7 DC Characterisation of 50nm T-Gate Devices on A1881 159

6.8 RF Characterisation of 50nm HEMT on A1881 162

6.8.1 Cold FET Analysis of 50nm HEMTs 162

6.9 Conclusions 169

6.10 References. 171

7.0 Transmission Electron Microscope (TEM) Analysis of HEMT Devices 172

7. 1 Cross-sectional Analysis by SEM 173

7.1 Sample Preparation by Focused Ion Beam Milling 174

7.2 Methods for TEM sample preparation 175

7.2.1 Trench Method 175

7.2.2 The Lift-out Technique 176

7.3 Sample preparation by the Lift-Out method 178
7.3.2 Sample Preparation using SiO_2 180
7.3.3 Preparation of samples using HSQ 184
7.4 Conclusions 186
7.5 References. 187
8.0 Development of a Sub-50nm T-gate HEMT process ***188***
8.1 Proposed T-gate process 189
8.1.1 UVIII/PMMA Bilayer 189
8.1.2 SiO_2 Gate Support Dielectric. 190
8.1.3 Lithography 190
8.1.4 Fabrication Plan 191
8.2 Development of processes for the fabrication of T-gates 193
8.2.3 Dielectric Deposition 193
8.2.4 Pattern Transfer in SiO_2 layer 195
8.2.5 Lithography 199
8.3 Process Flow of sub-50nm T-gate HEMTs 203
8.3.1 Selective Removal of SiO_2 layer 203
8.4 Device Fabrication 204
8.5 Conclusions 210
8.6 References 212
9.0 Conclusions ***213***
9.1 References ***216***
10.0 Further Work ***217***

Table of Figures

Chapter 2

Figure 2.1 Illustration of the structure of a HEMT.....6

Figure 2.2 Energy Band Diagram.....8

Figure 2.3 Epi-layer structure of HEMT with associated conduction band.....11

Figure 2.4 Schottky Model of Metal-Semiconductor Interface.....18

Figure 2.5 Bardeen Model of Metal-Semiconductor Interface..... 19

Figure 2.6 High and Low doped Metal-Semiconductor interface.....20

Figure 2.7 Band Diagram of Rectifying Junction.....22

Figure 2.8 Current/Voltage Characteristic of Rectifying Contact.....23

Figure 2.9 Schematic cross section of a HEMT.....25

Figure 2.10 Electron Velocity versus Electric field.....26

Figure 2.11 Output Characteristic of a ideal HEMT.....28

Figure 2.12 Intrinsic Lumped Element Equivalent Circuit.....30

Figure 2.13 Lumped Element Circuit with Parasitic Resistances and Inductances.....31

Figure 2.14 Full Lumped Element Equivalent Circuit.....32

Figure 2.15 HEMT biased in a common source configuration.....33

Figure 2.16 Illustration of Two Port Network.....36

Figure 2.17 MAG/MSG and k versus frequency.....42

Figure 2.18 Bandgap of $Al_xGa_{1-x}As$ compounds versus aluminium mole fraction.....44

Figure 2.19 Bandgap of $In_xGa_{1-x}As$ compounds versus indium mole fraction.....46

Figure 2.20 Energy bandgap on InGaAs and AlGaAs versus lattice constant.....46

Figure 2.21 Illustration of growth of mismatched crystal lattice.....47

Figure 2.22 Bandgap versus lattice constant for various semiconductor compounds.....48

Chapter 4

Figure 4.1 Photolithography process flow.....66

Figure 4.2 Schematic of electron beamwriter.....69

Figure 4.3 Flow of preparation from CAD to the beamwriter..... 71
Figure 4.4 Breaking of an C-C bond by interaction with an electron.....73
Figure 4.5 Forward and Backscattering of electrons through sample.....77
Figure 4.6 Energy Density versus radius from a point beam.....78
Figure 4.7 Illustration of Proximity Effect.....79
Figure 4.8 Schematic of an Electron Beam Metal Evaporator.....81
Figure 4.9 SEM image of "flags".....82
Figure 4.10 Illustration of undercut resist profile.....79
Figure 4.11 SEM Image of Photonic Crystal Device.....83
Figure 4.12 Process Flow of a typical HEMT.....84
Figure 4.13 Optical Micrograph of Completed HEMT.....88
Figure 4.14 Active Region of Completed HEMT.....88

Chapter 5

Figure 5.1 Fabrication of Self-Aligned Device.....93
Figure 5.2 Cap Layers of Low Temperature Ohmic Contacts Tests.....94
Figure 5.3 Schematic of TLM test structures.....97
Figure 5.4 Poor Surface Morphology for NiGeAu recipe.....99
Figure 5.5 Conduction Band Energy and Charge Density of A1409.....100
Figure 5.6 Schematic of revised TLM structures.....101
Figure 5.7 SEM image illustrating gate recess offset.....102
Figure 5.8 Results of succinic acid based etch.....105
Figure 5.9 Measured etch depth versus time.....106
Figure 5.10 3D surface plot of gate head.....106
Figure 5.11 DC Characteristics of Self-Aligned Device.....108
Figure 5.12 Gate Current versus Gate Voltage 108
Figure 5.13 DC Behaviour of HCL and HF treated HEMTs.....110
Figure 5.14 SEM image of pitted etch surface.....112
Figure 5.15 Cross Section SEM micrograph of Self-Aligned Ohmic Contacts.....114
Figure 5.16 DC behaviour of a 2x100μm Self-Aligned device.....115

Figure 5.17 Comparison of a Self-Aligned versus standard device.....116
Figure 5.18 Illustration of Ground-Signal-Ground probes and calibration standards...118
Figure 5.19 Composite optical micrograph of CPW calibration standards.....119
Figure 5.20 De-embedded |h21| and MAG/MSG for self-aligned GaAs pHEMT.....121
Figure 5.21 DC characteristics of self-aligned and standard 2x25μm devices.....122
Figure 5.22 Measured & modelled s-parameters of standard and self-aligned HEMT.124
Figure 5.23 De-embedded |h21| and MAG/MSG of 2x50μm standard and self -aligned HEMTs.....126

Chapter 6

Figure 6.1 Illustration of dose pattern of 120nm co-polymer gate.....135
Figure 6.2 SEM cross section of 120nm T-gate resist profile.....136
Figure 6.3 SEM micrograph of resist profile of a 70nm T-gate written at 50keV.....137
Figure 6.4 Sample of a dose test for 70nm T-gates.....141
Figure 6.5 SEM micrograph of metallised resist profile.....142
Figure 6.6 Planar view of etched gate within a source-drain gap.....143
Figure 6.7 DC characteristics of 2x20μm 70nm HEMT on A1602.....146
Figure 6.8 Measured and model s-parameters of 2x20μm 70nm HEMT.....147
Figure 6.9 De-embedded |h21| and MAG/MSG of 70nm HEMT.....148
Figure 6.10 SEM cross section of "70nm" device.....149
Figure 6.11 Illustration of focus problems arising from incorrect sample height measurement.....150
Figure 6.12 Optical micrograph of central alignment marker.....151
Figure 6.13 SEM micrograph of metallised 50nm T-gate structure where the head has become detached.....153
Figure 6.14 SEM micrograph of complete metallised 50nm T-gate.....154
Figure 6.15 SEM micrograph showing plan view of failed 50nm device.....155
Figure 6.16 SEM cross section of 50nm device operating in enhancement mode.....157
Figure 6.17 Output characteristic of a 2x20μm metamorphic HEMT.....159
Figure 6.18 Transfer characteristic of a 2x20μm metamorphic HEMT.....160

Figure 6.19 Histogram of Threshold voltage of 40 devices.....161
Figure 6.20 Comparison of output characteristics with different source-drain gaps.....162
Figure 6.21 Equivalent circuit model of pinched off cold FET.....164
Figure 6.22 Measured and modelled s-parameters of a 2x30μm 50nm HEMT.....165
Figure 6.23 De-embedded |h21| and MAG/MSG of 2x30μm 50nm HEMT.....166
Figure 6.24 Measured and modelled s-parameters of 50nm HEMTs with 2μm and 1.5μm source-drain separations.....167
Figure 6.25 De-embedded |h21| and unilateral gain, U of 50nm HEMTs.....168

Chapter 7

Figure 7.1 Tilting of the sample during final stages of milling.....175
Figure 7.2 FIB Image of TEM sample at various stages of fabrication.....176
Figure 7.3 TEM image of 50nm pHEMT.....178
Figure 7.4 TEM micrograph of T-gate lifted off surface and encased in platinum.....179
Figure 7.5 TEM micrograph of 50nm pHEMT protected by SiO_2.....180
Figure 7.6 TEM micrograph of the damaged channel region using SiO_2.....181
Figure 7.7 STEM image of epi-layers at the edge of the ohmic contacts.....182
Figure 7.8 EDX measured energy spectra of three positions within epi-layer.....183
Figure 7.9 TEM micograph of LMHEMT on InP sample prepared using HSQ.....185

Chapter 8

Figure 8.1 Illustration of proposed process for the fabrication for the fabrication of sub-50nm T-gates.....192
Figure 8.2 SEM micrograph illustrating problem of filling an etched gap in SiO_2.....193
Figure 8.3 SEM micrograph of tests of HF pattern transfer technique.....196
Figure 8.4 SEM cross section of dry etch test.....197
Figure 8.5 Wavemaker design of a sub-50nm gate.....198
Figure 8.6 SEM cross section of 33nm T-gate fabricated using single step

lithography process.....202
Figure 8.7 SEM micrograph of undercutting of PMMA by HF SiO_2 etch.....204
Figure 8.8 Optical micrograph of overexposed gate feeds written using single stage lithography process.....206
Figure 8.9 SEM image of fabricated sub-50nm HEMT using SiO_2 layer.....207
Figure 8.10 Optical micrograph of gate feeds written using bi-lithography process....208
Figure 8.11 SEM micrographs of 25nm T-gate fabricated using bi-lithography process.....209

1.0 Introduction

The waves in the electromagnetic spectrum in the range of 30 to 300GHz corresponding to a wavelength of between 1cm and 1mm are known as millimetre waves There are a number of interesting and varied applications that use the frequency bands within in this spectrum including indoor wireless LANs operating at 60GHz [1.1]. The applications are not however limited to communications. These include automotive radar, which operates at 77GHz [1.2] The W-band is of particular interest because of the low atmospheric absorption of the signal at these frequencies. This low absorption is used in passive imaging systems capable of seeing through fog and finding concealed weapons [1.3]. Above this frequency the G band (140-220GHz) is used for environmental and atmospheric monitoring from space [1.4]. Possible applications exist in the in the sub-millimetre regime. An application receiving a great deal of interest at present is digital circuits for the next generation of data communications that are expected to operate at a data rate of 160Gbit/s, which requires a technology with a cut-off frequency of in excess of 400GHz.

All these applications require circuits containing active devices that are capable of operating at these high frequencies. Many of these applications also require low noise i.e. in the case of passive imaging to give better temperature resolution and lower integration times. This requires the transistors with a very high operating frequency, high gain and low noise. There are a number of technologies capable of operation at frequencies in excess of 1GHz but only a few in excess of 100GHz. The main contenders are RF CMOS using silicon on insulator technologies [1.5], InP Heterojunction Bipolar Transistors (HBTs) [1.6] and High Electron Mobility Transistors. It is the latter that will be investigated in this work, as HEMTs have been

demonstrated with a cut-off frequency of 562 GHz[1.7] making them the fastest three terminal device in the world.

There has been a great deal of interest in HEMTs for over a decade. This initially concentrated on the initial development of the device, then transferring the process from AlGaA/GaAs system to InAlAs/InGaAs on InP for increased performance. In the early to mid nineties work was carried out investigating the increase of performance of a smaller gate [1.8]. Then the emphasis of research shifted the improvement of the performance of the devices with a gate length of around 100nm and the incorporation of these devices into Millimeter Monolithic Integrated Circuits (M^3ICs).

However in the past 3 years there has been a trend to attempt to achieve increased performance by the reduction of the gate length. This has lead to a new generation of ultra-small, ultra-fast transistors. This trend was started in 2001when Suemitsu et al. demonstrated a HEMT with a T-shaped gate with gate length of 35nm and an associated cut-off frequency of 350 GHz, then the device the fastest transistor in the world [1.9]. This prompted other research groups, including Glasgow to develop devices with a reduced gate length to improve the RF performance of HEMTs.

The aim of this project was part of this drive to develop advanced HEMT capable of operating deep into millimetre wave frequencies. This has concentrated on two areas of development; the reduction of parasitic resistances by the development of a self-aligned GaAs pHEMT and the enhancement of the intrinsic device performance by the aggressive reduction of the gate length. Throughout the course of this work a number of processes have been developed including non-annealed ohmic contacts and a selective wet etch enabling the fabrication of self-aligned GaAs pHEMT.

The research into the improvement of transistor performance by the scaling of the gate length included the development of two gate lithography processes, one intended to provide a straightforward process for the fabrication HEMTs with a sub-100nm T-gates and the second to produce HEMTs with a gate length of less than 50nm. These lithography processes have demonstrated a 50nm metamorphic HEMT demonstrating an f_t of 330GHz one of the fastest GaAs based transistors ever produced, in addition to this high performance the process demonstrated very high yield. The second fabrication process developed demonstrated a full process flow for a 25nm T-gate HEMT. To aid the analysis of these very small gate length devices a method was devised to prepare TEM cross sectional samples of the HEMT gate region. This process uses a spin on dielectric layer to protect the delicate gate head and allows the position of the cross-section sample to be placed with an accuracy of around 100nm.

1.1 Layout of Thesis

The thesis is laid out in the following manner. Chapter 2 introduces the HEMT and the theory of its operation. This is followed by a review of the current status of research into HEMT. Chapter 4 presents the methods and techniques that are used in the fabrication of the HEMTs described throughout the rest of this work. Chapter 5 deals with the development of the self-aligned HEMT including the realisation of the non-annealed ohmic contact and the selective wet chemical etch. The self-aligned HEMT is then presented and its performance discussed. Chapters 6, 7, and 8 deal with the scaling of the gate length of HEMTs. Chapter 6 reviews the need for the reduction of gate length as well as describing the development of the sub-100nm process concluding in the demonstration of a 50nm metamorphic HEMT fabricated using this process. Chapter 7 details the work carried out on the development of a technique for the preparation of cross sectional HEMT samples for examination by Transmission Electron Microscopy, a technique that becomes increasingly valuable as devices are scaled. Chapter 8 describes the process designed for the realisation of HEMTs with a gate length of less than 50nm and describes the problems associated with producing T-gate structures of this size.

1.2 References

[1.1] P. F.M. Smulders et al. "*A shaped Reflector Antenna for 60GHz Indoor Wireless LAN Access Points*" IEEE Transactions on Vehicular Transport, **50**, 2001, p584

[1.2] L.Raffaelli, "*Millimeter-wave automotive radars and related technology*", Microwave Symposium Digest, IEEE MTT-S, **1**, 1996, p17

[1.3] A, Lawrence, "*Millimeter-wave IC's open up new spectrum*", Compound Semiconductor, May 2001, http://compoundsemiconductor.net/articles/magazine/7/5/1/1

[1.4] G. Dambrine et al. "*An overview of Low Noise Circuits and Associated Devices for 100-200GHz Apace Applications*", Proceedings of GAAS2003, 2003, p473.

[1.5] T. Hirose et al."*A 185GHz f_{max} SOI DTMOS with a new Metallic Overlay-gate for Low Power RF applications*", IEDM 2001 Tech. Dig., 2001., p33.5.1

[1.6] M. Rodwell et al. "*InP Bipolar Transistors: High Speed Circuits and Manufacturable Submicorn Fabrication Processes*", Proc GAAS2003, 2003, p537

[1.7] A. Endoh et al., "*InP-Based High Electron Mobility Transistors with a very Short Gate-Channel Distance*", JJAP, **42**, 2003, p2214

[1.8] L.D. Nguyen et al. "*50-nm Self-Aligned-Gate Pseudomorphic AlInAs/GaInAs High Electron Mobility Transistors*" IEEE Trans. Elec. Dev., **39**, 1992, p2007.

[1.9] T. Suemitsu et al., "*30nm-Gate InAlAs/InGaAs HEMTs Lattice Matched to InP Substrates*", IEDM Tech. Dig., 2001, p223.

2.0 HEMTS

The device studied in this project is the High Electron Mobility Transistor (HEMT). A number of names are used to describe this type of device. Including Modulation Doped Field Effect Transistor (MODFET), Two Dimensional Gas Field Effect Transistor (TEGFET) and Selectively Doped Heterojunction Transistor (SDHT). The basis for these names shall become clear later. In this document the name HEMT shall be used throughout. The HEMT is particular type of Field Effect Transistor, (FET), in which, unlike the Silicon MOSFET or the GaAs MESFET, the charge modulated by the gate is confined to a channel between two heterojunctions.

2.1 Structure of a HEMT

A simplified illustration of the HEMT is shown in Figure 2.1

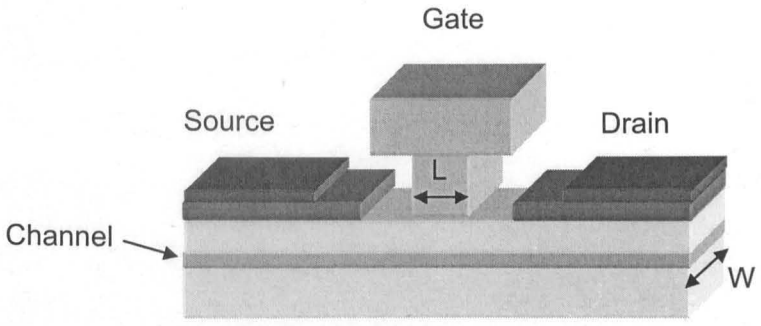


Figure 2.1 Illustration of cross section of a HEMT.

The structure consists of three metal contacts on the surface of the semiconductor substrate, the source, the drain and the gate. The important dimensions of the gate are shown. The gate length, L, is the length of gate contact parallel to the direction of current flow from source to drain, this is the smallest dimension of the gate. The width of the gate, W, is in the direction perpendicular to the current flow from source to drain.

2.2 Theory of HEMTs

Two types of metal contacts are used to make external connection to the HEMT, the Schottky gate contact, and ohmic source and drain contacts. The important difference between the two types of contacts is their current-voltage (I-V) characteristics. The Schottky contact is rectifying, allowing the current to flow in one direction and not in the other. In the case of the ohmic contact the current is able to flow equally in both directions. The semiconductor material contains the heterojunction that is the basis for this type of transistor.

2.2.1 The Heterojunction

A heterojunction is an interface between two dis-similar semiconductors. This structure was first suggested for use in transistors in 1948 as a means of increasing the emitter efficiencies in bipolar transistors [2.1]. However it was not until much later that epitaxial growth technologies matured sufficiently to enable the growth of these structures. Much of the early work concentrated on the heterojunction between doped germanium and gallium arsenide [2.2]. However it was discovered that the heterojunction formed at the junction of GaAs and AlGaAs could be important for the realisation of electronic devices.

In order to describe the mechanisms of heterojunction formation and the possible benefits of using them within the structure of a FET the conduction band diagram of such an interface should be considered. To construct this band diagram the Anderson model [2.3] will be used. This gives a simple method for aligning bands of the semiconductors.

Consider two different semiconductor materials, in this example these will be the wide bandgap $Al_{0.3}Ga_{0.7}As$ and the smaller bandgap material, GaAs. These materials have similar lattice constants, so it is possible to grow one material to be grown on top of the other without significant strain. The band gaps of the materials at room temperature are 1.75eV and 1.424eV for $Al_{0.3}Ga_{0.7}As$ and GaAs respectively. The materials also have a different electron affinity, χ, defined as the energy required to move a free electron from the conduction band to just above the surface of the semiconductor i.e. the vacuum level. Similar to the electron affinity is the work function, ϕ, this is the energy that is required to move an electron from the Fermi level to the vacuum level. Because the workfunction is relative to the Fermi level it is dependent on the doping levels of the semiconductor. In this example we shall assume that the materials are un-doped. The dielectric constants of the materials are assumed to be different.

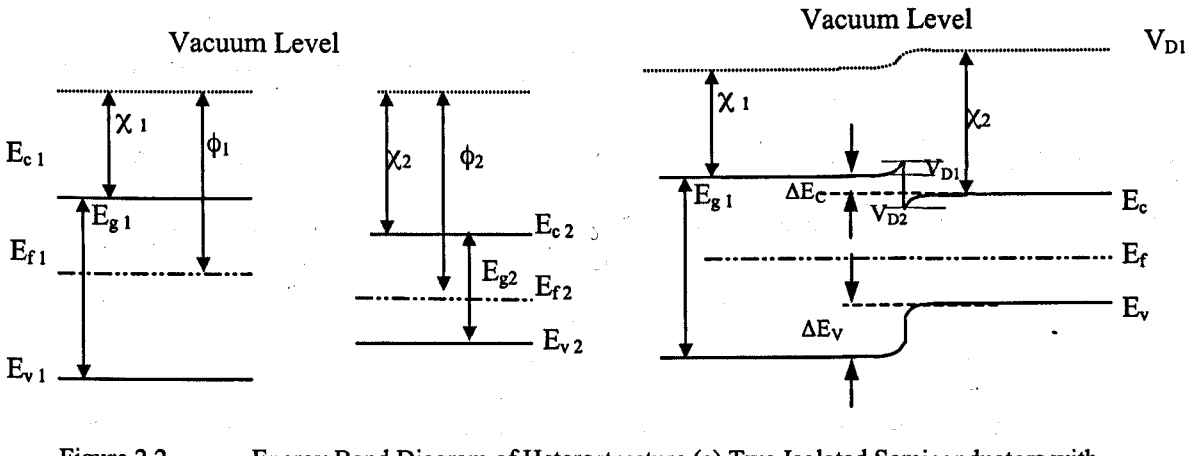


Figure 2.2 Energy Band Diagram of Heterostructure (a) Two Isolated Semiconductors with vacuum levels aligned (b) Semiconductors brought together in Equilibrium Condition.

The first step in creating the band diagram is to align the vacuum level of the two isolated semiconductors which is assumed to be constant before contact. This is shown in Figure 2.2 (a), it is seen that the Fermi level of the two material do not align. For the junction to be in equilibrium the Fermi levels must align, this condition forces electrons from the higher bandgap material to move across the interface. This depletion of electrons causes the conduction band to bend upwards, equally the accumulation of electrons in the lower bandgap material causes the band to bend downwards. The amount that each bands bends by can be found by considering the electrostatic potential supported in each material. The difference between the two workfunctions is the built in voltage, V_D.

$$V_D = V_{D1} + V_{D2} \qquad (2.1)$$

Where V_{D1} and V_{D2} are the potentials supported in each material. The ratio of potential in each side of the interface and hence the curvature of the band can be found using Poisson's equation, this is given by

$$(V_{D1}/V_{D2}) = N_2\varepsilon_2/N_1\varepsilon_1 \qquad (2.2)$$

Where N_1 and N_2 are the instrinsic carrier doping levels of each material. The difference in conduction band energies is given by

$$\Delta E_c = \chi_1 - \chi_2 \qquad (2.3)$$

and the energy gap in the valence band is

$$\Delta E_v = (\chi_1 - E_{g1}) - (\chi_2 - E_{g2}). \qquad (2.4)$$

2.2.2 Modulation Doped Heterojunctions

The structure just studied was an un-doped heterojunction, which adequately describes the principle of the formation of a heterojunction. However the electron density, n_D, at the interface is limited by the number of intrinsic donors in the AlGaAs layer. More electrons can be introduced by means of introduction Si atoms into the crystal as it is grown. Historically this was achieved by uniformly doping both the GaAs and AlGaAs layers in the heterojunction, thus increasing the number of electrons at the interface. This produced superlattices with a carrier density as high as $1x10^{18}cm^{-3}$ [2.4]. However, since the GaAs layer is doped as well as the AlGaAs layers the mobility of the electrons was badly degraded due to the large number of ionised donors in close proximity to the electrons. This problem was solved with the introduction of modulation doping of the superlattices, this method of doping only introduces silicon donors into the AlGaAs material leaving the GaAs material un-doped. This was first demonstrated in 1978 by Dingle et al. [2.5] who grew AlGaAs/GaAs superlattices by Molecular Beam Epiltaxy (MBE) with the Si source synchronised with the AlGaAs growth resulting in only the AlGaAs layer being doped. A further refinement was made to the process, which prevented doping in the region of 60Å from the edge of the AlGaAs layer. This has the effect of further separating the electrons from the ionised donors. This method allowed the growth of materials with comparable electron density, n_s, while simultaneously demonstrating electron mobilities, μ, as much as twice that of the uniformly doped structure.

The effect of modulation doping on the heterojunction of Figure 2.2 will now be discussed. The inclusion of doping into the AlGaAs has a number of effects, the ionised Si atoms will provide many more electrons which will diffuse across the heterojunction. These electrons will lose their energy by phonon interaction and get trapped in the well on the lower energy side of the interface. This movement of carriers across the interface continues until the electric dipole that is set up between the electrons and the ionised donors equal the conduction band offset, ΔE_c, of

the heterojunction. An electric dipole is formed between the ionised donors and the electrons, which attracts the electrons towards the donors, however because of the energy barrier they cannot pass through the interface. This produces a very narrow potential well with a thickness of between 5-10nm [2.6]. As this is of the order of the size of an electron wavelength, the motion of the electron therefore becomes quantised in the direction perpendicular to the interface. Due to the fact that the electrons are limited to travelling in two dimensions this accumulation of electrons has been labelled a Two Dimensional Electron Gas (2DEG). The properties of the 2DEG make it very useful as the basis for a high-speed transistor. Unlike the electrons within the GaAs MESFET the electrons of the 2DEG are confined to a specified distance from the surface. Most importantly, because the electrons are spatially separated from the donors, the mobility, particularly low temperature mobility can be very high due to the reduction of Coulombic scattering from the ionised donors.

2.2.3 *Modulation of the 2DEG*

For the 2DEG to be useful as the basis for a field effect transistor it must be possible for the number of electrons within the 2DEG to be modified by an applied voltage. This shall be investigated by studying the example of a AlGaAs/GaAs HEMT. This structure is shown in Figure 2.3. Starting from the surface and moving through the semiconductor structure, the first layer is the cap layer with a thickness c, which tends to be highly doped n-type, N_c, for reasons that will be discussed later.

Below this is AlGaAs supply layer, this has a doping of N_D and a thickness of d. The thin layer between the supply layer and the GaAs substrate is the spacer, undoped and has a thickness of s. The GaAs substrate is assumed to be un-doped and is much thicker than any other layers.

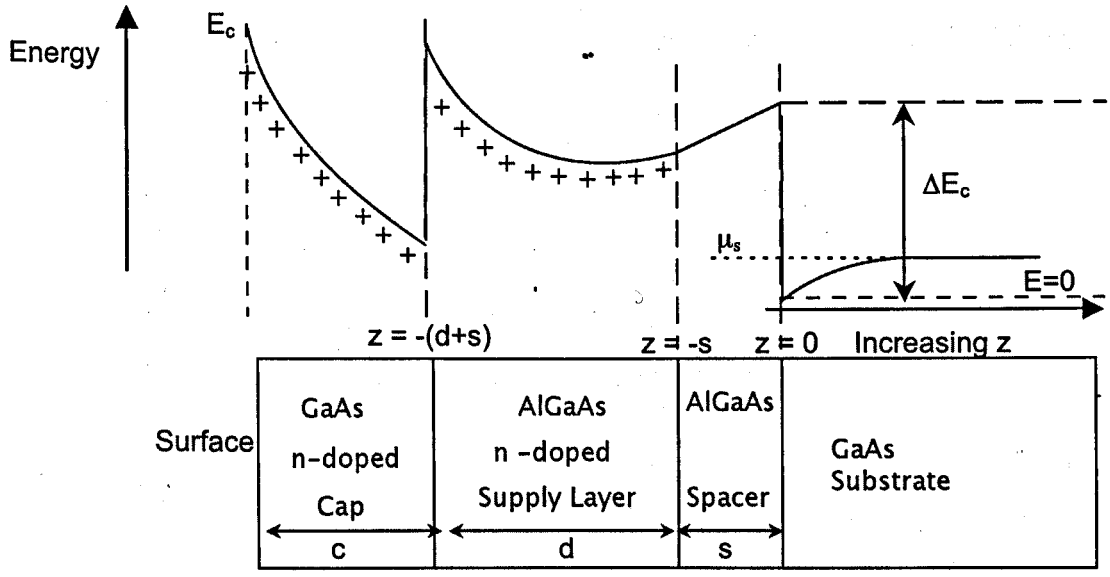


Figure 2.3 Structure of a Simple AlGaAs/GaAs HEMT structure with sketch of associated conduction band diagram.

The conduction band of this structure can be sketched starting from deep within the GaAs substrate. Far away from the heterojunction the electric potential will be zero, hence the energy band will be flat. As the interface is approached there will be an accumulation of electrons, this will cause the energy band to curve downwards, the curvature will continue to increase until the interface. When entering the AlGaAs spacer the band will step up by the conduction band offset ΔE_c. Since the spacer layer is un-doped there will be no charge to bend the band, hence the gradient will be almost identical either side of the interface with a slight modification due to the different permitivities as given in equation 2.2. In the AlGaAs doped supply region the ionised donors will cause the energy band to curve upwards, as the GaAs cap is entered the energy will step down by ΔE_c the band will continue to curve upwards due to depletion of electron from the cap. At the surface of GaAs there are surface states caused by the breaking of the periodic structure [2.7]. These surface states have the effect of "pinning" the Fermi energy at a certain energy below the conduction band [2.8]. This causes a depletion region at the surface of the cap, this depletion region will cause the band to bend upwards further as the band reached the surface.

It is possible to build up an expression that describes the effect on the 2DEG density of an applied bias at the surface. There are a number of assumptions that will be used to simplify the expression. These are the following:

1. All of the donors in the device are fully ionised. In the case of AlGaAs this is not always the case.
2. The relative permitivites of AlGaAs and GaAs are identical.
3. The only carriers in the 2DEG are electrons
4. The electrons in the 2DEG are degenerate and only occupy one sub-band. This happens only at low temperatures and low electron densities.

The electron density in the 2DEG will at first be assumed, and the applied voltage required to achieve this will be found. At the interface between the AlGaAs spacer and the GaAs substrate, the depth will be taken as zero. ($z = 0$)

In the GaAs at the interface the Electric field, F, from Gauss's law at $z=0^+$ will be:

$$F(0^+) = \frac{-qn_D}{\varepsilon} \tag{2.5}$$

On the other side of the interface, in the undoped AlGaAs spacer layer, Poisson's equation gives

$$\frac{d^2V}{dz^2} = 0\Big|_{z=0^-} \tag{2.6}$$

where V is the electrostatic potential, the total charge is zero because the material is undoped. However,

$$\frac{dV}{dz} = -F(z) \tag{2.7}$$

Therefore

$$\frac{dV}{dz} = \frac{qn_D}{\varepsilon}\Big|_{z=0} \tag{2.8}$$

Integrating (2.8) yields

$$V(z) = \frac{qn_D z}{\varepsilon} + const \qquad (2.9)$$

To find the constant, consider the potential either side of the interface, (z=0). Within the GaAs layer $V(0^+) = 0$ at the AlGaAs side interface i.e. z = 0 the potential is $V(0^-) = \frac{-\Delta Ec}{q}$ because of the change into the higher bandgap material.

Hence within the AlGaAs layer, the potential will be

$$V(z) = \frac{qn_D z}{\varepsilon} - \frac{\Delta E_c}{q} \qquad (2.10)$$

At $z = -s|_+$, the potential will be

$$V(-s_+) = -\frac{qn_D s}{\varepsilon} - \frac{\Delta E_c}{q} \qquad (2.11)$$

In the n doped AlGaAs layer Possion's equation gives

$$\frac{d^2V}{dz^2} = \frac{-qN_d}{\varepsilon}$$

Integrating gives

$$\frac{dV}{dz} = \frac{-qN_d z}{\varepsilon} + const \qquad (2.12)$$

Using (2.9) and equating the electric fields at either side of the interface ($z = -s$) gives

$$\frac{qN_d s}{\varepsilon} + const = \frac{qn_D}{\varepsilon}\Big|_{z=-s+}$$

The constant of integration can be found and substituted into (2.12) which yields

$$\frac{dV}{dz} = -\frac{qN_d z}{\varepsilon} - \frac{qN_d s}{\varepsilon} + \frac{qn_D}{\varepsilon} \qquad (2.13)$$

Integrating this gives the potential in the doped region with respect to z, the depth, the associated constant can be found by using equations 2.11 and 2.13 and considering the potential on each side of the supply/spacer interface i.e. $z = -s$. This yields the following expression

$$V(z) = -\frac{qN_d}{2\varepsilon}(z+s)^2 + \frac{qn_D}{\varepsilon} z - \frac{\Delta E_c}{q} \qquad (2.14)$$

Now consider the GaAs Cap/AlGaAs interface. Using the same method as above, but noting that the total charge in the AlGaAs layer will be N_c as opposed to N_d. Using Poisson's equation, integrating, setting the both sides to be equal at the interface $(z = -(d+s) = -e)$ and integrating again gives the following expression for the potential within the GaAs cap.

$$V(z) = \frac{-qN_c}{2\varepsilon}(z^2 - 2e) - \frac{qN_d}{2\varepsilon}(2e+s)z + \frac{qn_D}{\varepsilon}z + const \qquad (2.15)$$

The constant can again be found by equating the potentials at the interface (z = -e), this time taking into consideration the conduction band offset from the AlGaAs into the GaAs. By doing this and making the substitution $e = d+s$ an expression for the potential within the GaAs cap is given as

$$V(z) = \frac{-qN_c}{2\varepsilon}(z+(d+s))^2 + \frac{qN_d d}{2\varepsilon}(z+(s+\tfrac{1}{2}d))z + \frac{qn_D}{\varepsilon}z \qquad (2.16)$$

When the expression for the potential is known it can be used to give the conduction band energy $E_c(z) = -qV(z)$, this will give the energy of the region $z < 0$ just studied. The conduction band energy in the region $z > 0$ is still unknown.

It is now necessary to find the Fermi levels of both the gate metal and the 2DEG with respect to the zero energy at $z = 0$. In order to simplify the expression and because it is more realistic of devices the gate will be placed on the AlGaAs supply layer. The potential just under the gate metal is given by equation 2.10. The Fermi level, μ_m, will be pinned V_b below the conduction band, hence

$$\mu_m = \frac{q^2}{\varepsilon}\left(\tfrac{1}{2}(d+2s)^2 N_D + (d+s)n_D\right) - \Delta E_c - V_b \qquad (2.17)$$

The Fermi level of the 2DEG is given the energy of the bound state of the 2DEG plus the Fermi energy of the electron gas. Thus the Fermi level of the 2DEG, μ_s, is given as

$$\mu_s = E_1 - \left(\frac{\pi\hbar^2}{m}\right)n_D \qquad (2.18)$$

Where E_1 is the energy level of the bound state of the 2DEG, this is also a function of the electron density of the 2DEG. The applied gate bias is the difference between the two Fermi energies divided by the charge of an electron, q. Hence

$$V_g = \frac{\mu_s - \mu_m}{q} \tag{2.19}$$

Substituting (2.17) and (2.18) into (2.19) gives

$$V_g = \frac{E_1(n_D)}{q} - \frac{1}{q}\left(\frac{\pi\hbar^2}{m}\right)n_D + \frac{V_b}{q} - \frac{q}{\varepsilon}\left(\tfrac{1}{2}(d+2s)^2 N_D + (d+s)n_D\right) + \frac{\Delta E_c}{q} \tag{2.20}$$

The term for the Fermi energy of the 2DEG, $\left(\frac{\pi\hbar^2}{mq}\right)n_D$ can equally be written as $\frac{q}{\varepsilon}\left(\frac{\pi\varepsilon\hbar^2}{mq^2}\right)n_D$ and can be brought into the electrostatic expression. It is noted that the Bohr radius is given by $a = \left(\frac{4\pi\varepsilon\hbar^2}{mq^2}\right)$ hence it is seen that the effect of the density of states makes the 2DEG appear to be a/4 further into the material than would be suggested from the geometry of the layer structures. The expression for the applied gate bias can be rewritten as

$$V_g = \frac{E_1(n_D)}{q} + \frac{V_b}{q} - \frac{q}{\varepsilon}\left(\tfrac{1}{2}(d+2s)^2 N_D + n_D\left(d+s+\tfrac{a}{4}\right)\right) + \frac{\Delta E_c}{q} \tag{2.21}$$

This equation gives the applied gate voltage that is required to achieve a particular electron density, n_D. This will now be used to determine the threshold voltage.

Threshold Voltage

The threshold voltage is the applied voltage that is required to fully deplete the 2DEG i.e. this is the voltage necessary to set $n_D = 0$. The potential well will disappear so E_1 will be zero, hence.

$$V_T = \frac{V_b}{q} + \frac{\Delta E_c}{q} - \frac{q}{\varepsilon}\left(\tfrac{1}{2}(d+2s)^2 N_d\right) \tag{2.22}$$

2DEG Density

It is also possible to rearrange equation 2.21 to give an expression that shows how the 2DEG density varies with gate bias. There is however one term that is unknown at this point, the term

including the energy of the lowest energy state, E_1. The expression can be simplified by subtracting equation 2.22 from 2.21, giving the voltage, this gives

$$(V_g - V_T) = \frac{E_1(n_D)}{q} - \frac{qn_D}{\varepsilon}\left(d + s + \tfrac{a}{4}\right) \tag{2.24}$$

The expression resembles the equation for a capacitor, where,

$$\frac{1}{C_g} = \frac{1}{q}\frac{\partial(V_g - V_T)}{\partial n_D}$$

Differentiating equation 2.23 gives

$$\frac{\partial(V_g - V_T)}{\partial n_D} = \frac{1}{q}\frac{\partial E_1}{\partial n_D} + \frac{q}{\varepsilon}\left(d + s + \tfrac{a}{4}\right)$$

$$= \frac{q}{\varepsilon}\left(d + s + \tfrac{a}{4} + \frac{\varepsilon \partial E_1}{q^2 \partial n_D}\right) \tag{2.25}$$

The capacitance is therefore

$$C_g = \frac{\varepsilon}{\left(d + s + \tfrac{a}{4} + \frac{\varepsilon \partial E_1}{q^2 \partial n_D}\right)}$$

This is the equation for a parallel plate capacitor with the inclusion of the differential of the lowest bound state energy. Much in the same way that a/4 was included to account for the density of states this term may also treated as a distance, h, which accounts for the fact that the centre of the charge distribution of the potential well is not located at the interface but a distance into the well [2.9]. This distance is normally around 5-10nm. Using this term it is possible to simplify and rearrange equation 2.25 to give the sheet density as a function of gate voltage.

$$n_D = \frac{\varepsilon(V_g - V_T)}{q(d + s + \frac{a}{4} + h)} \tag{2.26}$$

In order to describe accurately the charge control of the 2DEG, the value for h must be found, it was seen that this depends on the energy of the lowest bound state in the 2DEG, E_1. This will depend on the exact shape of the quantum well. There are a number of approximations that are used to describe the quantum well, from the simplest case of the triangular well to more advanced approximations such as the Fang-Howard approximation [2.10]. In both cases the depth of the well will be a function of the density of electrons within it. However to find the electron density the bound state energy must be known. This problem requires that the Possion and Schrödinger equations be solved in a self-consistent manner. An approximation is found for the charge density, this is then fed back into the Schrödinger equation to find the potential of the quantum well. The potential is used to find a more accurate value of the charge density. This process is repeated for a number of iterations until the desired accuracy is achieved. Using this method is possible to develop an accurate model of the behaviour of the modulation-doped heterojunction.

2.2.4 *Metal – Semiconductor Interfaces*

In the last section the behaviour of semiconductor-semiconductor interfaces was studied. The other important type of interface in the HEMT is the metal-semiconductor interface. The importance of this interface is that if the properties of semiconductors are to be exploited it must be possible to connect to the wider world.

The Schottky Model

Schottky was the first to create a model of the metal-semiconductor interface.[2.11,2.12] which is described next. Consider the situation illustrated in Figure 2.4(a) a metal with a work function ϕ_m and a n-type semiconductor a workfunction, ϕ_s. For the case of the semiconductor it is more useful to use the electron affinity, χ_s rather than workfunction since this is a measure of the energy required to free an electron from the conduction band rather than the Fermi level. This can be written as

$$\phi_s = \chi_s + \zeta \qquad (2.27)$$

where ζ is the energy between the Fermi level and the conduction band.

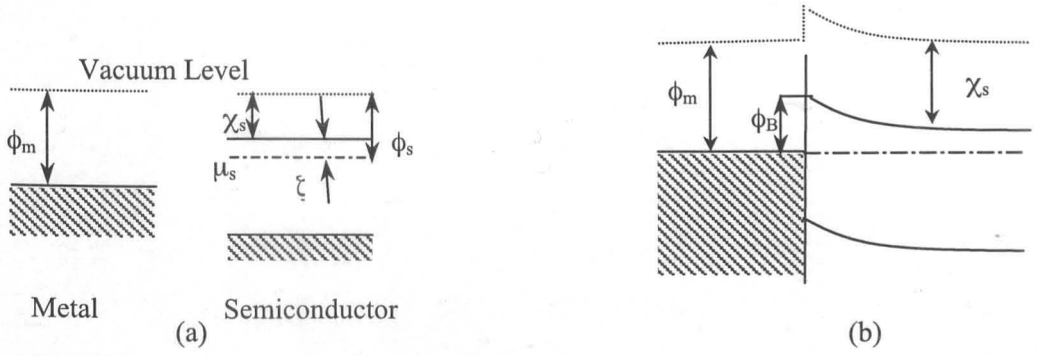


Figure 2.4 Illustration of Schottky model of metal-semiconductor interface (a) Equilibrium state with metal and semiconductors separated. (b) Equilibrium state when brought together. The work function for the metal is ϕ_m, and ϕ_s for the semiconductor. Semiconductor electron affinity is χ_s, ζ is the energy difference between the Fermi level and the conduction band.

When the metal and semiconductor are not in contact, the vacuum level is assumed to be constant. It seen that the Fermi level of the two materials do not align. As the metal and semiconductor are brought together electrons will flow from the semiconductor into the metal until, at equilibrium the Fermi levels are aligned. The transfer of electrons will cause the conduction band in the semiconductor to bend upwards as it approaches the interface creating a barrier as is shown in Figure 2.4(b). The height of the barrier is then

$$\phi_B = \phi_m - \chi_s \qquad (2.28)$$

The barrier height is therefore dependent on the workfunction of the metal and the electron affinity of the semiconductor. It therefore should be possible to control the barrier height by judicious choice of metal and doping of the semiconductor. It was however found experimentally that this was often not the case. Hence another model of the metal interface was required.

The Bardeen Model

It was stated in section 2.2.2 in the analysis of the heterojunction that when a metal and semiconductor are brought into contact the Fermi level is pinned to a value of V_b below the conduction band. This follows from the Bardeen Model which was proposed in 1947 [2.13].

This model suggested that if there were a large number of surface states at the semiconductor-metal interface then the barrier energy would become independent of the work function of the metal. (Surfaces states are electronic states at the surface of the semiconductor produced by the interruption of the periodicity of the lattice.)

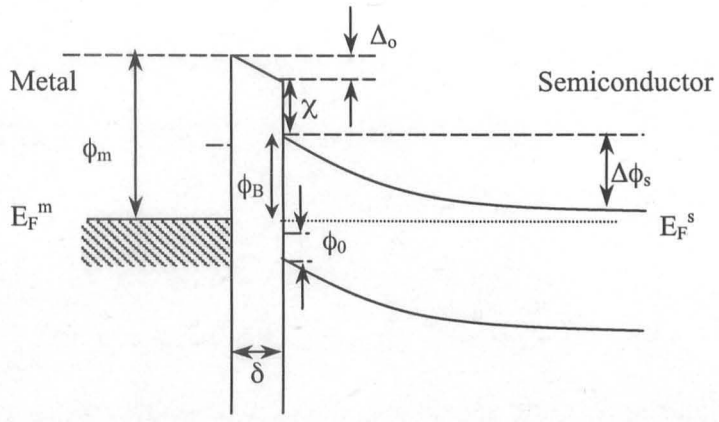


Figure 2.5 Diagram of general metal-semiconductor interface with interface layer with a thickness δ. Potential across interface layer is Δ_0, the diffusion potential $\Delta\phi_s$, the barrier height, ϕ_B, ϕ_0 is the energy above the valence band that the surface states are required to be filled above to achieve charge neutrality at the semiconductor surface.

In Bardeen's model it assumed that at the interface there exists a thin interfacial layer. This is thin enough to be electron transparent but it is able to support a potential across it. This situation is shown in Figure 2.5. As before, the workfunction of the metal is ϕ_m, the electron affinity of the semiconductor is χ and the bandgap of the semiconductor is E_g. A new term is introduced, ϕ_0, this is the level above the valence band that the surface states must be filled to ensure that the surface is neutral. If the density of surface states is D_{ss} and is uniform the charge due to the surface states will be

$$Q_{ss} = -eD_{ss}(E_g - \phi_0 - \phi_B)$$

There will be an equal and opposite image charge in the metal. If the density of the surface states is large enough the entirety of the potential difference, ϕ_m-ϕ_s, will be supported by the surface states. There is no need for any diffusion of electrons across the interface; hence the barrier height will be independent of the metal workfunction. The barrier height will be

$$\phi_B = E_g - \phi_o \qquad (2.29)$$

This is often referred to as Fermi level pinning as the Fermi level is "pinned" to an energy below the conduction band, or rather, the conduction band is forced to be at an energy above the equilibrium Fermi level by the amount E_g-ϕ_0. In GaAs this value is typically 0.7eV.

In general the real barrier energy lies between these approximations. Cowley and Sze [2.14] were the first to analyse the general case. They showed that the barrier energy with zero electric fields is given by

$$\phi_B = \gamma(\phi_m - \chi_s) + (1-\gamma)(E_g - \phi_0) \qquad (2.30)$$

where $\gamma = \varepsilon_i/(\varepsilon_i + qN_{ss}\delta)$, ε_i, and N_{ss} being the permittivity of the interface layer and the density of the surface states. The barrier energy depends on two terms. A Schottky term, that depends on the workfunction of the metal and the electron affinity of the semiconductor. A Bardeen term that involves the difference between the bandgap energy, E_g, and the energy to which surface states are filled to, ϕ_0. The magnitude of the weighting term, γ, depends on the nature of the interface layer on the surface.

A diagram of the metal-semiconductor interface for two different semiconductor materials is given below. Figure 2.6(a) shows the metal in contact with a highly n-doped low bandgap material such as GaAs. Figure 2.6(b) shows the case of an undoped higher bandgap material such as AlGaAs

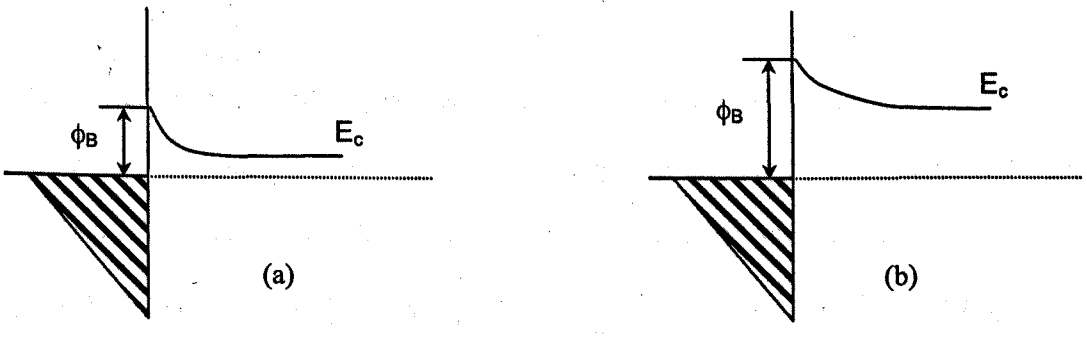


Figure 2.6 Illustration of metal-semiconductor interface for (a) highly doped low band gap material (b) undoped higher bandgap material.

In the situation described in Figure 2.6(a), the highly doped low bandgap semiconductor, as is seen in Equation 2.29 the barrier height, ϕ_B, will be low due to the small bandgap. In addition to this, the incorporation of doping will reduce the width of the depletion region that is formed at the surface. In the case illustrated in Fig 2.6(b), the undoped larger bandgap material the barrier height will be greater. Since the material is undoped the depletion region will also be wider.

Electronic Transfer

There are a number of ways that electrons can pass through barriers like those described above. These are:

Thermonic Emission, when the electrons have an energy greater than the barrier energy $E > e\phi_B$, they are able to pass over the barrier.

Field Emission, this is where the electrons have an energy below the barrier energy, $E < e\phi_B$, they are able to pass through the barrier by quantum mechanical tunnelling. The probably of the electron tunnelling though depends on the barrier energy and the thickness of the barrier and is estimated by the WKB approximation.[2.15]

$$T = \exp\left[-\left(\frac{2m\phi_B}{\hbar^2}\right)^{1/2} d\right] \qquad (2.31)$$

where d is the barrier thickness. This is true for electrons with an energy at the Fermi level. For electrons with a greater energy, the probability will depend on the energy difference between the electron and the top of the barrier. At low temperatures the dominant transport mechanism is field emission, at high temperatures thermionic emission is the dominant mechanism.

Rectifying Contacts

In the case of the rectifying contacts the main process of electron transfer is by thermionic emission. In this case the barrier height is much greater than the thermal energy of the electrons ($\phi_b >> kT$). The barrier is wide hence the tunnelling current is low. The energy diagram for such a example is shown in Figure 2.7(a-c) in the case of equilibrium, forward bias and reverse bias.

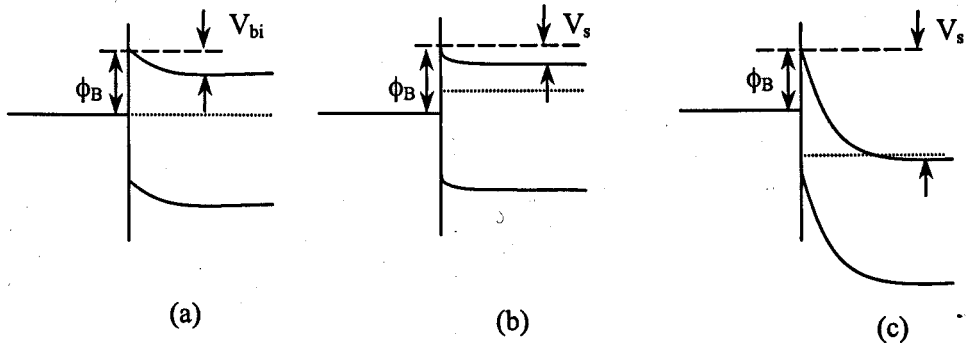


Figure 2.7 Band diagram for rectifying junction (a) Equilibrium State (b) Forward bias and (c) Reverse bias conditions. The barrier height, ϕ_B, and the built in voltage, V_{bi}, are shown.

The barrier height ϕ_b is given by Equation 2.30. This depends on the interface state density, in the case of III-V compounds the surface state density is high, hence the barrier energy will be largely independent of the metal. The built in voltage, V_{bi}, is the barrier caused by the depletion of electrons. If the interface is forward biased the Fermi level in the semiconductor is raised as shown in Figure 2.7(b). There will be negligible current flow when the applied bias is less that the built in voltage, V_{bi}. However as the applied voltage is increased the Fermi level will reach the height of the barrier and many electrons will flow into the metal causing the observed exponential increase of current with voltage. The application of a reverse bias will pull down the conduction band of the semiconductor, resulting in a larger barrier to current flow. In the reverse bias regime, a limited number of electrons will still be able to pass though the barrier by tunnelling, which will lead to a small current, referred to as leakage current.

The current-voltage characteristic of the rectifying contact was first derived by Beth [2.16]. The current density, J, for an applied voltage is given by

$$J = A^* T^2 \exp(^{-q\phi_B}/_{kT})[\exp(qV/kT) - 1] \qquad (2.32)$$

where $A^* = 4\pi q m_e^* k^2 / h^3$ and m^*_e is the electron effective mass. A^* is called the Richardson constant and depends on the semiconductor.

The I-V characteristic of the contact is sketched in Figure 2.8.

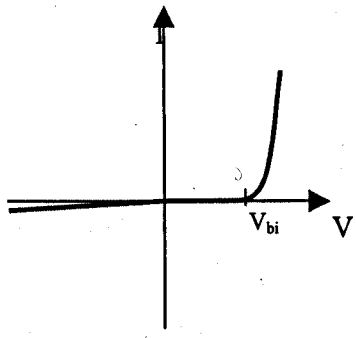


Figure 2.8 Current-Voltage Characteristic of a rectifying contact. This shows the asymmetric behaviour of the contact. A small leakage current is shown in the negative bias regime.

Ohmic Contacts

The purpose of the ohmic contact is to provide a low resistance path from the 2DEG to the surface via the metal contacts. The current voltage behaviour of the contact must be independent of the direction of current flow and have a low resistance. It is seen above that in the case of carrier transport purely by thermionic emission flow of current is highly influenced by the polarity of voltage applied.

In order to produce a non-rectifying contact the primary means of electron transfer must be tunnelling. It is seen from the WKB approximation (Equation 2.31) that probability of electron tunnelling is inversely exponentially dependent on the thickness and the square root of the barrier height. It has been seen in Equation 2.30 that the barrier height is largely dependent on the semiconductor bandgap but is also dependent on the choice of metal. Using a semiconductor layer structure with a small bandgap material for the cap layer it is possible to minimise the barrier height. By highly doping ($N_c \sim 4x10^{18}cm^{-3}$) the low bandgap semiconductor cap layer, it is possible to form a very thin depletion region at the cap surface. For these reasons, a low bandgap, highly doped cap layer is used in the epi-layer of a HEMT to minimise the ohmic contact resistance.

The contact resistance can be reduced by the choice of metals used to form the contact and by annealing. Annealing in the process of heating the ohmic metals to encourage diffusion of the metal, which increases the doping of the semiconductor material reducing the barrier width further. This will be discussed in Section 5.2.1.

2.3 Behaviour of a HEMT

It is possible to use the expressions that have been derived above to describe the simple characteristics of the HEMT. The charge control model first derived for the HEMT by Delagebeaudeuf and Linh will be studied first. [2.17] This was the basis for other more refined models, which study the situation in greater detail. [2.18,2.19]

2.3.1 Current in a HEMT

The schematic cross section of a HEMT consisting of a Schottky gate electrode placed between source and drain ohmic contacts is given in Figure 2.9. This consists of a gate electrode placed in between the source and drain contacts. At the surface of the semiconductor beneath the gate, a depletion region will form. This has the effect of increasing the gate length as experienced by the 2DEG from the physical gate length, L_{ph}, to effective length, L_{eff}. The transistor can be split into three sections. Section I is the region from the source contact to gate. Section II is the region under the gate (the effective gate length). Section III is the region from the gate to the drain.

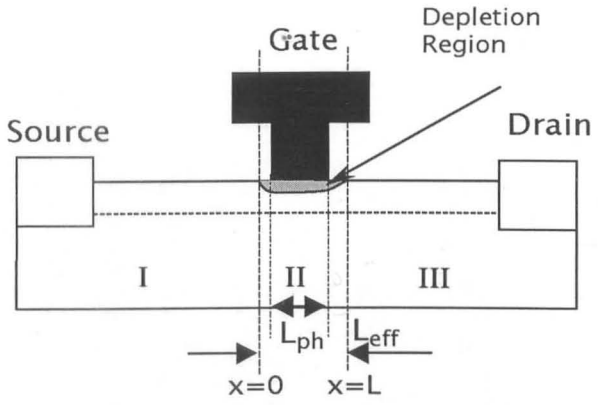


Figure 2.9 Schematic cross section of a HEMT. This source and drain ohmic contacts with a gate between them.

Using the above diagram and the equation derived for the electron sheet density (2.26) is possible to find an expression for the current that flows through the channel.

Equation 2.26 gives the electron density of the 2DEG as a function of the applied voltage between the gate and the channel. When there is no voltage between the source and the drain this will just be V_g. If however there is a voltage applied between the source and drain the voltage in the channel will depend on the position x. The effective gate voltage, V_{eff}, will then become

$$V_{eff} = V_g - V_c(x)$$

where $V_c(x)$ is the channel voltage at the point x. Substituting into (2.26) gives

$$n_D = \frac{\varepsilon(V_g - V_c(x) - V_T)}{q(d + s + \frac{a}{4} + h)} \qquad (2.33)$$

The current that will flow through the channel will depend on the electron density, the electron drift velocity in the channel, $v(x)$, and the width of the HEMT, W. It can be written as

$$I = qn_D(x)Wv(x) \qquad (2.34)$$

The electron velocity within the 2DEG is dependent on the applied electric field. However the dependency is not simple. At low electric fields (below 2 kV cm^{-1}) the electron velocity rises rapidly with increasing electric field with the rate of increase being the low field mobility. As the field increases the rate of change of velocity with field decrease, the effective mobility

decreases. This decrease continues and the differential mobility can become negative. This behaviour is shown in Figure 2.10 taken from Masselink et al. [2.20]

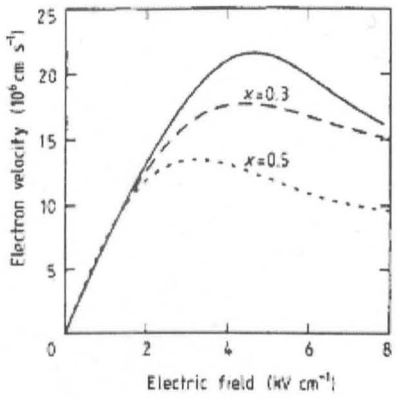


Figure 2.10 Electron Velocity versus electric field in bulk GaAs (full line) and GaAs/$Al_xGa_{1-x}As$ heterostructures with x=0.3 and 0.5.

The electron velocity dependence on electric field can be modelled to the first order by the following assumptions. At low electric field the electron velocity is given by the product of the low field mobility and the applied field. This situation continues until a certain field strength then the velocity saturates becoming independent of applied field. i.e.

$\upsilon = \mu F$ $F < F_c$ Low Field Region

$\upsilon = \upsilon_s$ $F > F_c$ Saturation Region

Where F_c is the critical field, and μ is the low field mobility.

In the low field region, the channel current can be found by substituting (2.33) into (2.34), using the simplification that the channel voltage $V_c(x)$ is small and its gradient can be approximated as the potential difference across the gate divided by the gate length, L.

$$I = \frac{\mu W \varepsilon (V_g - V_T - V_c(x))}{(d + s + \frac{a}{4} + h)} \frac{dV_c}{dx} \tag{2.35}$$

$$= \frac{\mu W \varepsilon (V_g - V_T)}{(d + s + \frac{a}{4} + h)} \frac{V_c(L) - V_c(0)}{L}$$

The channel voltage on either side of the gate is given by

$V_c(0) = R_s I$ (2.36)

$V_c(L) = V_D - R_d I$

Where R_s and R_d are the source and drain access resistances respectively.

Hence the low field resistance is given by

$$\frac{V_D}{I} = R_s + R_d + \frac{L(d + s + \frac{a}{4} + h)}{\mu W \varepsilon (V_g - V_T)} \tag{2.37}$$

In the high field region, the channel current becomes independent of the mobility and is determined by the saturation velocity of the channel. Equation 2.37 can be integrated and rearranged to give the channel voltage, $V_c(x)$.

$$V_c(x) = V_G - V_T - \sqrt{(V_G - V_T - V_C(0))^2 - \frac{2(d + s + \frac{a}{4} + h)Ix}{\mu W \varepsilon}} \tag{2.38}$$

By differentiating (2.38) and substituting (2.36), an expression for the electric field, $F(x)$, can be found.

$$F(x) = \frac{(d + s + \frac{a}{4} + h)}{\mu W \varepsilon}\left((V_G - V_T - R_s I_s)^2 - \frac{2(d + s + \frac{a}{4} + h)Ix}{\mu W \varepsilon}\right)^{-\frac{1}{2}}$$

Assuming that current saturation occurs when the field at the drain side of the gate ($x=L$) reaches the critical field, F_c the equation for the channel current is given as

$$I_s = \frac{W \varepsilon v_s}{(d + s + \frac{a}{4} + h)}\left(\sqrt{F_c^2 L^2 + (V_g - V_T - R_s I_s)^2} - F_c L\right) \tag{2.39}$$

Notice that the current is dependent on the both the gate width and gate length.

2.3.2 Transconductance

The Transconductance of a transistor, g_{mo}, is a measure of how much the channel current I_{ds} is modulated by the application of a gate voltage in the saturation region, this can be written as

$$g_{mo} = \frac{dI_{ds}}{dV_{gs}} = \frac{W \varepsilon v_s}{(d + s + \frac{a}{4})} \frac{V_{gs} - V_c(0) - V_T}{\sqrt{(V_{gs} - V_c(0) - V_T) + (F_c L)^2}} \tag{2.40}$$

This term also depends on the gate width and inversely proportional to the gate length and the gate-channel separation. The Transconductance of a device is very important as it will be seen shortly how it determines the gain of a transistor and subsequently how it relates to the RF performance of a device.

2.3.3 Output Conductance

The second important dc term that relates the channel current to an applied voltage is the Output Conductance. This is a measure of the dependence of the channel current within the saturation region on the applied source drain voltage. As the source drain voltage increases the electric fields within the channel increase, particularly at the drain end of the gate. The electrons within the channel heated by these large electric fields are therefore able to pass into the buffer layer, adding to the channel current. The simple model described above does not account for the electric fields and is therefore inadequate to describe the observed output conductance. The general behaviour of the HEMT in the saturation region is that the channel current will increase linearly with the source drain voltage, the gradient of this increase is the output conductance and can be written.

$$g_{ds} = \frac{dI_{ds}}{dV_{ds}} \tag{2.41}$$

2.3.4 Breakdown

Figure 2.11 shows the output characteristics of a HEMT, it is seen that the behaviour can be split into 3 sections, the low field, saturation and breakdown regime. The first two regimes have been described in section 2.3.1 however the simple model that was used is inadequate to describe the effect of breakdown.

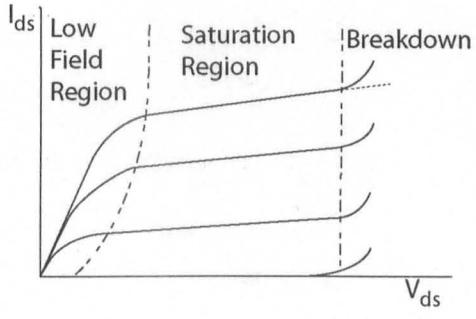


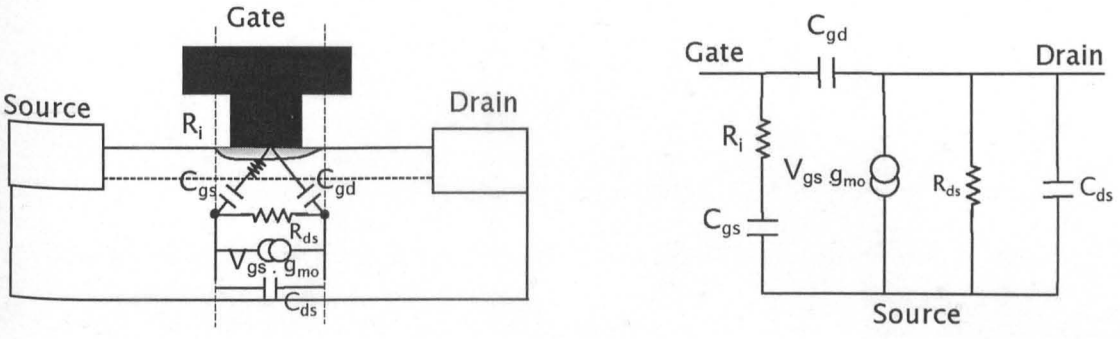
Figure 2.11 Illustration of output characteristic of a HEMT showing, low field, saturation and breakdown regions.

A HEMT is said to have entered the breakdown region when the constant output conductance that characterises the saturation region gives way to a increasing output conductance leading to a large increase in the current. The point at which this happens is called the breakdown voltage. Two breakdown voltages are defined, the on-state breakdown, BV_{on} and the off–state breakdown, BV_{off}, when the channel is open and pinched off respectively. BV_{on} which is defined in [2.21] as the voltage at which the drain current, I_{DS}, exceeds the maximum value of the linear region of I_{DSmax}. This definition although illustrative on the output characteristic plot is somewhat ambiguous, for instance, particularly for non-ideal devices where I_{DSmax} is not obvious. BV_{off} defined as the maximum value of V_{DS} that can be applied whilst not exceeding a certain current [2.22]. The mechanisms that cause breakdown in each state are also thought to be different. In off-state the mechanism is thought to be thermionic emission, from the gate to the drain electrode [2.23,2.24]. The dominant mechanism in the on-state is due to holes created by impact ionisation travelling through the barrier layers to the gate electrode [2.22]. The internal mechanisms are believed to be identical for both AlGaAs/GaAs and InGaAs/InAlAs HEMTs, however AlGaAs/GaAs have higher breakdown voltages, BV_{on}, and BV_{off} due to the higher Schottky barrier height of AlGaAs compared to InAlAs, and the higher impact ionisation rate in the InGaAs/InAlAs devices, due to a smaller bandgap.

2.3.5 *Small Signal Equivalent Circuit*

The preceding sections have described the DC current voltage behaviour of a HEMT. This is very useful in understanding the operation of the HEMT but it is of limited use in the description of the HEMT at the microwave and millimetre wave frequencies. It is however possible to produce a simple lumped element circuit to model the RF behaviour of the HEMT. This model uses as it basis the current voltage relationships derived in 2.3.2 and 2.3.3 together with a consideration of the physical structure of the HEMT. The construction of the model is described below.

The core for the small signal model for a HEMT is the voltage controlled current source. This is described by the transconductance of the transistor, or in other words, the change in drain voltage caused by a small gate voltage modulation. In parallel to this voltage controlled current source is the output conductance, g_{ds}. In the circuit this is modelled by a resistor, R_{ds}, connected across the current source, the value of which is equal to the inverse of the output conductance. Previously it has been noticed that the gate region of the device resembles a parallel plate capacitor. This capacitance can be broken in two section, the gate – source capacitance, C_{gs}, and the capacitance between the gate and the drain, C_{gd}. The relative values of these capacitances will depend on the detailed geometry of the depletion region under the gate, which is dependent on the applied bias. The formation of the model for the GaAs MESFET is described in greater detail in [2.24], and this analysis also holds for the HEMT. Figure 2.12(a) shows the cross section that was shown in Figure 2.9 with the addition of the intrinsic elements of the model. In addition to the voltage controlled current source, the output conductance and gate capacitances there will also be an input resistances, R_i, this arises from the resistance of the channel in the intrinsic region of the gate. As well as the capacitance associated with the gate there will also be capacitance between the source and the drain, C_{ds}. Figure 2.12(b) shows this circuit presented in schematic form.



(a) (b)

Figure 2.12 (a) Circuit showing a voltage controlled current source with input resistance, R_{in}, and a gain g_{mo}, connected to resistor R_{ds} modelling output conductance. (b) Voltage controlled current source with the addition of capacitances associated with the channel.

Parasistic Elements

The elements of the equivalent circuit considered so far have been the intrinsic elements associated with region of the device under the effective gate length, L_{eff}, in Figure 2.9. In order to model accurately its behaviour, the whole device needs to be considered. There will be additional access resistances and extra inductances arising from the metal structures on the surface such as the ohmic, and gate contacts and the RF bondpads.

The intrinsic device model shown in Figure 2.12(b) must now be modified to account for the parasitic elements of the circuit. The source and drain will have a series resistance associated with them, coming from the ohmic contact resistance in addition to the access resistance through the semiconductor. The gate will also have an associated resistance; this will be dependent on the physical geometry of the gate.

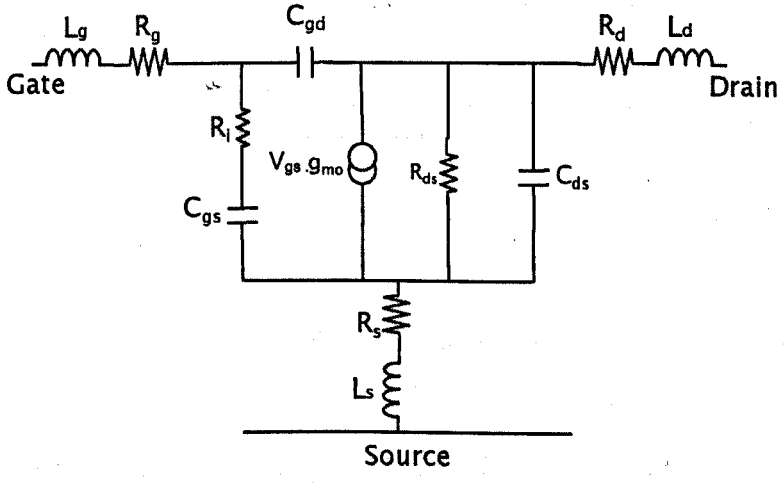


Figure 2.13 Lumped Element circuit with parasitic resistances and inductances.

In addition to the parasitic resistances, each terminal will have an associated inductance, arising from the metal structures of the source, drain and gate. These inductances will be connected in series with the parasitic resistances. When these parasitic elements are added to the intrinsic circuit the model shown in Figure 2.13 is produced. This is the simplified extrinsic model of the HEMT, and can be used to model the small signal behaviour of the HEMT at high frequencies.

It is possible to refine this model (further) by taking into account for example the capacitances that arise between the metal structures (pads) on the surface that make up the HEMT, and the capacitance of the gate feeds. A delay time is also incorporated into the voltage controlled current source in order to account for the propagation of the signal along the width of the gate. This model, shown in Figure 2.14 is widely used to model the behaviour of HEMTs at frequencies in up to 100 GHz.

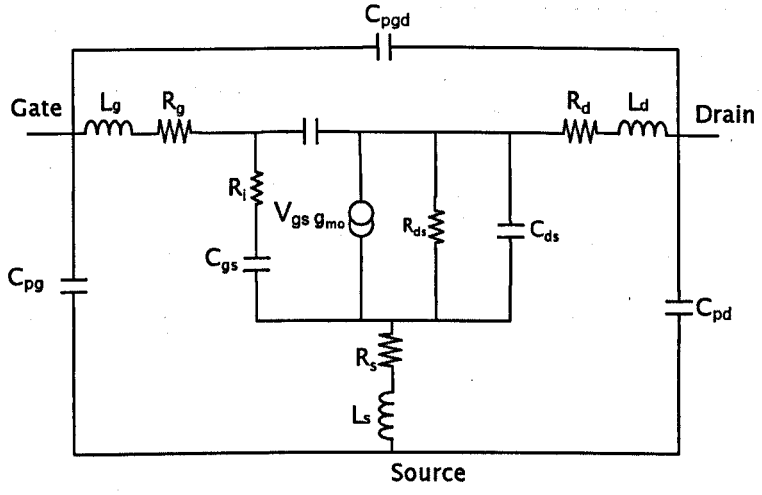


Figure 2.14 Full small signal model including parasitic drain capacitance, C_{pd}, parasitic gate capacitance, C_{pg}, and parasitics capacitance between the gate and drain, C_{pgd}.

Effects of the Parasitic Resistances

In the previous section parasitic resistances were introduced, in this section the impact of these resistances will be discussed.

Consider the model for the intrinsic HEMT at DC with the addition of the parasitic resistances R_s, R_d and R_g, . Assume that there are external voltages applied to the circuit, V_{gs} is connected between the gate and the source and V_{ds}, is connected between the drain and the source of the device. This situation is illustrated in Figure 2.15

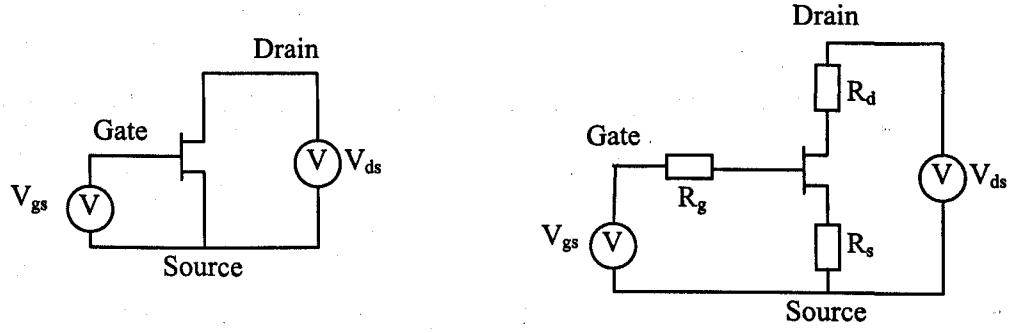


Figure 2.15 (a) HEMT biased in common source configuration. (b) With Parasitic Resistances included.

The applied voltages can be written

$$V_{ds} = V_d' + I_{ds}(R_s + R_d) \qquad (2.42)$$

and $$V_{gs} = V_g' + I_{ds}R_s + R_g I_{gs}$$

where V_d` and $V_g$` are the voltages across the intrinsic device. By differentiating (2.42) and substituting $g_{mo} = \frac{dI_{ds}}{dV_{gs}}$ the expression for the extrinsic transconductance, g_{mext}, is found to be

$$g_{mext} = \frac{g_{mo}}{1 + g_{mo}R_s} \qquad (2.43)$$

hence as the source resistance decreases the extrinsic transconductance decreases. The effect of this will be seen in the following section.

2.3.6 *Figures of Merit*

It is often beneficial to be able to quickly compare different transistors without examining their performance in great detail. For this purpose various Figures of Merit. (FOM) are used.

A number of these have already been discussed previously in the description of the behaviour of the HEMT - this section will provide a brief review.

Transconductance: This is a measure of the variation of the channel current with respect to the applied gate voltage.

Output Conductance: This is a measure of how the channel current varies with the applied drain voltage in the saturation.

Threshold Voltage: The threshold voltage is the voltage that needs to be applied to the gate to fully deplete the 2DEG.

Breakdown Voltage: The voltage at which the saturation region ends and breakdown starts.

Further DC figures of merit that have not been met yet.

Channel Current Density: This is the measure of the current that flows in the channel at a particular biasing condition, normalised to be independent of the width of the transistor and is usually defined in mA/mm. This is bias dependent so the bias condition is also given.

Voltage Gain, A_v, is a measure of the variation of the drain voltage with small modulation of the device and is defined as

$$A_v = \frac{\partial V_{ds}}{\partial V_{gs}}$$

It is seen that from the above equations for g_{mo}, and g_{ds} that

$$A_v = \frac{g_{mo}}{g_{ds}}. \qquad (2.44)$$

It can be seen that for a large voltage gain to be achieved a large transconductance and low output conductance is required. Although this equation is a good indicator of voltage gain at low frequencies it is of limited use at higher frequencies. However from this and considering the physical structure of the transistor it is possible to develop a lumped element circuit that can model adequately the behaviour of transistor at RF frequencies and above.

2.3.7 RF Figures of Merit

Figures of Merit are also widely used to compare the high speed performance of transistors at microwave frequencies and above. As with the DC FOM's there are a number, however, there are two main figures of merit used throughout this work. These are the cut-off frequency, f_T, and the maximum frequency of operation, f_{max}.

Before the RF figures of merit can be studied the behaviour of the electrical signals at high frequencies must first be briefly reviewed and the parameters that are used to measure devices must be introduced.

At RF frequencies the wavelength of the signal becomes comparable with the length of the device. The phase of the current and voltage signals depend where on the device the measurement is taken. This means that the standard circuit theory that is used to analyse circuits at low frequencies becomes inadequate. It is possible for the circuit to be analysed using Maxwell's equations and considering the electric and magnetic fields throughout the device. However the solutions to this problem can be very complex and time consuming for all but the simplest devices.

Transmission line theory provides a bridge between the full electromagnetic analysis and the standard circuit theory. This describes the voltage and the current throughout a device to be a waveform composed of the forward and backward travelling waves. The details of the theory will not be discussed here (a detailed description is given by Pozar [2.25]). However the main results will be briefly reviewed.

The main point that differentiates the transmission line theory from standard circuit theory is the characteristic impedance of a transmission line and the reflections caused by a unmatched load. The characteristic impedance, Z_0, of a transmission line depends on the capacitance and inductance of the line per unit length. In the case of Coplanar waveguides used as the terminals of a HEMT this arises from the physical geometry of the waveguides. If a load with impedance, Z_L, is attached to end of the line some of the voltage wave will be reflected back from the load. The ratio of the reflected to incident voltage is given by the voltage reflection coefficient, Γ given by [2.26]

$$\Gamma = \frac{Z_L - Z_0}{Z_L + Z_0} \qquad (2.45)$$

The maximum power transfer will occur when there is zero reflection which occurs when the load impedance is equal to the characteristic impedance of the line and equal to the source impedance. There are two other cases to note, namely when the load is either a short ($Z_L=0$) or open ($Z_L=\infty$) circuit, in this case the reflection coefficient is unity, and standing waves are set up along the line. On a transmission line terminated by a mismatched load, the impedance of the line repeats (periodically) every $\lambda/2$ where λ is the wavelength of the signal on the line. It is possible therefore to choose a length of line that will provide either maximum or minimum reflection. However, due to the dependence on the wavelength, it is not possible to use this approach to cover a wide frequency range as λ depends on the frequency. This is one of the problems overcome by the use of scattering parameters.

Two Port Networks and Scattering Parameters

To simplify measurements at microwave and millimetre-wave frequencies the device under test is considered to be a simple two port network as shown in Figure 2.16. The device is is fully characterised by the measurement of parameters, which describe how the network responds to excitation of the two ports. The parameters measured can be based on various characteristics of the network such as conductance parameters (y-parameters), resistance parameters (z-parameters), and hybrid parameters which are a mix of both conductance and resistance parameters (h-parameters).

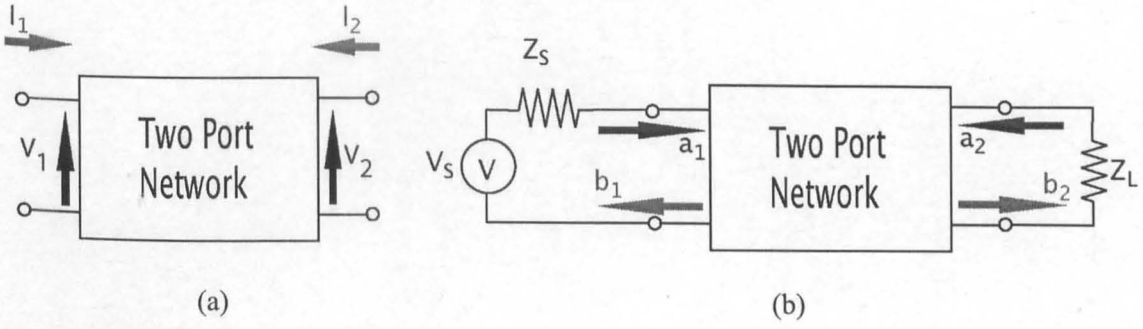


Figure 2.16 Illustration of a Two Port Network (a) With voltage and current as variables (b) Two Port network connected to a voltage

However these parameters require one of the ports to be opened and shorted during measurement. As the previous section has shown, this can be difficult to achieve, particularly

over a wide range of microwave and millimetre wave frequencies [2.27]. This problem can be overcome by using scattering parameters, or s-parameters.

S-parameters measure the scattering of the travelling waves at each of the terminals of the network. Figure 2.16(b) illustrates this, at the input (Port 1) of a two port network is a voltage source, V_s, and a load with an impedance of Z_L is connected at the output (Port 2). At port 1 there are two travelling waves the incident signal, a_1, and the reflected signal, b_1. Equally at port two there will be the incident wave into port 2, a_2, and the associated reflected wave, b_2. The independent variables can be related to their respective travelling voltage waves by the characteristic impedance of the ports, Z_0. For example the variables at port 1 can be written

$$a_1 = \frac{V_{i1}}{\sqrt{Z_0}} \qquad b_1 = \frac{V_{r1}}{\sqrt{Z_0}}$$

where V_{i1} and V_{r1} are the incident and reflective voltage waves at Port 1. The variables can be described with respect to each other in the following way.

$$b_1 = s_{11}a_1 + s_{12}a_2$$

$$b_2 = s_{22}a_2 + s_{21}a_1$$

The parameters s_{11}, s_{12} etc are the scattering parameters or s-parameters. They are defined as follows.

$s_{11} = \left.\frac{b_1}{a_1}\right|_{a_2=0}$ = Input Reflection Coefficient with the output port terminated by a matched load.

i.e $Z_L=Z_0$

$s_{22} = \left.\frac{b_2}{a_2}\right|_{a_1=0}$ = Output Reflection Coefficient with the input terminated with a matched load.

$s_{21} = \left.\frac{b_2}{a_1}\right|_{a_2=0}$ = Forward Transmission Coefficient with output terminated with a matched load.

$s_{12} = \left.\frac{b_1}{a_2}\right|_{a_1=0}$ = Reverse Transmission Coefficient with input terminated with a matched load.

The use of these scattering parameters allows the straightforward measurement of two-port systems with the advantage that they can be easily transformed into other parameters that give more information about the network under test.

Cut-off frequency, f_t

The cut-off frequency is also known as the gain-bandwidth product and defined as the frequency at which the short circuit current gain falls to unity. The short circuit current gain is defined as the ratio of the output current to the input current when the output is shorted. It is possible to use the intrinsic small signal model to find an expression for the cut-off frequency.

The short circuit current gain is given by

$$G_i^{SC} = \left| \frac{i_{out}}{i_{in}} \right|$$

from figure 2.11(b) with the output shorted where V is the input voltage applied between the gate and source.

$$G_i^{SC} = \left| \frac{g_{mext} V}{i_g} \right| = \frac{g_{mext}}{\omega(C_{gs} + C_{gd})} \qquad (2.46)$$

setting equal to unity and rearranging gives the equation for the cut-off frequency.

$$f_t = \frac{g_{mext}}{2\pi(C_{gs} + C_{gd})} \qquad (2.47)$$

It is seen that the cut-off frequency is inversely dependent on the gate capacitance. However it is seen in (2.43) that the extrinsic transconductance is largely dependent on the source access resistance of the device. If equation is substituted into (2.43) the cut-off frequency is

$$f_t = \frac{g_{mo}}{2\pi(1 + g_m R_s)(C_{gs} + C_{gd})} \qquad (2.48)$$

From this is clear that if the cut-off frequency of the device is to be maximised the gate capacitance and the source resistance must be minimised as well as maximising the transconductance.

The cut-off frequency of the device can be written in another way that hides the importance of the parasitic elements but is more informative about the high frequency behaviour of the device. If we consider the two port network shown in Figure 2.16 (a), the input (gate) current and output (drain) current may be written as.

$$i_{in}=y_{11}V_g + y_{12}V_d \qquad i_{out}= y_{21}V_g + y_{22}V_d$$

where y_{11}, y_{12}, etc... are the y-parameters. The short circuit current gain can now be written as

$$G_i^{SC} = \left|\frac{y_{21}}{y_{11}}\right| = h_{21} \qquad (2.49)$$

however, by reference to the transforms relating the y, h and s parameters that the short circuit current gain can be described by a single parameter h_{21}. This is seen to decay at –20dB/decade extrapolating to unity gives the cut-off frequency, f_t.

Maximum Frequency of Operation, f_{max}

The maximum frequency of operation is defined as the frequency at which the Unilateral Power, U, gain drops to unity. The unilateral power gain is a measure of the maximum gain that is achievable from a transistor. In order to achieve the maximum possible gain from a transistor the input and output ports must be conjugately matched to the input and output of the transistor, in order to eliminate any reflection caused by impedance mismatch. The input and output of the transistor must be isolated. This is not typical in a transistor, this is normally a degree of feedback provided by the gate-drain capacitance as well as other parasitic capacitances. A network must be added to cancel this feedback. In the technical literature it is not unusual for f_{max} to be taken at the point where the Maximum Available Gain (MAG) reaches unity. Although not technically correct the difference between the two method is very slight [2.28].

As with the cut-off frequency, f_t, it is possible to use the small signal model to derive an expression for f_{max}. The model that will be used is that given in Figure 2.12(b) with the gate-drain capacitance assumed to be zero to achieve the unilateral case. The power gain is defined as

$$G_P = G_V . G_I$$

where G_V and G_I are the voltage and current gain respectively. The voltage gain is defined as

$$G_V = \left|\frac{v_{out}}{v_{in}}\right|$$

Analysis of the small signal model shows that

$$\frac{v_{out}}{v_{in}} = \frac{g_m R_L}{1 + jw(C_{gs} + C_{gd})(R_i + R_g)}$$

Hence the voltage gain is

$$G_v = \left|\frac{v_{out}}{v_{in}}\right| = \frac{g_m R_L}{\left[1 + \omega^2 (C_{gs} + C_{gd})^2 (R_i + R_g)^2\right]^{1/2}}$$

When $\omega^2 (C_{gs}+C_{gd})(R_i+R_g) >> 1$ then

$$G_v = \frac{g_m R_L}{\omega(C_{gs} + C_{gd})(R_i + R_g)} \qquad (2.50)$$

Combining (2.46), (2.47) and (2.50) gives the following expression for the power gain

$$G_p = \frac{g_m R_L}{\omega C(R_i + R_g)} \cdot \frac{g_m}{\omega C} = \left(\frac{f_t}{f}\right)^2 \frac{R_L}{(R_i + R_g)}$$

Maximum power transfer to the load when the load resistance is equal to the output resistance, Hence

$$G_P = \left(\frac{f_t}{f}\right)^2 \frac{R_{ds}}{4(R_g + R_i)}$$

Setting the power gain equal to unity and rearranging for f, gives f_{max}.

$$f_{\max} = \frac{f_t}{2\left(\frac{R_g + R_i}{R_{ds}}\right)^{1/2}} \qquad (2.51)$$

It is seen that the maximum frequency of oscillation is strongly dependent on the cut-off frequency, f_t, and also the gate resistance. Usually a high f_T is achieved by scaling down the length of the gate which can result in an increase in the gate resistance, however for a high f_{max} the gate resistance must remain small. This gives rise to the characteristic shape of the T-gate used in the HEMT, the foot can be made very small to provide high f_T, the large head of the gate keeps the resistance of the gate low.

The full extrinsic small signal model can be analysed in the same way, and gives an expression for f_{max} as [2.29]

$$f_{\max} = \frac{f_t}{2\left(\frac{R_g + R_i + R_s}{R_{ds}} + 2\pi f_t R_g C_{gd}\right)} \tag{2.52}$$

In a similar manner to the current gain, the maximum available gain can also be expressed in terms of the admittance parameters, (y parameters)

$$G_{\max} = \left|\frac{y_{21}}{y_{12}}\right|\left(k - \sqrt{k^2 - 1}\right)$$

where k is the stability factor, given by

$$k = \frac{2\mathrm{Re}(y_{21})\mathrm{Re}(y_{22}) - \mathrm{Re}(y_{12}y_{21})}{|y_{12}y_{21}|}$$

When k>1 the transistor is unconditionally stable, and will not oscillate with any passive loads attached, this is normally only the case at very high frequencies (>100 GHz) for short gate length HEMTs usually k<1 and the transistors must be stabilised with shunt resistors on the input and output. The gain in this region is referred to as the maximum stable gain and is given by

$$G_{ms} = \left|\frac{y_{21}}{y_{12}}\right|$$

These two gains are shown in Figure 2.17 alongside the stability factor *k*. It is seen that the G_{max} curve decays at −10dB/decade in the MSG region, then as k=1 a knee forms and the slope increases. This is now the MAG region and the curve can be extrapolated to unity to determine f_{max}.

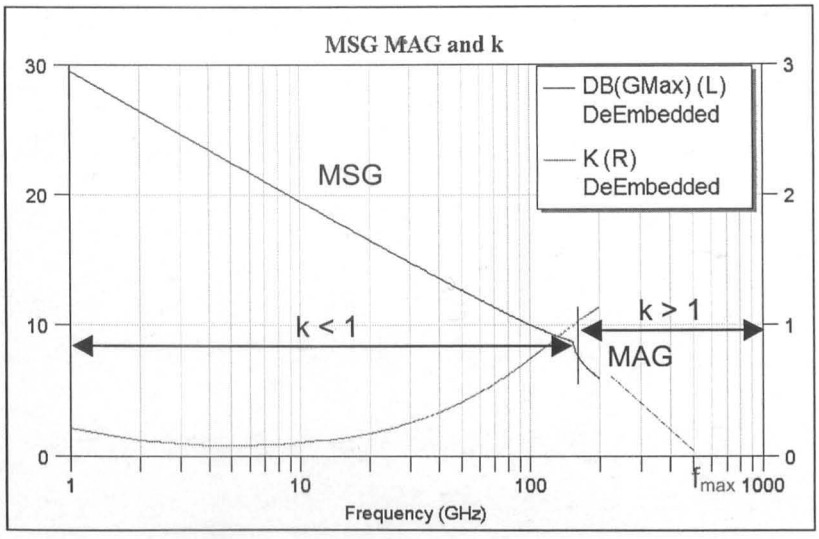


Figure 2.17 Plot of the maximum stable gain, MSG, and maximum available gain, MAG, plotted alongside the stability factor k. The determination of f_{max} by extrapolation of MAG is shown.

Extrapolation to find f_{max} requires a great deal of caution, since the HEMTs tend to be only conditionally stable except at very high frequencies. It is usually not possible to determine the frequency at which k>1 and thus determine MAG in the usual measurement range.

2.4 *Different Types of HEMTs*

The HEMT that has been described in the preceding sections is the GaAs/AlGaAs HEMT, the first type of HEMT fabricated [2.30]. Its simple structure makes it ideal to describe the general operation of the device. The performance of this type of transistor was slightly disappointing however, the increased mobility didn't translate into such large improvement in the RF performance. The maximum frequency of oscillation of the early AlGaAs/GaAs HEMTs were lower than equivalent GaAs MESFET's technology [2.31].

The lower than expected performance was for two related reasons. It has already been noted in section 2.3.1 that the electron velocity is very important in the behaviour of the HEMT. It influences the channel current and transconductance of the device. It was found that the peak electron velocity of the GaAs/AlGaAs heterostructure was lower than that in a GaAs MESFET [2.32].

Two effects cause this reduction in electron velocity. The first is by **k** space transfer. When an electron is placed in an electric field it will gain energy. The amount of energy gain will depend on the magnitude of the field, the mobility and the energy relaxation time of the material. As the electron gains energy they can pass from the lower energy Γ valley into the higher energy L valley. (The L valley has a lower band curvature than the Γ valley; this leads to a larger effective electron mass and hence lower effective mobility.) This effect occurs in many III-V semiconductors. However the effect is amplified in the selectively doped heterojunction because the confinement of the electrons with in the 2DEG effectively lower the energy separation between the Γ and L valleys from 0.31 to 0.27eV[2.33].

The second method that is responsible for the reduction of peak electron velocity in the MODFET structure is real space transfer. As the electrons within the channel are accelerated by the electric field produced by the source-drain bias they gain energy, it was stated in the previous paragraph that it was possible for these heated electrons to transfer into higher valleys in the GaAs. It is also possible; as the electrons get hotter that they are able to pass in to the L and Γ valleys within the barrier and spacer layers [2.34]. This is referred to as real space transfer. The mobility of the AlGaAs layer is much lower than GaAs, for example the undoped mobility of $Al_{0.3}Ga_{0.7}As$ is 3500 cm^2/Vs^{-1} compared to 8500 cm^2/Vs^{-1} for GaAs. The barrier layers are also highly doped which reduced the mobility and hence electron mobility further.

Due to the low conduction band offset of the GaAs/AlGaAs system early HEMTs were badly effected by real space transfer and its associated reduction of electron velocity.

The RF performance of GaAs/AlGaAs HEMT was also limited by the electron sheet density. The access resistance of the device is strongly influenced by the electron density of the channel. As (2.43) shows, the access resistance strongly influences the extrinsic transconductance. It has been seen above that the depth of the electron well and therefore the sheet density are dependent on the conduction band offset. The larger the conduction band offset of the interface the larger

sheet densities that are achieved leading to lower access resistances and hence the higher the extrinsic RF performance of the device.

It therefore advantageous to increase the conduction band offset of the heterojunction, (2.3) and (2.4) give the band offset for the conduction and valence band respectively. It is seen from these equations that the conduction band offset is dependent on the work function and the difference between the band gaps of the material. Figure 2.18 shows the dependence of the band gap of $Al_xGa_{1-x}As$ on the aluminium mole fraction, x, it is seen that the as the concentration of aluminium increases so does the energy gap. This seems to be a way of achieving a greater discontinuity.

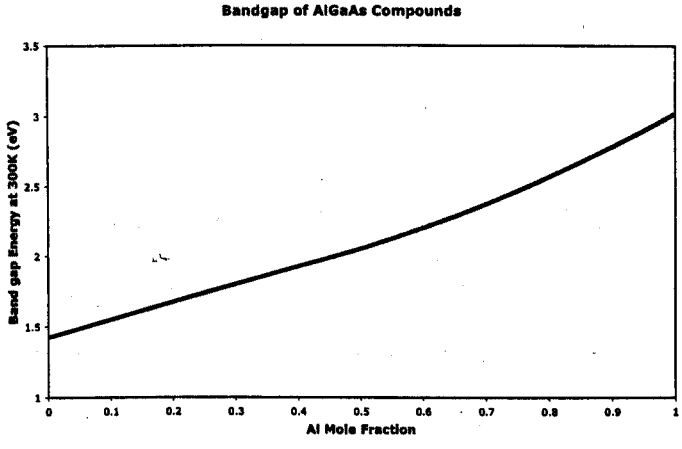


Figure 2.18 Bandgap of $Al_xGa_{1-x}As$ compounds as a function of the aluminium mole fraction, x.

However, when the aluminium content of the compound increases another problem arises that of DX centres.

DX Centres

DX centres are deeply bound donors within the $Al_xGa_{1-x}As$ caused by a large relaxation of the crystal lattice around the donor. This releases energy, which is then dissipated into the crystal [2.36]. The binding energy of this state is very large, up to 160meV, and increases with the aluminium content, x. [2.36] When $x > 0.22$ the DX centre state is the lowest energy state and

thus the stable state. [2.37] For a doped material at thermal equilibrium the free electron concentration is determined by the donor binding energy. The result of the increase in binding energy is that in $Al_xGa_{1-x}As$ as x is increased the fraction of donors which are ionised is greatly reduced.[2.38] Therefore to achieve the desired free electron concentration the number of doping atoms needs to be increased as x is increased. In addition, DX centres also cause a shift in the threshold voltage at low temperatures [2.39] and, transient shifts in the threshold voltage at room temperature.[2.40]

The existence of DX centres in $Al_xGa_{1-x}As$ limits the choice of aluminium fraction, the benefits of increased conduction band offset must be weighed against the problems of DX centres. Typically a value of 30% aluminium is used.

2.4.1 Pseudomorphic Growth

It was shown that as the indium content of InGaAs compounds is increased the band gap of the material decreases. [2.41] The variation of the energy gap versus the indium concentration is given in Figure 2.19 using the equation suggested in the previous reference. It is seen that the energy gap of the compound varies from a value of 1.425eV for GaAs (x=0) to 0.36eV for InAs (x=1).

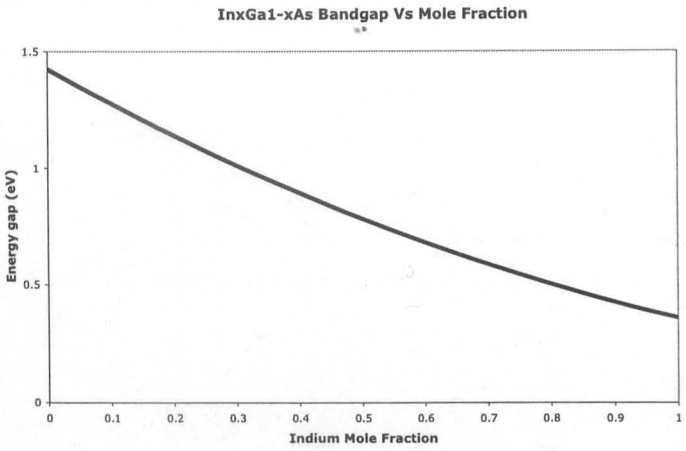


Figure 2.19 Energy Bandgap of $In_{x+}Ga_{1-x}As$ with respect to Indium mole fraction, x. Based on work by Nahory et al. [2.41]

Unlike $Al_xGa_{1-x}As$, $In_xGa_{1-x}As$ shows a clear lattice constant dependence on relative composition of indium and gallium. Figure 2.20 shows the energy gap for InGaAs and AlGaAs as a function of the lattice constant. The lattice constants were calculated assuming Vegard's law.

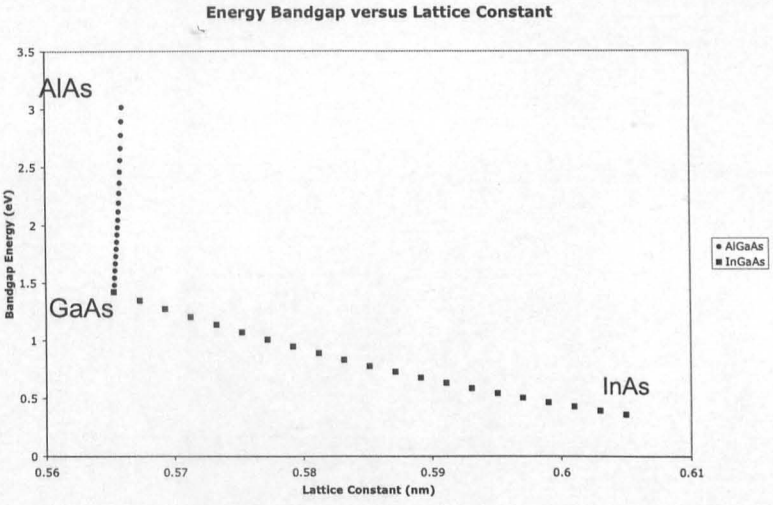


Figure 2.20 Energy Bandgap of InGaAs and AlGaAs as a function of lattice constant.

The graph shows that as the indium concentration is increased to give a greater conduction band offset, the lattice constant quickly changes significantly. Hence the growth of InGaAs layers on GaAs is not straightforward. It is however possible. It was suggested [2.42] that if the layers were thinner than a critical thickness, L_c, the lattice mismatch can be accommodated. This

critical thickness is dependent on the strain, which, in turn is dependent on the indium concentration.[2.43]

The process that allows this accommodation of differing strain is as follows. When a layer of material is grown on a substrate, the grown material is able to expand or contract to match the lattice parameter of the substrate material. This is illustrated in Figure 2.21. In this situation, the lattice constant of the grown material is larger than the substrate, and the lattice is compressed to accommodate the mismatch.

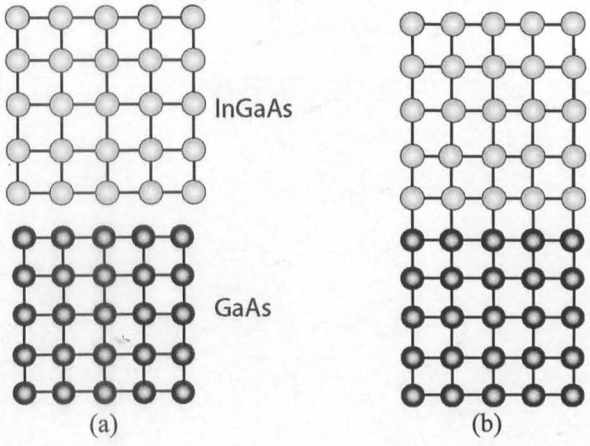


Figure 2.21 Illustration of the growth of mismatched crystal lattice (a) Separate crystals with dis-similar lattice constants. (b) The grown material (InGaAs) distorts its lattice to fit that of the substrate lattice constant

As the thickness of the layer increases the effects of the strain within the grown layer increases which causes defects such as misfit dislocations in the crystal. These dislocations interrupt the periodic structure of the crystal. This has been shown to degrade electron transport in such a structure.[2.44,2.45] This method of growth is referred to as Pseudomorphic growth.

2.4.2 *Lattice Matched Growth*

Figure 2.20 shows that it is possible to grow any composition of $Al_xGa_{1-x}As$ on a GaAs with almost no lattice mismatch. This case is referred to as lattice matched growth. Figure 2.22 shows the bandgap versus lattice constant of $In_xGa_{1-x}As$, $In_yAl_{1-y}As$ and InP. It is seen that for a

composition of around 50% the lattice constant of InGaAs and InAlAs are very close to that of InP.

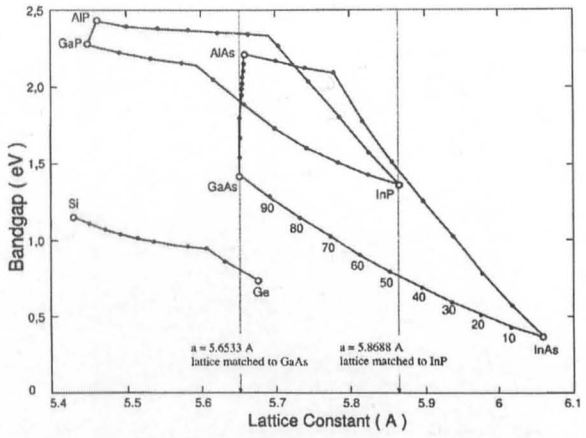


Figure 2.22 Bandgap versus Lattice Constant for a number of common semiconductor compounds including InGaAs, InAlAs an InP

For the compounds $In_{0.53}Ga_{0.47}As$ and $In_{0.48}Al_{0.52}As$ the lattice parameter is identical to that of InP. Figure 2.22 also shows that the difference in the bandgap of the two materials is very large compared with that of the GaAs/AlGaAs system studied previously. This makes the InGaAs/InAlAs structure very important in the realisation of high speed HEMTs.

2.5 *Conclusions*

This chapter has introduced the High Mobility Transistor (HEMT), its physical structure, and the theory of its operation. The electrical behaviour of the HEMT was discussed including the introduction of a number of figures of merit, which allow various transistors to be discussed and compared. The work introduced in this chapter will be used throughout the remaining chapters of this thesis.

2.6 References

[2.1] W.Shockley, *"The Path to conception of the Junction Transistor"*, IEEE Trans. Elec. Dev. Vol. ED-23, No.7,1976.

[2.2] A.G. Milnes and D.L. Feucht, *"Heterostructures and Metal-Semiconductor Junctions"*, London, Academic Press, 1972.

[2.3] R.L. Anderson, *Germanium-gallium arsenide Heterojunctions*, IBM Journal of Research and Development, 4, p283.

[2.4] L. Esaki and L.L. Chang, *"New Transport Phenomenon in a Semiconductor "Superlattice""*, Physics Review Letters Vol. 33, 1975. p495,

[2.5] D. Dingle et al., *"Electron mobilities in modulation doped semiconductor Heterojunction superlattices"*, Applied Physics Letters, Vol 33, 1978. p665,

[2.6] F. Stern and S.D. Sarma, *"Electron levels in GaAs-GaAlAs Heterojunctions"*, Physics Review B, **30**, 1984, p840

[2.7] R.H. Kingston, *"Semiconductor Surface Physics"*, Philadelphia, University of Pensylvania Press,1957

[2.8] Crowell et al., Applied Physics Letters, Vol. 4, p91, 1964

[2.9] J.H. Davies, *The Physics of Low-Dimensional Semiconductors, Cambridge, Cambridge University Press, 1998, p336*

[2.10] F. F. Fang and W. E. Howard, *Negative Field-Effect Mobility on (100) Si Surfaces*, Phys. Rev. Lett. 16, 1966, p797

[2.11] W. Schottky, Zeits f. Physik, Vol. 41, 1940, p570

[2.12] W. Schottky, Zeits f. Physik, Vol. 118, 1942, p539

[2.13] J. Bardeen, *Surface States and rectification at metal-semiconductor contact*, Phys. Rev. 71, 1947, p714

[2.14] A.M. Cowley, S.M. Sze, *Surface States and Barrier Height of Metal Semiconductor Systems*, Journal of Applied Physics, Vol.36, 1965, p3212

[2.15] J.H. Davies, "*The Physics of Low-Dimensional Semiconductors*", Cambridge, Cambridge University Press, *1998, p269*

[2.16] H.A. Bethe, "*Theory of Boundary Layer of crystal rectifiers*", MIT Radiation Report, 1942, p12-44

[2.17] D. Delagebeaudeuf and N.T. Linh, "*Metal-(n)AlGaAs-GaAs Two Dimensional Electron Gas FET*", IEEE Transactions on Electron Devices, Vol. ED-29, No. 6, 1982, p955

[2.18] K. Lee, "*Current-voltage and capacitance-voltage characteristics of modulation doped field effect transistors*", IEEE Trans. Elec. Dev., Vol ED-30, 1983, p207

[2.19] R. Gupta et al., "*A new depletion dependent analytical model for sheet carrier density of InAlAs/InGaAs heterostructure, InP based HEMT*", Solid-State Physics, Vol. 47, 2003, p 33,

[2.20] W.T. Masselink, "*Electron Velocity in GaAs : bulk and selectively doped heterostructures*", Semiconductor Sci. Tech, **4,** 1989, p508

[2.21] J. Dickmann et al. "*Breakdown Mechanisms in Psuedomorphic InAlAs/$In_xGa_{1-x}As$ High Electron Mobility Transistors on InP II : On State*", JJAP, **34,** 1995, p1805.

[2.22] J. Dickmann et al."*Breakdown Mechanisms in Psuedomorphic InAlAs/$In_xGa_{1-x}As$ High Electron Mobility Transistors on InP I : Off State*", JJAP, **34,** 1995, p66.

[2.23] G. Meneghesso et al., "*On state and Off state Breakdown in GaInAs/InP Composite channel HEMT's with variable GaInAs channel thickness*" IEEE Trans. Elec. Dev., **46,** 1999, p6

[2.24] P. Ladbrooke, *"MMIC Design GaAs MESFETs and HEMTs"* , Boston, Artech House, 1989 p112

[2.25] D. Pozar, "Microwave Engineering", New York, John Wiley & Sons, 1998, p56-103

[2.26] D. Pozar, "Microwave Engineering", New York, John Wiley & Sons, 1998, p66

[2.27] Hewlett Packard Test and Measurement Application Notes, 95-1

[2.28] F. Schwierz and J.J. Liou, "*Semiconductor Devices for RF Application: evolution and current status.*:, Microelectronics Reliability **41**, 2001, p145

[2.29] S. M. SZE, "*Physics of Semiconductor Devices*"

[2.30] T. Mimura et al., "A New field effect transistor with selectively doped GaAs/n-$Al_xGa_{1-x}As$ heterojuncttions" JJAP, **19:L**, 1980, p225

[2.31] F. Schwierz, J.J. Liou, "*Semiconductor Devices for RF Application: evolution and current status.*:, Microelectronics Reliability **41**, 2001, p149

[2.32] W.T. Masselink, "*Electron Velocity in GaAs : bulk and selectively doped heterostructures*", Semiconductor Sci. Tech, **4**, 1989, p509

[2.33] W.T. Masselink, "*Electron Velocity in GaAs : bulk and selectively doped heterostructure"s,* Semiconductor Sci. Tech, **4**, 1989, p510

[2.34] R.Menozzi, "*Hot Electron effects and degradation of GaAs and InP HEMTs for microwave and millimetre-wave applications*", Semiconductor Science and Technologies, 13,(1998), p1083

[2.35] J.H. Davies, "*The Physics of Low-Dimensional Semiconductors", Cambridge, Cambridge University Press, 1998, p337*

[2.36] N. Chand, et al., "Comprehensive analysis of Si-doped Al xGa 1- xAs (x=0 to 1): Theory and experiments " Phys. Rev. B, **30**, 1984, p4481,

[2.37] P.M. Mooney, *"Deep donor levels (DX centers) in III-V semiconductors"*, J. Appl. Phys, **67**, 1990, R1

[2.38] T.F. Kuech, et al. *"Silicon doping of GaAs and Al xGa 1- xAs using disilane in metalorganic chemical vapor deposition*", Journal of Crystal Growth, **68**, 1984,, p48

[2.39] P.M. Mooney et al. Inst. Phys. Conf. Ser., **74**, 1984, p623

[2.40] M.I. Nathan et al., "*Room-temperature electron trapping in Al 0.35 Ga 0.65 As/GaAs modulation-doped field-effect transistors*" Appl. Phys. Lett., **47**, 1985 p628

[2.41] R.E. Nahory et al., "*Bandgap versus composition and demonstration of Vergard's law for $In_{1-x}Ga_xAs_yP_{1-y}$ lattice matched to InP*", Appl. Phys. Lett. **33**, 1978, p661

[2.42] J.W. Matthews and A.E. Blakeslee "*Defects in epitaxial multilayers *1 : I. Misfit dislocations*" Journal of Cyrstal Growth, 27, 1974, p118

[2.43] I.J. Fritz et al.,*"Dependence of critical layer thickness on strain for $In_xGa_{1-x}As/GaAs$ strained layer superlattices", Applied Physics Letters, 46, 1985, p967*

[2.44] R. Hsing et al., "*Influence of misfit dislocations on the mobility in pseudomorphic high electron mobility transistors based on In xAl 1-x As / In 0.75 Ga 0.25 As / InP structures.* ", Proc. 42nd Elec. Mat. Conf., 2002

[2.45] D. Zhao and K. Kuhn, *"Dislocation Scattering in n-Type Modulation Doped $Al_{0.3}Ga_{0.7}As/In_xGa_{1-x}As/Al_{0.3}Ga_{0.7}As$ Quantum Wells" Trans. Elec. Dev., 38, p2582, 1991*

3.0 Review of Current Research

As was seen in chapter 2, most III-V HEMTs utilise an $In_xGa_{1-x}As$ channel to provide enhanced transport properties, increased carrier density and hence improved high-speed performance. However, the composition of the channel, the materials used in the heterojunction and the substrate used vary. In this review of the current technologies, the types of HEMT will be split into four according to the channel composition and the substrate on which the active layers are grown. These are as follows:

Pseudomorphic HEMT grown on GaAs substrate (GaAs pHEMT), this is a device that has a strained $In_xGa_{1-x}As$ channel with x typically in the range of 20-30%.

Metamorphic HEMT on GaAs (mHEMT) this device is also grown upon a GaAs substrate but utilises a strain accommodating thick buffer layer which allows an unstrained channel without constraint on the indium content.

Lattice Matched HEMT on InP (LMHEMT) has an unstrained $In_{0.53}Ga_{0.47}As$ channel grown on an InP substrate.

Pseudomorphic HEMT on InP. (InP pHEMT) This has a strained channel, using strained material on InP substrates and allows an indium concentration in excess of 53%. InP pHEMTs have been realised with an indium concentration of 100% i.e.an InAs channel [3.1].

In addition to these there are other types of HEMT under active investigation. These include GaN based HEMTs for high power applications, [3.2,3.3], SiGe HEMTs on Si [3.4,3.5] as well as the growth of metamorphic HEMTs grown on a Ge substrate. However they are not considered in this review, which is limited to GaAs and InP based technologies.

3.1 GaAs pHEMT

The GaAs pHEMT has been developed over many years since its initial development in the late 1980's. In this early period the record value for the cut-off frequency, f_t, was in the 110GHz [3.6] to 150GHz [3.7] range for 0.1μm gate length devices. There was then an extended period of consolidation as studies were carried out on the reliability of the pHEMT [3.8-3.11]. The technology was used to build Microwave Monolithic Integrated Circuits (MMIC) operating in the Ka band (26-40GHz) [3.12]. There are now several industrial foundries with automated processing that use GaAs pHEMT wafers with a diameter up to 6 inches.

The majority of the development now being carried on the GaAs pHEMT technology is less concerned with the ultra high speed aspect of the devices but the improvement of the power characteristics.[3.12,3.13].

Over the past 15 years the improvement of the high speed performance has been very slight. This is due to the limitation of the material system, namely the small conduction band offset and the peak electron velocity imposed by the indium content of the channel. There has been some improvement however recently by the aggressive scaling of the gate length. The current record for the cut-off frequency for a GaAs pHEMT is 200 GHz. This work was performed by the author together with colleagues in Glasgow in 2003. These devices used a 50nm T-shaped gate and a scaled layer structure to minimise short channel effects [3.14]. The highest reported maximum frequency of oscillation, f_{max}, is 290GHz [3.15], a record which stands from 1990.

3.2 Metamorphic on GaAs

Over the past 15 years there has been extensive research into metamorphic devices on GaAs substrates. This is because it is a very promising technology that is capable of growing any composition of InGaAs on a GaAs substrate without the introduction of strain in the active layers. This promises to give the performance of InP devices but without the need for small,

expensive, brittle InP substrates. In the development of metamorphic devices there has been numerous areas of study in order to improve the performance of the devices. Some of these are: the Indium content of the channel and how this affects the high speed performance. [3.16] Different types of buffer layer have been tested in order to provide the best confinement of the dislocations within the buffer layer. Some of these have included step grading the buffer [3.17] linearly grading the buffer [3.18], and using different compound for the material of the buffer such as AlGaAsSb [3.19]. There has also been work done with creating an over shoot in the Indium content within the buffer and then stepping back down to the required composition. This has been shown to reduce the dislocation density within the channel [3.20-22].

This extensive research has improved the quality of the material grown and new fabrication processes have been developed that give an enhanced performance. The performance of the metamorphic HEMT is now comparable with those devices based on InP substrates for a similar gate length. Recent results have been published using a metamorphic layer structure include a metamorphic HEMT with a T shaped gate with a gate length of 50nm with a $In_{0.53}Ga_{0.47}As$ channel layer. This had a cut-off frequency, f_t, of 260Ghz, and a maximum oscillation frequency, f_{max}, of 490Ghz [3.23]. The record cut-off frequency was increased yet further by researchers in Glasgow with the fabrication of a HEMT with a gate length of 50nm and having an f_t, of 350GHz [2.34]. Similar performance was achieved by the author using an alternative process flow, as will be described in chapter 6. These recent devices currently have the highest reported operating frequencies for a transistor on a GaAs substrate.

3.3 Lattice Matched and Pseudomorphic HEMT's on InP

Despite the intense development and the impressive high speed performance of metamorphic HEMTs, lattice matched and pseudomorphic HEMTs grown on InP are still the material system of choice for ultimate high speed and low noise devices. This is because the InP based HEMT

has been in development as long as the GaAs pHEMT, and wafer growers have a long experience in producing very high quality layer structures. Since InP based layer systems do not have the strain caused by the lattice mis-match present in metamorphic layer it is easier to grow materials with very smooth interfaces. Rough interfaces introduce scattering that lowers the mobility [3.25]. The InP based systems have demonstrated the best high-speed performance and so often when a new shorter gate process is developed, it is the InP system that benefits from this technology first.

At present a large amount of research is being carried out into every aspect of the technology. This includes research into development of the commercial production of InP based devices, including the development of larger InP wafer sizes (4 and 6 inch) to increase production efficiency[3.26] and improvements in wafer mechanical strength [3.27] as fragility of the InP wafers has traditionally been a limitation in the production of InP devices. A large amount of work has been carried out in the fabrication of various MMIC circuits operating in excess of 100GHz such as the amplifier that operates in the frequency range between 150 and 205GHz [3.28] and amplifier designed to operate in the G-band using 70nm gate technology [3.29]. A large amount of work is also being carried out on developing InP based integrated circuits suitable for 40Gbit/s and above optoelectronic communications systems [3.30-33]. In addition to the factors relating to the design and manufacture of integrated circuits there is also related research concerned with the reliability of discrete devices and the physical mechanisms that cause the devices to fail [3.34], a topic of great interest for applications.

Despite the large amount of research being carried out with the purpose of commercialising the technology, the pace of improvement in high-speed device performance has not waned, in fact in recent years the pace has quickened. There is limited recent work published on the subject but an example of this development can be found in a brief review of the literature regarding fabricated devices. This shows that the typical contact resistance is very low, between about 0.07 to 0.15 Ω.mm using different metals and annealing strategies.[3.23,3.24,3.31,3.35-37] This

suggests that there has been a large amount of research tailoring ohmic strategies to the particular layer structures used.

There has also been work in the development of different etch processes for the gate recess. These tend to be selective, either stopping on InAlAs or InP in order to provide better threshold voltage control over the wafer. This selective etch is achieved either by wet chemical dry etching [3.38]. The different selective wet etch solutions are based on either succinic acid [3.39] or citric acid and tend to be pH balanced. The details of the selective gate etching will be discussed in greater detail in the following chapters. Research has also been carried on the use of a double recess in which a second narrow and shallow etch is used in the etch stop layer. The second recess plays its role in controlling the electric fields below the gate and can provide better charge control [3.40]. An additional benefit of this process is in the fabrication of ultra-small gate length devices. This will also be discussed later.

A major topic for device research in recent years has been the development of reduced gate length T shaped gates. Is was seen in section 2.3.2 that for a well scaled device the transconductance and hence f_t and f_{max} are inversely proportional to the gate length. It was first shown by Nguyen et al. in 1992 that ultra high frequency performance could be achieved by the use of 50nm T-shaped gate. Using such a gate process an $In_{0.8}Ga_{0.2}As/In_{0.52}Al_{0.48}As$ HEMT had an f_t of 340 GHz [3.41]. At this time the typical gate length was over 100nm. In the following years this performance was not surpassed with many researchers working to optimise their existing 100nm gate process and incrementally work down to smaller gate lengths [3.42]. However in 1998 Suemitsu and colleagues at NTT in Japan presented results of a 30nm gate length lattice matched HEMT on InP. This device had an f_t of 350 GHz [3.43] comparable to the results of Nguyen using an 80% pseudomorphic channel. This prompted other research labs to develop equivalent processes such as Chen et al. at Glasgow [3.44] who produced 30nm T-gates using single step UVIII/PMMA layer as opposed to the two stage ZEP520 and fullerene based process used by Suemitsu et al. However these were fabricated on a planar substrate rather than

of actual devices. Fujitusu Labs reported in 2001 the fabrication and device performance of a HEMT with a 25nm gate length with a cut-off frequency of 396GHz [3.45]. This still remains the shortest gate length T-gate reported. Since then work has been carried out to properly scale the devices in order to attain the performance that these ultra small gate lengths should provide. This has seen the cut-off frequency rise to 500 GHz for a lattice matched HEMT, and a record 562 GHz for a pseudomorphic HEMT on InP in 2002 [3.46]. This is the highest f_t reported for any transistor. The results presented in this recent work have concentrated on the cut-off frequency of the transistor; the maximum frequency of oscillation is only briefly mentioned. When figures are given for the f_{max} for modern devices it is around 500GHz. This is lower than the record value of 600GHz that has been achieved using 100nm gate length HEMTs [3.46]. Although it should be noted that the extrapolation used was questionable as the maximum stable gain (MSG) was extrapolated at –6dB from 110GHz. This doesn't take into account the knee and increased gradient of MSG that is usually witnessed at high frequencies. If modern devices with shorter gate lengths were extrapolated in this manner, the f_{max} obtained would be likely to exceed 1000GHz.

This review of the current research has shown the various methods that have been developed to realise high speed HEMTs. This included the use of different compositions of the channel layer grown of various substrates, and the reduction of gate length. The following chapter will introduce the methods and techniques that were used in this work to realise High Speed HEMTs.

3.4 References

[3.1] A. Chin et al. "*$In_{0.52}Al_{0.48}As/InAs/In\ xAl\ 1\text{-}x\ As$ pseudomorphic HEMT's on InP*", IEEE EDL, **18**, 1997, p157

[3.2] M.A. Khan et al., *"High electron mobility transistor based on GaN-$Al_xGa_{1-x}N$ heterojunction"* App. Phys. Lett., **63**, 1993, p1214

[3.3] Y.F. Wu et al., *"Measured Microwave Power Performance of AlGaN/GaN MODFET"*, IEEE Elect. Dev. Lett., **17**, 1996, p455.

[3.4] U. Koenig et al., *"Enhancement mode n-channel Si/SiGe MODFET with high intrinsic transconductance"*, Electronics Letters **28**, 1992, p160

[3.5] M. Zeuner et al. *"90GHz ft SiGe HFET with fully optical self-aligned sub 100nm gate"*, Proc. Device Research Conference, DRC technical Digest **II B 8**, 2002.

[3.6] A.N. Lepore, et al. *"0.1μm gatelength MODFET's with unity current gain cut off frequency above 110GHz"* Electronics Letters, **24**, 1988, p364

[3.7] L.D. Nguyen at al. *"Characterisation of ultra-highspeed pseudomorphicAlGaAs/InGaAs (on GaAs)sMODFET's"* Trans. Elec. Dev., **36**, p2243. 1989

[3.8] C.Canali et al., *"Enhancement and Degradation of Drain Current in Pseudomorhic AlGaAs/InGaAs HEMT's Induced by Hot Electrons"*, Proc. IEEE International Reliability Physics 1995, p205

[3.9] B. Lambert et al., *"Breakdown voltage of AlGaAs/InGaAs HEMT subjected to life-tests in impact ionization regime"*, GaAs Reliability Workshop 2000, p3

[3.10] W.T. Anderson et al. *"Life Testing and Failure Analysis of PHEMT MMICs"*, GaAs Reliability Workshop 2000, p47

[3.11] M.J. Delaney et al. *"GaAs PHEMT and InP MMIC Requirements for Satellite based Communications Systems"* GaAs MANTECH 2001,

[3.12] B. *Kraemer" Power PHEMT module delivers 12 watts, 40% P.A.E. over the 8.5 to 10.5 GHz band"* IEEE MTT-s International 1994, p1155

[3.13] S.H.Chen et al., " *2.4 V-operated enhancement mode PHEMT with 32 dBm output power and 61% power efficiency*" Proc. Asia Pacific Microwave Conference 2001, p1291

[3.14] X.Cao et al., "*mm-wave Performance of 50nm T-gate AlGaAs/InGaAs Pseudomorphic High Electron mobility Transistors with f_T of 200GHz*", GaAs2003, 2003, p13

[3.15] T.K. Tan et al., "*94GHz 0.1μm T-gate low noise pseudomorphic InGaAs HEMT*", Elect. Dev. Lett, **11**, p585 1990

[3.16] S. Bollaert et al., "*The Indium Content in metamorphic $In_xAl_{1-x}As/In_xGa_{1-x}As$ HEMTs on GaAs substrate: a new structural parameter.*" Solid State Electronics, **44**, 2000, p1021

[3.17] M. Haupt et al. "*Growth of high quality $Al_{0.48}In_{0.52}As/Ga_{0.47}In_{0.53}As$ Heterostructures using strain relaxed $Al_xGa_yIn_{1-y}As$ buffer Layers*"Applied Physics Letters, **6**, 1996, p412

[3.18] W.E. Hoke et al., "*Molecular beam epitaxial growth and device performance of metamorphic high electron mobility transistor structures fabricated on GaAs substrates*", Journal Vacuum Science and Technology B, **17** (3) 1999, p1131

[3.19] D. M. Gill et al., "*High Performance 0.1μm InAlAs/InGaAs High Electron Mobility Transistor on GaAs*" IEEE Electron Device Letters, **17**, 1996, p328

[3.20] S.I. Molina et al. "*Strain relief in linearly graded composition buffer layers: A design scheme to grow dislocation-free and unstrained epilayers*", App. Phys. Lett. **65**, 1994, p2460.

[3.21] Y. Cordier et al., "*AlInAs/GaInAs Metamrophic HEMT's on GaAs substrate from material to device*", Proceedings of Indium Phosphide and Related Materials, 1998, p211

[3.22] M. Chertouk et al., "*Metamorphic InAlAs/InGaAs HEMTs on GaAs substrates with composite channels and f_{max} of 350Ghz*" Proceedings of Indium Phosphide and Related Materials, 1995

[3.23] S. Bollaert et al., "*0.06μm Gate length metamorphic $In_{0.52}Al_{0.48}As/In_{0.53}Ga_{0.47}As$ HEMT on GaAs with high f_t*" Proceeding of 2001 International Conference on Indium Phosphide and Related Materials, 2001, p192

[3.24] X.Cao et al. "*mm-wave performance of 50nm T-gate In $_{0.52}Al_{0.48}As/In_{0.53}Ga_{0.47}As$ Metamorphic High Electron Mobility Transistors*", Proceedings of GaAs2003, 2003, p197

[3.25] T. Ando, A.B. Fowler and F.Stern, "*Electronic Properties of two dimensional systems*", Reviews of Modern Physics, **54,** 1982, p502

[3.26] P.M. Smith et al. "*Advances in InP HEMT technology of High Frequency Applications*", Proc. of IPRM 2001, 2001 p9

[3.27] Sumitomo product data sheet for 4-inch semi-insulating InP-VCZ substrate.

[3.28] C. Pobanz el al., *"High performance MMICs with sub-millimeter performance InP-based HEMTs"* Proc. of IPRM 2000, 2000, p67

[3.29] R.Lat et al., "*InP HEMT Amplifier Development for G-band (140-220 GHz) Applications*", Proc. of IEDM 2000, p175

[3.30] C. Meliani, et al. "*A high gain-bandwidth product InP HEMT distributed amplifier with 92 GHz cut-off frequency for 40 Gbit/s applications and beyond",* Proc. of GaAsIC Symposium 2002, p103

[3.31] K. Murata et al., "*100-Gbit/s Logic IC using 0.1μm-gate-length InAlAs/InGAAs/InP HEMTs*", Proc. of IEDM 2002, *p937*

[3.32] K. Sano et al., "*1.7-W 50-Gbit/s InP HEMT 4:1 multiplexer IC with a multi-phase clock architecture*", *", Proc. of* GaAsIC Symposium 2002, p159

[3.33] T. Enkoi, "*InP-based HEMT technologies toward 100Gbit/s ICs*" Proc. of IPRM 2002, p439

[3.34] M. Hafizi and M.J. Delaney, "*Reliability of InP-based HBT's and HEMT's: Experiments, failure mechanisisms, and statistics*", Proc. of IPRM, 1994, p229

[3.35] H. Rohin et al. "*A 0.1μm MHEMT millimetre-wave IC technology designed for manufacturability*", Solid State Electronics, **43**, 1999, p1645

[3.36] Y. Yamashita, et al. "*High f_t 50nm Gate InAlAs/InGaAs high electron mobility Transistors Lattice Matched to InP substrates*", JJAP, **39**, 2000, pL838

[3.37] K.Shinohara et al., "*Extremely High-Speed Lattice Matched InGaAs/InAlAs High Electron Mobility Transistors with 472GHz cutoff frequency*" JJAP, **41**,2002, p L439

[3.38] H. Maher et al., "Dry Etch Recess of an InGaAs/InAlAs/InP HEMT like structure using Low Energy High Denisty $SiCl_4$ Plasma (ICP)" Proc. Of IPRM 1998, p793,

[3.39] H. Fourre et al., "*Selective wet etching of lattice matched InGaAs/InAlAs on InP and metamorphic InGaAs/InAlAs on GaAs using succinic acid/hydrogen peroxide solution*", J. Vac. Sci. Technol. **B14**, 1996, p3400

[3.40] T. Suemitsu et al."*30nm two-step Recess Gates InP Based InAlAs/InGaAs HEMTs*", IEEE Trans. Electron Devices, **49**, 2002, p1694

[3.41] Nguyen et al. "*50-nm self Aligned-Gate Pseudomorphic AlInAs/GaInAs High Electron Mobility Transistors*" IEEE Trans. Electron Devices, **39**, 1992, p2007

[3.42] D.L. Edgar et al., "*millimeter-wave performance of InAlAs/InGaAs HEMTs using a UVIII/PMMA bilayer for 70 nm T-gate fabrication*", European Microwave Conference 2002, 2002.

[3.43] T. Suemitsu et al. "30-nm-Gate InAlAs/InGaAs HEMTs Lattice Matched to InP Substrates", Proc. of IEDM 1998, p223

[3.44] Y. Chen et al., "*Fabrication of 30nm T gates using SiN_x as a supporting and definition layer*", J. Vac. Tech., **B18,** 2000, p3521

[3.45] A.Endoh et al. "*Fabrication Technology and Device Performance of Sub-50nm-Gate InP-Based HEMTs*", Proc. of IPRM 2001, p448

[3.46] A.Endoh et al. "*InP-based High Electron Mobility Transitors with a very Short Gate Channel Distance*", JJAP, **42**, 2003, p2214

[3.47] P.Smith et al., "*W-Band High efficiency InP-based Power HEMT with 600 GHz fmax*", IEEE microwave and Guided Wave Letters, **5**, 1995, p23

4.0 Fabrication Technologies

Chapter 2.0 introduced the physical structure of the HEMT, and discussed the theory of the operation of the HEMT. This chapter will describe the techniques used to realise these devices. This will include the growth technologies employed to produce the epi-layer structures then concentrating on the lithography and pattern transfer processes that allow the fabrication of the HEMTs with critical dimensions below 100nm in the main body of the chapter. Finally the process flow of a HEMT will be introduced.

4.1 Wafer Growth

As shown in section 2.2.1 the basis for the HEMT is the heterojunction formed between a high and low bandgap material. In addition, the electrical behaviour of the HEMT is strongly dependent on the respective thickness of the numerous layers in the HEMT structure. This requires a technology that is capable of growing the desired layers to atomic monolayer precision. There are a number of methods that are used to do this, this most notably of which are Molecular Beam Epitaxy (MBE) and Molecular Organic Vapour Deposition (MOVCD).

The theory of MBE growth is relatively simple. Atoms or molecules of the desired elements are produced by heating up a solid source. Under Ultra High Vacuum (UHV) conditions these molecules migrate as a beam towards the sample. When they reach the sample they diffuse, react and form a new surface. The various compositions of the grown material can be controlled by controlling the flux of the beams or by closing the shutters between the source and the sample.

In MOVCD the growth is achieved by reacting gases that contain the constituent components of the desired crystal on the surface of the sample. Typical gases that are used are group III metal alkyls such as $(CH_3)Ga$ and group V hydrides e.g. AsH_3. When these react, GaAs is formed and a by-product of methane is produced. Other gases can be used to produce many types of semiconductors such as InGaAs, AlGaAs or InP. Dopants such as Si are also added in the form of reactive gases. Altering the reaction gases used, changes the composition of the grown material.

This work was carried out using material grown by MBE. This is because the MBE process allows layers to be grown with great precision and with very sharp interfaces. Due to the nature of the MBE process highly accurate doping profiles can be achieved. This degree of control

allows epi-layers to be grown as they are designed and therefore provides good material for the fabrication of high speed HEMTs.

4.2 Lithography

Lithography is the process of transferring designed patterns on to a material such as a semiconductor wafer grown by MBE.

There are a number of different methods of lithography available, the majority of which involve the transfer of the desired pattern on to a polymer layer, referred to as resist, which is then transferred to the substrate by either deposition of addition material such as metal or etching the substrate. The various lithography techniques differ by how the pattern is initially transferred into the resist. This can be achieved by exposing the resist to various energetic wave-particles such as UV photons, electrons or X-rays. This exposure changes the composition of the resist allowing selective removal by a suitably chosen solvent. A second method of patterning involves a mechanical process, either stamping the patterning into a thick polymer layer [4.1] or dipping the patterned stamping tool into a polymer and then on the surface of the sample. [4.2]

It is the exposure method that is almost universally used, both in industrial and R&D applications, the later method is still in its infancy with a great deal of active research refining this technique. The most favoured techniques of lithography use either ultraviolet light or electrons to modify the resists, these are referred to as photolithography and electron beam lithography respectively.

4.2.1 Photolithography

Photolithography uses UV light to modify the chemistry of the polymer chain changing the solubility of the polymer. The process flow is illustrated in Figure 4.1.

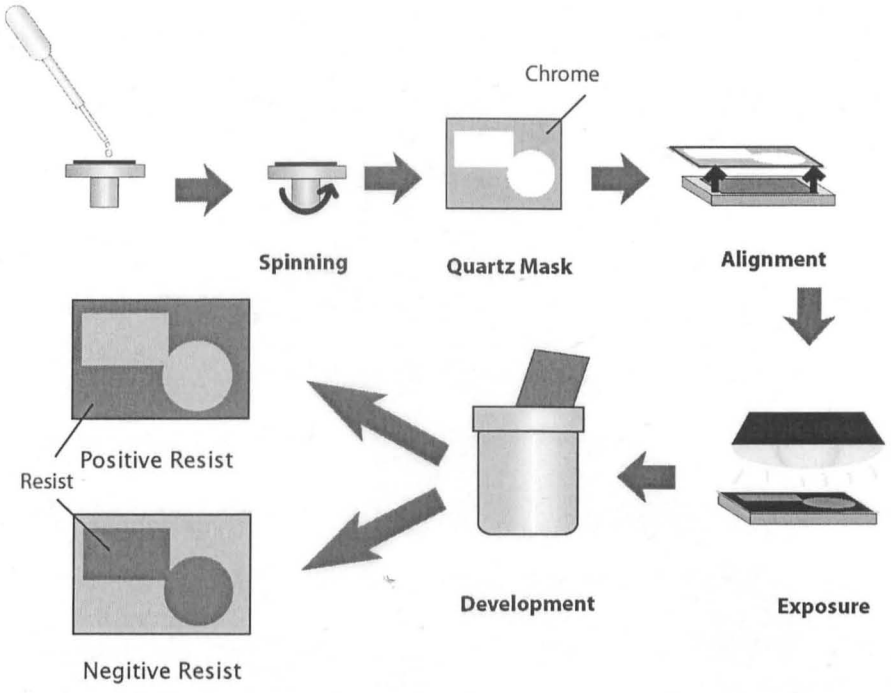


Figure 4.1 Step by step illustration of the photolithography process.

The first stage in photolithography is the making of the mask. The pattern is transferred from the mask onto the substrate in the following manner. A solution of resist in solvent is applied on the surface of the material. The sample is then spun at high speed (typically between 1000 and 6000rpm), the centrifugal action causes the excess resist to flow off leaving a thin and uniform layer of polymer. The substrate is then baked to drive off the solvent. The mask containing the pattern shown in Figure 4.3, is made by etching the pattern into a chrome film on a quartz or glass plate. During exposure the UV light produced by the lamp passes though the mask and onto the sample, modifying the resist. The final step is to develop the sample in a suitable solvent, one is able to differentiate between the exposed and unexposed areas. In the case of a positive resist the developer removes the areas of resist that have been exposed to light, whilst the developer for a negative resist will remove only the areas of resist that have not been exposed.

Various wavelengths of UV light are used in photolithography, each represent the various stages of development of equipment with an trend towards the definition of smaller and feature sizes. The photolithography equipment with the University of Glasgow has a light source wavelength of 365nm capable of producing features sizes of 1.12μm, whereas extensive work is being carried out in contemporary fabrication facilities to bring production lines based on 157nm wavelengths into production of devices with critical dimensions of 65nm and below. [4.3]

The main advantage of photolithography is that it is a parallel process, which provides high throughput and cost effectiveness. The main disadvantage of photolithography is the ultimate resolution that is possible to achieve, despite the recent great advances in photolithography pattern size will be limited by the wavelength of light. In addition the cost of the plate plates that are required to produce sub-100nm patterns is very high. If a modification is made to the design a whole new set of mask plates needs to be produced. Since this work is primary concerned with the reduction of critical dimensions of the HEMT, which requires rapid prototyping this work was carried out using electron beam lithography. Electron beam lithography is capable of producing much smaller features and it is possible to transfer the CAD pattern directly from the computer onto the substrate.

4.2.2 *Electron Beam Lithography*

Electron beam lithography, as its name suggests is the process of writing patterns into a electron sensitive polymer. Electron beam lithography is capable of much higher resolution than photolithography. Instead of the blanket exposure of the mask during the photolithography process the electrons in electron beam lithography are focused into a beam. The width of the beam used depends on the tool but is typically around 10nm. The beam is scanned over the surface of the sample which has been coated in a electron sensitive resist producing the pattern.

The use of a very small beam allows very high-resolution patterns with sub-10nm geometries to be produced [4.4,4.5].

In order to gain a better understanding of the process of electron beam lithography the basic operation of an electron beam writer will be studied.

4.2.3 Electron Beam Writers

There are several types of electron beam writers available. These range from the small scale converted Scanning Electron Microscopes (SEM) and Transmission Electron Microscopes (TEM) to production systems that can be loaded with many large mask plates simultaneously. Use of converted electron microscopes dates back to the mid-1960's. These systems still find use in many research laboratories worldwide as they provide small beam spot sizes at relatively low cost. An advanced example of this kind of machine is the converted Jeol 100CXII TEM at Glasgow University. This has been fitted with Raith scan generators, focus control and a stage capable of movement of 3μm. This tool is able to produce a spot size of 3nm and has demonstrated wires with a thickness of 3nm [4.6] and a grating with a period of 23nm. [4.7]. Example of the commercial machines are those made by companies such as Jeol and Leica. The work of this thesis was primarily carried out on this type of machine, a Leica EBPG5 HR100. This tool can be operated at both 50 and 100keV and is capable of writing substrates as large as 6 inch mask plates and producing spot sizes as small as 12nm at 100keV. The maximum size of pattern that can be written without moving the stage is 800μm and the smallest stitching possible between two adjacent fields is 30nm.

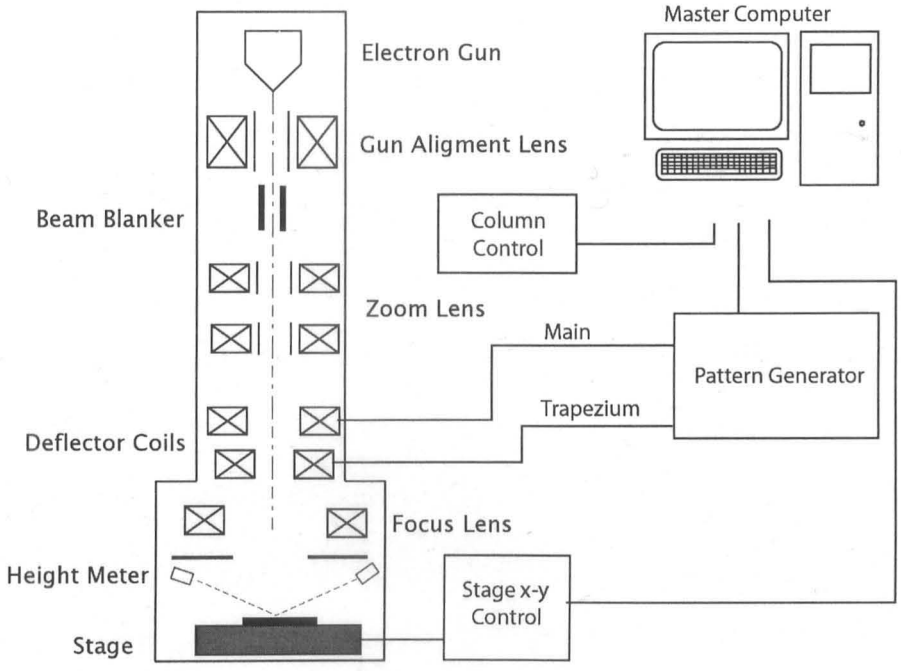


Figure 4.2 Schematic of the column of an electron beamwriter showing the various electo-optics in the column.

The schematic of a typical beamwriter is shown in Figure 4.2, this shows the different elements that are part of the column of a beam writer. The first element is the electron gun, which produces the beam of electrons. There are two ways of extracting electrons out of a metal. The first method is to heat the metal until the electrons have a greater energy than the workfunction and are free to escape, this is called thermionic emission. The second method is to apply a very large electric field, strong enough to allow the electrons to tunnel through the barrier, this is referred to as field emission. The EBPG5 uses a thermionic LaB_6 source, which has the advantage of producing a much brighter beam than the equivalent tungsten thermionic source [4.8]. The emission properties of the gun can be controlled by using two additional electrodes [4.9].

The next stage is the gun alignment lens, which aligns the beam produced by the gun down the axis of the column. The following stage is the beam blanker whose purpose is to allow the beam to be switched on and off. The beam blanker consists of a pair of plates set up as an electrostatic deflector. When a voltage is applied across the plates the electrons are deflected off axis and stopped by a downstream aperture. Care is taken when designing the blankers that they

operate fast enough not to drag the beam across the sample and thereby exposing resist. The zoom lenses de-magnify the beam to reduce the spot size of the beam from the size of beam produced by the source. The deflectors are used to steer the electron beam in the XY plane, this in addition to the beam blankers allow the pattern to be written. There are two deflector coils, the main deflector and the trapezium deflector. The main deflector is capable of producing deflections of ±400μm. This positions the beam within a block the trapezium deflector is used to move the beam around the trapeziums that make up a pattern (the maximum deflection possible is ±3.15μm). The final stage is the final focus lens; this uses the information collected by the laser height monitor to ensures that the beam is in focus on the surface of the sample. The sample is mounted on a stage, isolated from any vibrations and capable of moving with an accuracy of 10nm.

If the beam was swept in a raster over the whole of the 400x400μm block writing each pixel in turn the time to write the pattern would be very long. To avoid this, the pattern is "fractured" into smaller shapes called trapeziums. These shapes are written using the trapezium deflection coils to write the shape. When this shape has been written the beamwriter moves to the next trapezium in the block. Due to the fact that in the majority of patterns there is a very small written area this method greatly reduces the time taken to write each job.

Block Size and Stitching

The size of the fine pixels in the pattern is determined by the "resolution", which, on the EBPG5 can be varied between 5nm and 312.5μm. The number of pixels that make up a block is determined by the number of bits in the Digital to Analogue Converters (DACs) that control the scan coils, the EBPG5 has 15bit DACs allowing 32000 pixels. This means that when using the finest resolution the block size is reduced to 160μm. The maximum block size possible is limited by the main deflector coils, at a beam energy of 50keV this is 800μm reducing to 560μm at 100keV due to the electrons increased energy.

Patterns which are larger than the block size need to be stitched together. There is a limit to how accurately the blocks can be placed together and aligned, resulting in "stitch errors". These can often be minimised by designing the pattern to allow for block boundaries. Often the pattern itself is smaller than the block size but lies across a block boundary. By design and careful choice of the extent of the whole design, the patterned areas can be moved away from the boundaries. A number of effects increase the stitch errors, one of most significant is height errors. A difference of 1μm in height can increase the block size by 10nm, this will increase the stitch error by 20nm. Therefore it is important that the back of the sample is clean and that the sample is mounted as flat as possible.

Pattern Transfer from CAD design to substrate.

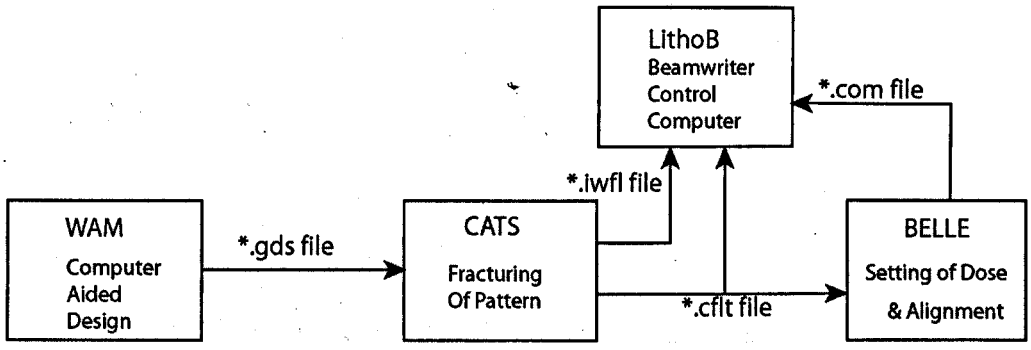


Figure 4.3 Flow of preparation of pattern design from CAD to the beamwriter.

As was discussed in section 4.2.1, one of the advantages of electron beam lithography is the ability of being to write the pattern on the substrate from the Computer Aided Design (CAD) program. The process flow of transferring the pattern from the CAD program to the e-beam tool is illustrated in Figure 4.3. The first step is the design of the pattern, this can be done using a program called Wavemaker, often referred to as WAM. This is a simple design program that allows patterns to be designed with great accuracy on up to 255 layers. The design is saved in GDSII format which is referred to as a *.gds file. This file is then "fractured" by a program called CATS. At this stage the layer(s) that are to be written are selected, and the resolution and the beam energy are set as these affect how the pattern is "fractured". The output files of this program are referred to as *.iwfl and *.cflt files, these are transferred to the beamwriter control

computer, LithoB. The final stage is to set the dose that will be used to write the pattern, set the spot size and detail the necessary distances required for alignment. This requires an input *.cflt and outputs a *.com file which is transferred to the beamwriter control computer which then writes the pattern on the substrate.

In this section the basic operation of an electron beam writer has been introduced as well as factors affecting the pattern quality that arise from the design of the beamwriter. The following section will consider the resists that allow the patterns defined the beamwriter to be transferred on to a substrate.

4.2.3 *Electron Beam Resists*

Electron beam sensitive resists use the modification of the size of the polymer chain length due to exposure.[4.10] This is enabled by electrons breaking the bond of a C-C link thus reducing the length of a chain. This process is illustrated in Figure 4.4, an energetic electron with an will ionise one of the electrons within the C-C bond thus breaking the bond. It has been found by studies using Energy Loss Spectroscopy that the energy required to cause a scission event is around 20eV.

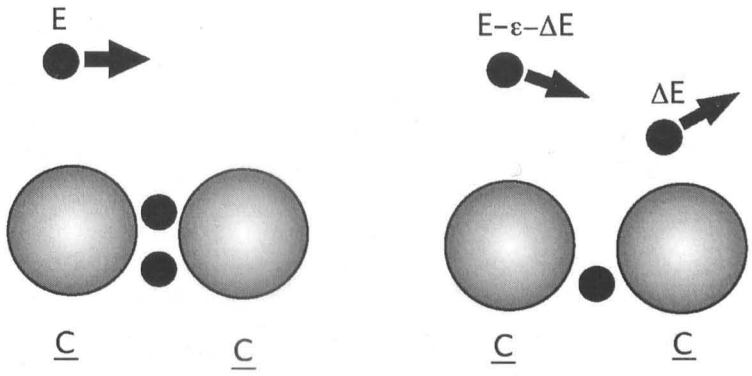


Figure 4.4 Diagram showing the breaking of an C-C bond by the interaction with a energetic electron.

It is important to note that the majority of the exposure reactions are not caused by the high energy primary electrons but by the lower energy secondary electrons that are emitted as the primary electrons pass though the resist and from back scattering of the electrons from the substrate. These electrons are important to the ultimate resolution that can be achieved in PMMA.

Development

The selective removal of the exposed areas is referred to as development. Development is usually achieved using 2 components. One of these is a good solvent of a particular polymer irrespective of its molecular weight. The second is not a solvent of the polymer regardless of the molecular weight. By creating a mixture of these with a particular ratio it is possible to tailor the developer to be a solvent for polymers below a particular molecular weight but be a non-solvent for polymers above the critical weight. For a developer this critical molecular weight should lie between the weight of the initial polymer and the average weight of the exposed polymers.

In the late 1960's and the early 1970's work was carried out to investigate the suitability of different polymers to be used as high resolution electron beam resists. One such study was carried out by Haller et al. [4.11] in which four different polymers were evaluated, namely Cellulose Acetate, Poly-isobutylene, Poly-(α-methyl-styrene) and Poly-(methacrylate) (PMMA). It was seen that of the polymers tested, PMMA was best suited as a high resolution e-

beam resist because it required the lowest dose of the resists tested to be fully developed; it provided the best resolution (less than 0.8μm) and in addition had the best resistance to etchants such as HF and HCl. The developer used was a mixture of Metha-isobutyl Ketone (MIBK) and Isopropyl alcohol (IPA), this was mixed in ratio of 30:70 MIBK:IPA. This was a very promising resist/developer system back in the 1960's and despite 30 years of intense development of other resists it is still used throughout the world today.

Doses and Resist Sensitivities

The number of electrons incident on an area of resist for the area to be fully removed by the developer is called the "clearance dose" and is measured in $\mu C/cm^{-2}$. The dose required to clear out a pattern in a given resist will depend on a number of parameters such as the developer, the development time and the temperature of the developer. Normally a standard development process that is followed and only the number of incident electrons is varied.

Different resists will behave differently to the electron exposure and they will therefore require a different dose to clear out. The lower the dose the resist requires, the more sensitive the resist and the faster it is. PMMA was mentioned above as being a standard e-beam resist. This is available is two molecular weights in the University of Glasgow these are denoted by ELV2010 and ELV2041 which have molecular weights of 90,000 and 420,000 respectively. The names are not related to the weight nor the sensitivity but are part of the trade name. The resist with the higher molecular weight will require a larger number of scission events and thus requires a greater dose.

Another polymer resist often used is a copolymer of Methyl-methacrylate and Methacrylic Acid, P(MMA-MAA), this is often referred to as Co-Polymer. Co-Polymer can also be developed in MIBK/IPA developer and has a even greater sensitivity than 2010 PMMA. This large difference in sensitivity over 2041 PMMA allows the fabrication of T-shaped gates. This process will be studied in greater detail in the subsequent chapters.

Although resists such as a 2010 and P(MMA-MAA) have better sensitivity than 2041 PMMA they are still relatively slow. In order to be able to increase the throughput needed for mass production new types of resists are constantly being investigated in Glasgow. Initially developed for use in deep ultra violet photolithography Chemically Amplified Resists (CAR)'s can be used in electron-beam lithography.

Chemical Amplified Resists (CARs)

CARs operate in a slightly different way from the traditional resists such as PMMA. The incident electrons do not themselves change the solubility of the exposed areas rather the electrons (or photons in the original case of photolithography) generate acids on exposure. These acids act as a catalyst to reactions in the resist that alter the solubility of the exposed areas.[4.12] This change can be used to remove the exposed areas or in the case of the negative resist leave the exposed areas.

Due to the reactions that take place in the resists the processing of CARs tend to slightly more complicated. In order to allow the acid catalysts to diffuse though the resist it is necessary to include a Post Exposure Bake (PEB). This PEB is a critical part of the development process as it determines how the acid catalyst diffuses though the resist. The duration and temperature of the PEB influence the sensitivity of the resist. In the case of a common resist, UVIII manufactured by Shipley, increasing the baking temperature increases the sensitivity of the resist. Increasing the PEB time can further increase the sensitivity [4.13]. If there is a delay between the PEB and development this can lead a to difference in the profile of the developed resists. These issues can however easily be controlled by careful consideration of timing of the fabrication process and monitoring of the development variables.

Examples of CAR's include UVIII already mentioned. This is a very sensitive positive electron beam resist, with a clearing dose of around 80μC/cm^{-2} compared to 300μC/cm^{-2} for PMMA, the fundamental resolution limit of UVIII is around 60nm [4.14]. In addition, UVIII has a good dry

etch resistance, better than PMMA [4.14]. Another positive CAR is the ZEP series of resists, this is a very high resolution resist, capable of features below 20nm with good contrast at writing speeds much faster than PMMA. The dry etch resistance is also very good, however removal is problematic, requiring an overnight soak in dedicated remover. Examples of negative CAR's include NEB-31 a very high resolution resist which has demonstrated features of 40nm [4.15].

Scattering and Resolution Limits

Electron beam writers are capable of producing a beam with a spotsize as small as 0.5nm [4.16] However the smallest features that have been fabricated by direct writing are of the order of 6nm. [4.4,4.5] This is because of the widening of the effective energy distribution of the beam within the resist [4.17] caused by scattering in addition to the production and path of secondary electrons in the resist. [4.18]

As the high energy electron beam passes through the resist it produces a large number of low energy secondary electrons. These electron travel perpendicular to the direction of the primary beam as they travel, they scatter and exposure the resist. Since the electrons have low energies their range is very small (~5nm in PMMA). These electrons have the effect of increasing the developed size of the pattern by around 10nm and therefore must be consider when realising pattern of 50nm or less.

As stated before the spot size of the electron beam can be made very small, however due to small aberrations in the lenses the energy distribution of the beam will not be an ideal delta function, rather it will be better described as a Gaussian distribution. Most of the energy will be located in the peak but there will be small but significant tails on either side. This effect has to be considered when producing very high-resolution patterns, however the effects are small compared to the next problem, scattering.

Scattering is the deflection of electrons from the beam axis by either resist or substrate atoms. This deflection of electrons causes the effective beam width to widen. There are two forms of scattering; forward scattering and back scattering, these are illustrated in Figure 4.5. Forward scattering takes place in the resist, incident electrons are deflected as they take part in scission events that expose the resist. Since the incident electrons are very energetic the deflection experienced will be a small and is inversely dependent on the energy of the electrons. By using thin resists and an accelerating voltage of at least 50keV the effects of forward scattering can largely be ignored for most patterns [4.19].

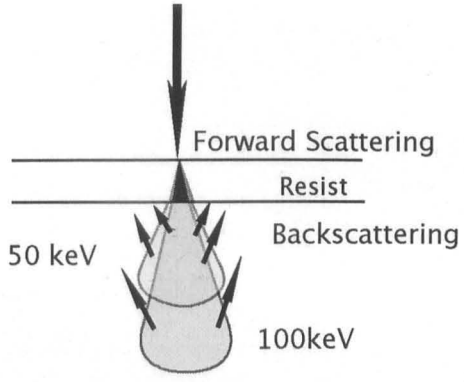


Figure 4.5 Illustration of the two forms of scattering from a point beam. Forward scattering occurs from scattering in the resist. Back scattering arises from the scattering in the substrate.

The second form of scattering is caused by the high energy electrons that have passed through the resist, scattering in the substrate material. The electrons penetrate deep into the substrate and scatter in all directions, some electrons scatter backwards and re-enter the resist and expose the resist from the substrate upwards. This scattering regime is therefore called backscattering and is the main component of the exposure of resist. The range of backscattering is much longer than that of forward scattering. The higher the beam energy the deeper into the substrate the electron will penetrate, the range of the scattering will be greater but the peak intensity will be lower.

If the two components of scattering are brought together, the effective deposited energy spread of a point source can be found. This is usually modelled by a double Gaussian distribution, one Gaussian models the forward scattering and the second models the backscattering.

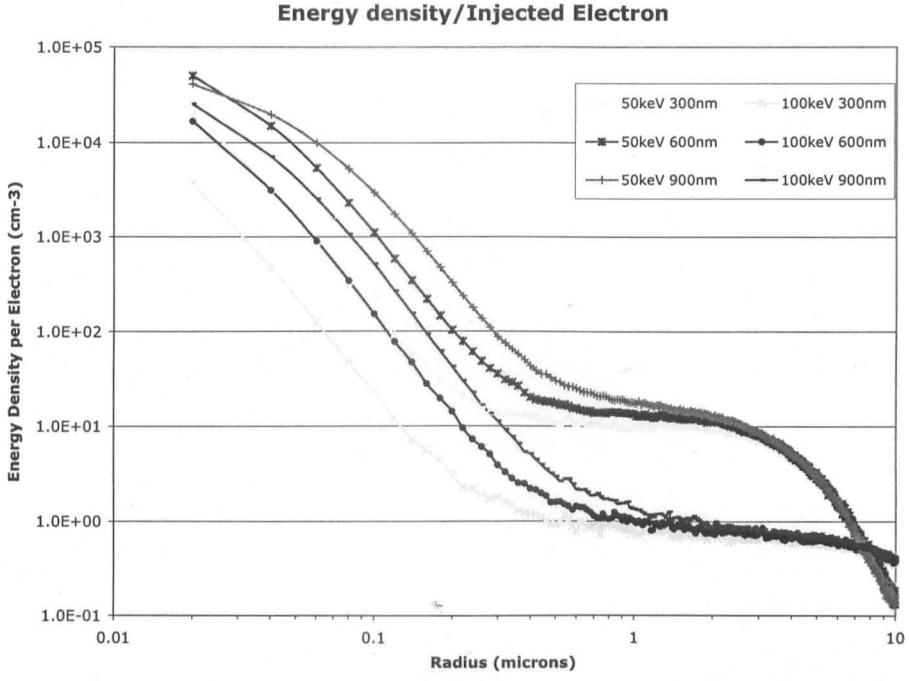


Figure 4.6 Plot of Energy density versus distance away from point beam. Plot shows the energy density distribution at a depth of 300, 600, 900nm from the surface of a 1000nm layer of PMMA for both 50 and 100keV.

The details of the distribution can either be found experimentally [4.20] or by Monte Carlo modelling, where the scattering of a large number of electrons is modelled. A resist structure of 1000nm of PMMA on a GaAs substrate was modelled using the "sceleton" Monte Carlo simulator for an electron energies of 50 and 100keV, the number of electrons were simulated was 10^6 in each case. The result of this simulation is shown in Figure 4.6. The distribution can be split into two parts, each of which can be described by a Gaussian distribution. It is seen that close to the beam axis the intensity falls off sharply, this is the regime controlled by forward scattering, the Gaussian for this section is tall and narrow, it is also seen that the 100keV curves fall off much quicker than at 50keV, this equates to the smaller forward scattering. The second feature of the plot is due to backscattering, this has a much greater range and is particularly clear at 50keV. This takes the form of a second, much smaller but wider Gaussian. At 100keV the

backscattered contribution is less pronounced. As the electrons scatter further, and so the peak intensity is smaller.

In addition to the comparison of the two energies the graph also shows the dependence of the distribution for a range of depths into a 100nm thick resist film. The graph shows the energy spread for 300, 600, and 900nm from the top surface of the resist. Moving further into the resist increases the width of the peak caused by increased forward scattering as the electrons travel further through the resist.

Proximity Effect

The effective widening of the beam caused by scattering has two main effects; firstly it imposes limits on the smallest feature size, as there will always be a widening of the electron beam due to scattering. This effect is not critical in many cases but becomes very important as the features size is reduced to below 30nm.

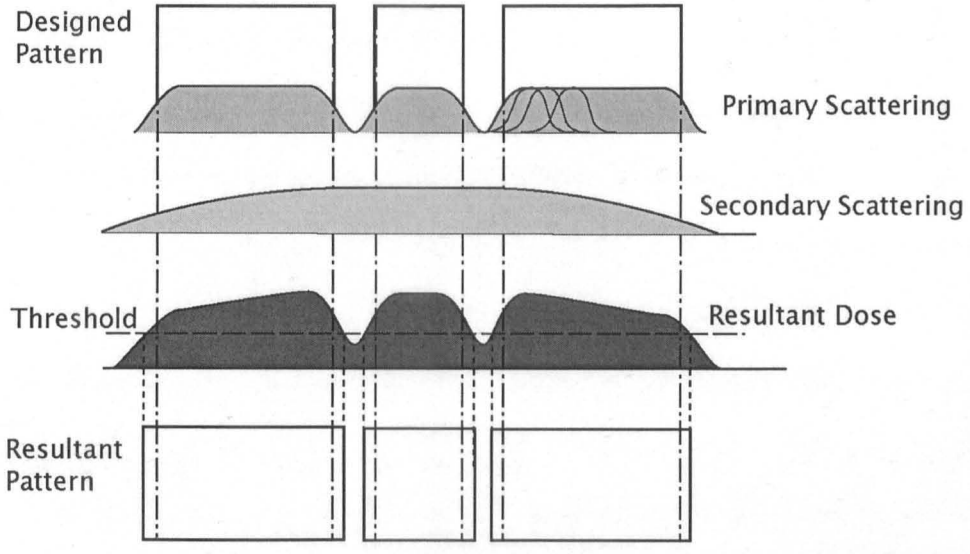


Figure 4.7 Illustration of the proximity effect. Forward and Backscattering each contibute to increasing the size of 3 closely spaced blocks. (After Proxecco Proximity Correction Website).

The second effect is the proximity effect, and is a more significant problem. It is caused by long range secondary electron scattering. Figure 4.7 illustrates the problem. If two large blocks are written close to each other (typically 2μm apart or less), tails of the energy distribution overlap, causing the gap in the middle to be developed, or at least become narrower than the

designed width. This effect can be reduced by using a process called proximity correction. A model of the electron scattering is produced for a resist of given thickness and sensitivity on a certain substrate material. This information is fed into a software program called PROXCECO, which breaks up the fractured patterns into smaller sections and assigns each of these a relative dose to allow the pattern to be developed out as designed. This process is used in the writing of the ohmic contacts of a HEMT to ensure the source and drain ohmic contacts are well defined.

4.3 Pattern Transfer

When the desired pattern has been produced in the resist the next stage is to transfer this pattern on to the surface of the substrate. This can either be achieved by etching away material or by lift-off. Lift-off is the process of depositing a material (usually metal) on to a substrate and using the resist as a shadow mask.

4.3.1 Metal Deposition and Lift-off

Various methods of metal deposition can be used for the lift-off technique. These include electroplating, thermal evaporation and electron beam evaporation. Electroplating involves placing the sample in a solution containing the metal that is to be deposited. A potential is applied between the substrate and the solution inducing a current. This breaks down the compound depositing the required metal on the surface of the sample. This method is usually only used when a very thick layer of metal is to be deposited, because of the length of time that it take to plate a sample and the fact that a thin seed layer needs to be deposited before deposition as well as the relatively inaccurate control of deposition depth. An example of an application of electroplating is the formation of the pillars for airbridges used in HEMTs and MMICs.

The more common process of metal deposition is by evaporation. This involves heating a metal source above its evaporation temperature by either an electrical element or by an electron beam.

The evaporated metal is deposited on the sample. A schematic of an evaporator is shown in Figure 4.7. This shows metal in the crucible, which is heated by an electron gun. Two shutters separate the sample and the crucible. Between the two shutters is a quartz crystal oscillator. As metal is evaporated, the resonant frequency of the crystal will fall and this change is linear with the metal deposition thickness. This is used to monitor accurately the rate of the metal deposition and the thickness of metal that has been deposited. Referring again to Figure 4.8, after the first shutter has been opened and the desired rate has been achieved the second shutter is opened allowing the metal to be deposited on the sample. The crystal frequency is monitored and when the desired thickness has been deposited the upper shutter is closed. The evaporators used in the course of the work of this thesis were electron beam evaporators produced by Plassys with computer controlled rate and thickness control. The accuracy in thickness of layer is better than 1nm for layers between 10nm and 100nm[4.21].

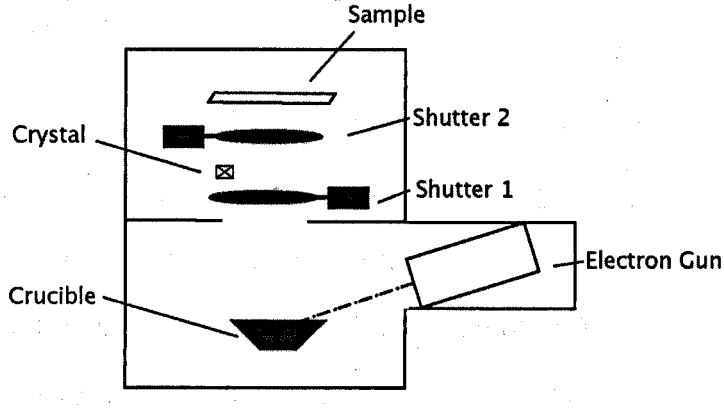


Figure 4.8 Schematic of an Electron Beam Metal Evaporator. The metal is evaporated in the crucible by the electron beam and is deposited on the sample. The rate is monitored by the piezoelectric crystal oscillator shown.

Lift-off Resist Structure

In order that the deposited metal is well defined and has the desired geometry after lift off, the resist profile has to be optimal. For example, if the resist profile of a line has sloping sidewalls the metal will be deposited up the wall of the resist. This makes it difficult for the solvent to get under the metal and dissolve the resist and if lift-off does occur the metal coating the side wall will form what are called "flags". "Flags" are strips of metal along the edge of a metalised area,

which can be many times taller than the thickness of the metal. An SEM image of a flag is shown in Figure 4.9.

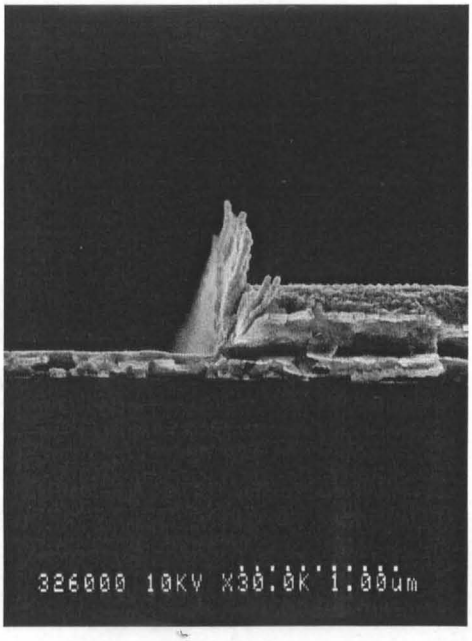


Figure 4.9 SEM cross section of a device showing large flags on the edges of the CPW bond pads. The height of the flags is on the same order as metal thickness, 450nm.

In order to avoid flags a slightly undercut resist profile is desired, as shown in Figure 4.10. This can be achieved using a two layer resist process. The upper resist layer defines the pattern by acting as a shadow mask. The lower resist layer is thicker this separates the metal deposited on the substrate from that on the surrounding resist. This profile can be easily achieved by using the different sensitivities of PMMA discussed in the previous section.

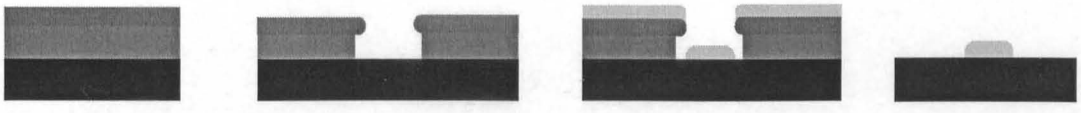


Figure 4.10 Illustration of undercut profile for lift-off using two sensitivities of resists. The less sensitive upper layer acts as a shadow mask for the metal deposition. The separation of the metal and resist prevents flagging.

4.3.2 Etching

Various etch processes can be used depending on the application, the material used and the depth of etch that is required. Two main types of etching are used in semiconductor manufacturing. In dry etching a plasma of energetic ions is used to etch the material. The gases used are chosen to be reactive with the material to be etched. Dry etching can achieve high resolution, high aspect ratio patterns as illustrated Figure 4.11, a Photonic Crystal etched in an GaAs/AlGas structure to a depth of around 2μm with a minimum feature size of 200nm.

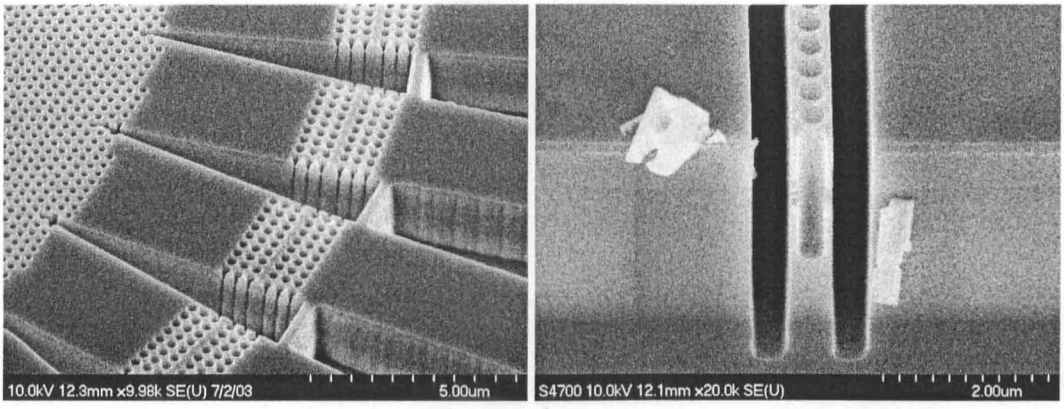


Figure 4.11 SEM images of Photnic Crystal device showing deep etches with high aspect ratio achieved using Dry Etch. (Courtesy of A. Jugessur, PBG Group, University of Glasgow)

With careful control, the dry-etch process can be tailored to selectively remove materials in an epilayer stack e.g. etching a GaAs layer but stopping on a AlGaAs layer. The disadvantage of the dry-etch process, is the damage that can occur due to ion bombardment of the sample which can degrade the electrical transport properties of the material, particularly in shallow structures such as HEMTs.

Wet etching of a semiconductor occurs in a solution of reactants usually dissolved in water. These reactants are tailored for the particular material that is to be etched. For instance for non-selective etching in the GaAs/AlGaAs material system, the active components are ammonia and hydrogen peroxide whereas for InGaAs/InAlAs structures, orthophosphoric acid and hydrogen peroxide are used. The etch rate of these solutions can be controlled by dilution in water. Using

this method the etch rates for the chemistries above can be as low as 1nm per second and therefore capable of good depth precision.

By careful choice of components it is possible to produce a selective wet etch which removes one material in an epilayer stack, stopping on a second material of different composition. Using this method it is possible to produce etches with an exact depth and adjustable width. The primary use for these etches is for the gate recess etch which will be discussed in greater detail subsequently.

4.4 Fabrication of a HEMT

The fabrication of a HEMT requires all the techniques that were discussed above and requires over 100 individual process steps and five separate lithography stages. The full process sheet for a GaAs pHEMT device is given in Appendix A. This section will explain the process flow of fabrication and discuss each stage in turn.

4.4.1 Process flow

The process flow leading to the realisation of a typical HEMT is shown in Figure 4.12. Each step will be briefly discussed, the techniques used and the importance and requirements of each stage will be explained.

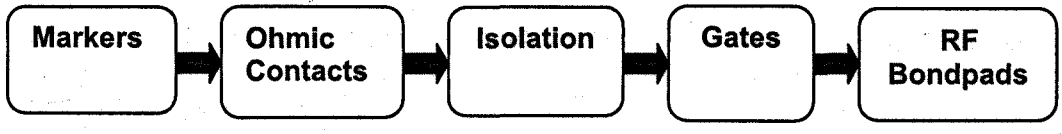


Figure 4.12 Process flow of a typical HEMT, although the Ohmic Contacts and Isolation levels can be interchanged without altering process.

Markers There are 5 stages of lithography that need to be completed in the fabrication of a HEMT. Each of these layers has to be accurately aligned with previous levels. The alignment required varies between layers; the ohmic and isolation levels can be misaligned by as much as 1μm, but the gate must be positioned in the centre of the source and drain contacts,

to an alignment better than 500nm and ideally no more than 100nm. This alignment is performed by the beamwriter by finding four gold markers that are placed at each corner of the pattern. These are then used as a reference point allowing the centre of the pattern to be positioned with great accuracy. Markers are realised using a lift off process described in section 4.3.1. In order that there is high enough contrast between the markers and the substrate for the markers to be located accurately the thickness of the metal layer needs to be more than of 100nm. Typically 20nm of titanium and 130nm of gold are used.

Ohmic Contacts The need for ohmic contacts was discussed previously. Like the markers they are fabricated using the lift-off method. The uniformity of the separation of the contacts is important as any variation along the width of the device can alter the electric fields in the channel and therefore the behaviour of the device. The contacts are written in positive resist so there is a small area of resist between two large open blocks. It was shown previously that the scattering of electrons can lead to the proximity effect and in the case of the ohmic contacts this has the effect of narrowing the gap in the centre of the contacts. Using PROXECCO reduced this effect. The metals that are deposited depend on the substrate used, typically for InP based HEMT the recipe is based on Ni/Ge/Au. The height of the ohmic contacts will influence the resist thickness in the gate lithography step, this therefore has to be considered when developing an ohmic recipe. After the ohmic contacts have been deposited they are typically annealed. This involves heating the metals to promote their diffusion into the semiconductor, providing more doping and reducing the semiconductor-metal barrier. The optimum temperature is normally found by experiment.

Isolation Isolation is necessary to separate one device from another and to ensure the current between source and drain flows under the gate. In these devices the removal of un-needed areas of active material serves to isolate the device. The pattern is written in positive resist where all of the cell is written with the exception of the device area. This protects the device while the rest of the material is removed by a non-selective wet etch. A typical etch

solution for InP based devices is 100ml water, 1ml orthophosphoric acid and 1ml hydrogen peroxide, a very dilute etch solution that etches at a rate of around 1nm/sec. The depth of the etch is monitored either electrically or by actual depth as measured by either a Dektak 6 surface profiler or an Atomic Force Microscope. The electrical method measures the current flowing between two 150μm square pads separated by 10μm before and after each etch step. When the current is in the range ~100nA for an applied gate voltage of 2V the sample is considered to be isolated. Using the depth method, the required depth is based on the layer structures. The target is to remove the cap, supply layers and the channel. The resist thickness is measured before etch and after each etch step. Electrical monitoring is preferable due to uncertainties caused by variations of resist thickness across the sample.

Gates The gate level is the most critical level in the whole fabrication process. It is the gate that ultimately controls the behaviour of the device. The gate must have a T-shaped profile to provide low resistance as well as small gate length. Gate lengths ideally must be below 100nm and be continuous along the whole gate width, typically up to 100μm. The T shaped profile is achieved by using up to three layers of resists with different sensitivities. The lithography of gates will be discussed in detail in Chapters 6 and 8.

The gate metal is deposited within an etch trench. In order for the gate to be separated from the highly doped cap the trench must be slightly wider than the gate. This gap is called the recess offset. (2.39) shows that the channel current is sensitive to the gate-channel separation so the depth of the etch has to be controlled carefully. The width of the recess offset modifies the depletion region and affects the breakdown voltage of the device and must also be controlled. This offset is achieved by using a selective etch discussed in section 4.3.2.

The metal deposited for the gate contains three layers, the first layer titanium is the "sticky" layer to give good adhesion to the substrate. The majority of the gate metalisation is a thick layer of gold to provide low resistance. Between these layers is an anti-diffusion layer, either

palladium or platinum. This layer is to prevent the gold diffusing into the semiconductor and degrading the Shottky contact.

RF Bondpads The purpose of the RF bondpads is to allow on-wafer characterisation of the device at DC and RF frequencies. The bondpads are in coplanar waveguide (CPW) configuration with the common source used as the ground layer. The waveguide is designed to have a characteristic impedance of 50Ω. The pads are formed by metal deposition and lift-off. In order to reduce losses at RF frequencies the metal used for the bonds is significantly thicker than used for other layers, up to 1.2μm. To allow for this, the thickness of the resist stack must be increased. This is done by using a resist solution with a slightly higher concentration of ELV2010 PMMA, 15% instead of the 12% used in the rest of the process. This gives an overall resist thickness of 1.3μm and enables thick layers to be lifted off.

This completes the high level description of device fabrication. Figure 4.13 shows an optical micrograph of a completed device, at x20 magnification and converted to greyscale for clarity. The CPW are labelled and the active section of the device is in the centre of the image. The vertical lines either side of the device are markers to enable accurate placement of the RF probes used for on-wafer measurements.

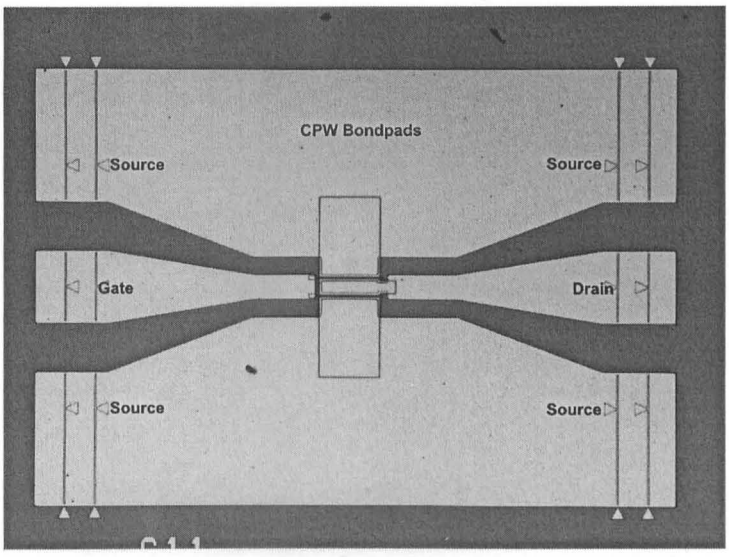


Figure 4.13 Optical Micrograph of a completed HEMT, the CPW bondpads are labelled as the source, drain and gate of the device.

Due to the large variations of dimensions within a HEMT structure (30nm to 400μm), several images at various magnifications are needed to fully illustrate the HEMT. Figure 4.14(a) shows the active region in greater detail.

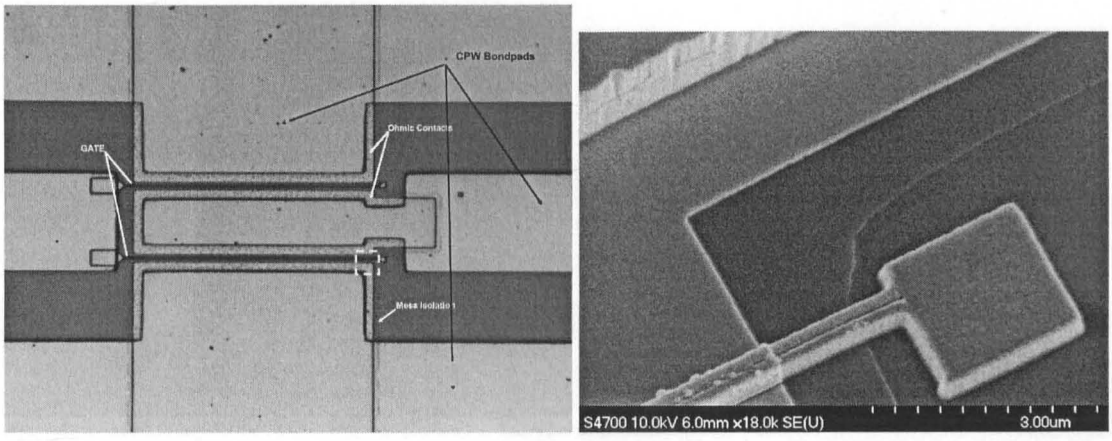


Figure 4.14 Images of the active region of a HEMT. (a) Optical Micrograph of a 50nm HEMT. (b) SEM image of a self-aligned HEMT showing the end of the gate (corresponding to dashed square in 4.14(a))

Figure 4.14(a) shows the active region in greater detail, this is an optical micrograph taken with a magnification of x100. The speckled morphology of the ohmic contacts can be seen this is caused by the annealing process. The isolation etch can be seen as a faint line around the ohmic level, it is clearer in the bondpad area. At this magnification it is possible to see the gate within the source-drain gap. However it is not possible to obtain accurate information on the alignment or the continuity of the gate. If more information is required on a device other analytical

methods must be used, such as AFM or SEM. Figure 4.14(b) shows a SEM image of the end of a self-aligned gate. The isolation level is much clearer now and it is it also possible to see the separation of the gate and the ohmic levels even though it is of the order of 5nm.

This chapter has introduced the technologies and the process that will be used in the following chapters for the realisation of advanced III-V HEMTs. This has included the technologies that are used to create the epi-layer material that HEMTs are fabricated upon and the lithography and pattern transfer techniques that are used to realise the devices in the following chapters .

4.5 References

[4.1] S.Y. Chou et al. " *Imprint of sub-25 nm vias and trenches in polymers*", Appl. Phys. Lett., **67**, 1997, p3114

[4.2] Y. Xia and G.M. Whitesides "Soft Lithography" Angewndte Chemie Int. Ed., **37**, 1998, p550.

[4.3] F. Schellenburg, "A little light magic" IEEE Spectrum, September 2003, p34, 2003.

[4.4] W.Chen and H. Ahmed, '*Fabrication of 5-7nm wide etched lines in silicon using 100keV electron-beam lithography and polymethaacrylate resist*", Appl. Phys. Lett. **62**, 1993, p1499

[4.5] F. Lehmann et al., "*Fabrication of sub-10-nm Au_Pd structures using 30keV electron beam lithography and lift off*", Micro. Eng. **65**, 2003, p327

[4.6] D.R.S. cumming et al, "*Fabrication of 3 nm wires using 100 keV electron beam lithography and poly(methyl methacrylate) resist*" Appl. Phys. Lett. **68**, 1996, p322.

[4.7] K.Lister et al. "*Pattern transfer of a 23nm period grating and sub-15nm into CVD diamond*", To be published in Proc. of MNE2003, 2003.

[4.8] SPIE Handbook of Microlithography, Micro machining and Microfabrication, §2.22

[4.9] M. Gesley, "Thermal Field Emission for nanolithography", Journal Appl. Phys., **65**, 1989, p914

[4.10] I.Haller et al. "*High Resolution Positive Resists for Electron Beam Exposure*", IBM Journal, 1968, p252.

[4.11] G. Wallraff et al."*Thermal and acid catalysed deprotection kinetics in candidate deep alutraviolet resist materials*", J. Vac. Sci. Technol. **12**, 1994, p3857

[4.12] L.Grella et al. "*Nanolithography performances of ultraviolet chemically amplified positive resist*", J. Vac. Sci. Technol. **B 15**, 1997 p2569.

[4.13] Cornell Univeristy Website, Resists and Processing, http://www.cnf.cornell.edu/ebeam/resist.html

[4.14] D.MacIntyre, S.Thoms, "*High Resolution Electron Beam Lithography Studies on Shipley Chemically Amplified DUV Resists*", Micro. Eng. **35**, 1997, p215.

[4.15] SPIE Handbook of Microlithography, Micro machining and Microfabrication, §2.22

[4.16] Cornell Univeristy Website, Resists and Processing, http://www.cnf.cornell.edu/ebeam/resist.html

[4.17] A.N. Boers, "Resolution Limits for Electron-beam Lithography", IBM Journal of Research and Development, **32**, 1988, p502

[4.18] S.A. Rishton and D. P. Kern, "*Point Exposure distribution measurements for proximity correction in electron beam lithography on a sub 100nm scale*", J. Vac. Sci. Technol., **5**, 1987, p135

[4.19] B. Wu and A, Neureuther, "*Energy deposition and transfer in electron-beam lithography*", . Vac. Sci. Technol., **19**, 2001, p2508.

[4.20] S.A. Rishton and D.P Kern., "Point exposure distribution measurements for proximity correction in electron beam lithography on a sub-100 nm scale "J. Vac. Sci. Tech. **B5**, 1987, p135

[4.21] Communications with Fiona McEwan

5.0 Development of Self-Aligned HEMT.

Section 2.3.7 showed that parasitic resistances and capacitances degrade the high frequency performance of a HEMT. Equations 2.48 and 2.52 for f_t and f_{max} respectively (both key indicators of high speed performance) demonstrate that the gate resistance, R_g, the source resistance, R_s, and the gate capacitance are particularly important in determining the RF performance of a HEMT. This is confirmed by Monte Carlo modelling of 120nm pHEMTs that a reduction in source and drain resistances significantly increases f_{max} [5.1]. For these reasons, Mishra proposed the Self-Aligned HEMT in order to reduce these parasitic elements [5.2] This work demonstrated a HEMT with a 100nm gate length on lattice matched InP substrate using the T-shape gate to define the ohmic contacts. Self-Aligned HEMT's have also been demonstrated with gates lengths of 65 and 50nm by researchers at Hughes [5.3,5.4]. It was decided to develop a self-aligned process for GaAs pHEMTs based on the University of Glasgow 120nm T-gate process. This chapter details the fabrication technologies developed to realise these new type of devices.

5. 1 Self-Aligned HEMT

A self-aligned device is realised by fabricating the T-shaped gate before the source and drain contacts. The gate is used as a shadow mask for the evaporation of the ohmic metal. Since the gate head is used to define the source-drain gap, the source-gate and drain-gate separation are reduced to approximately half the length of the gate head, this is illustrated in Figure 5.1.

Figure 5.1 Illustration of the fabrication of a self-aligned device showing the formation of the ohmic contacts using the T-gate as a shadow mask.

The advantage of this type of device is that since the source-gate distance is reduced the source resistance is also reduced. A secondary but important effect is that extra metal deposited for the ohmic contacts increases the cross sectional area of the T-gate thereby reducing the gate resistance.

5.2 Development of Non-annealed Ohmic Contacts

For the ohmic contact process to be compatible with a self-aligned process flow it must meet the following criteria:

1. The metal must be thin enough to prevent shorting of the gate to the ohmic contacts. The height of the gate stalk of the 120nm PMMA/P(MMA/MAA) gate process is typically 110nm. This means that the total thickness of the ohmic metal must be less, ideally around 100nm.

2. To ensure that the Schottky contact of the gate is not degraded by diffusion of the gate metal into the barrier layer, the processing temperature of the contacts must be low, ideally below 200°C.

3. In order to achieve the possible benefits of the self-aligned technology the resistance between the ohmic metal and the 2DEG (the contact resistance) must be at least comparable with the resistance of the standard pHEMT process. The resistance of the standard process is 0.12 Ω.mm.

5.2.1 Layer Structures for non-annealed ohmic contact.

Three layer structures were designed by Khaled Eglaid of the Ultra-Fast Systems group to investigate the feasibility of thin, low thermal budget Ohmic contact process and were grown by MBE at Glasgow University. These layer structures were based on the standard 120nm GaAs pHEMT layer structure given in Appendix B. All the layer structures were identical with the exception of the cap layers which were designed with the aim of providing a low contact resistance with low temperature processing. The design of the cap layer of each wafer is shown in Figure 5.2.

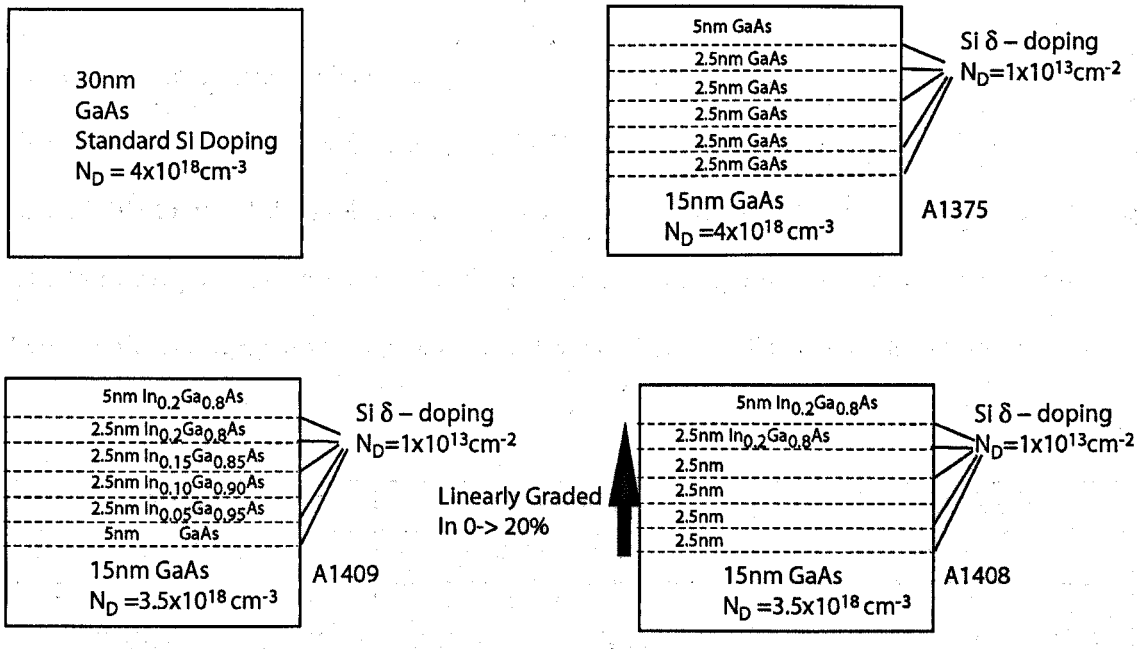


Figure 5.2 Cap layers of the wafers used in the development of low temperature ohmic contacts.

It was shown in section 2.2.4 that to produce a good ohmic contact it is important to minimise the depletion region that forms at the metal-semiconductor interface. This is achieved in a number of ways in the structures shown in Figure 5.2, as follows. In the wafer A1375 the GaAs cap layer contains 5 layers of $1x10^{13}cm^{-2}$ of Si δ-doping, each is separated by 2.5nm of GaAs.

This doping strategy increases the local doping to higher than can be achieved by uniformly doping the GaAs cap layer. Below this delta doped region is 15nm of Si doped GaAs with a doping concentration of $4x10^{18}cm^{-3}$. Wafers A1408 and A1409 used an identical doping strategy, however these wafer also contained various compositions of InGaAs in the upper section of the cap to lower the barrier height at the metal/semiconductor interface. The maximum In concentration used was 20% to minimise the lattice mismatch. In wafer A1408 the indium content is increased linearly from zero to 20%, in A1409 however the indium content increases in steps of 5% with each step having a thickness of 2.5nm. The uppermost layer in both wafers was a 5nm thick layer of $In_{0.2}Ga_{0.8}As$. The grading of the In concentration is designed to minimise the discontinuities between the conduction band of GaAs and $In_{0.2}Ga_{0.8}As$ which could hamper current flow.

5.2.2 *Ohmic Metal Recipes*

Several ohmic contact recipes were tested on each wafer. The Au-Ge-Ni based contact was developed in the 1960's by Braslau et al. [5.5]. The Ge-Pd based contact was developed in the late 1980's to avoid the problems of non-uniform morphology of the Au-Ni-Ge system [5.6]. The third recipe is based on the Au-Ni-Ge system but is more complex, using layers of gold between the surface, germanium and nickel layers. The metallisation strategies are summarised in Table 5.1.

Metals	Thickness	Total Thickness
Ge/Pd/Au	20/30/30nm	80nm
Ni/Ge/Au	10/50/80nm	140nm
Au/Ge/Au/Ni/Au	14/14/14/11/50nm	103nm

Table 5.1 Table of the ohmic recipes tested.

These metallisation recipes vary in their details but the methodology is similar in each situation, being based on the diffusion of the germanium through the metal-semiconductor interface. The diffused germanium displaces the gallium from the lattice providing extra n-doping at the surface, narrowing the barrier at the interface and increasing the tunnelling probability, thus

decreasing the contact resistance. The Ni layer aids the diffusion of the germanium into the crystal [5.7] as well as forming a conductive NiAs layer at the interface. [5.8] The palladium layer in the GePdAu scheme reacts with the Ge to form PdGaAs during deposition and then reacts with Ge upon annealing, the excess germanium diffuses and displaces the Ga in the GaAs cap. [5.5] The amount of excess Ge depends on the ratio of Pd to Ge, hence the process can be controlled by the thickness of the Pd layer.

The contact resistance of the ohmic contacts depends not only on the metallisation but also on the annealing temperature. The optimum temperature is normally found by measuring the contact resistance using a number of annealing temperatures and plotting against annealing temperature. The normal behaviour is that a minima is seen in the resistance, this process can then be repeated using temperatures around this value to find the lowest contact resistance. This technique was used to find the annealing temperature for the above recipes. Each were annealed at three temperatures, 260, 300, and 360°C for 60 seconds on all four wafers.

The resultant contact resistance was measured using the Transmission Line Method.

Transmission Line Method

The Transmission Line Method (TLM) is a technique to determine the contact resistance, R_c, through a planar semiconductor contact that allows for a non-uniform current flow. This was first proposed by Shockley [5.9], and was later improved independently by Berger [5.10] and Murmann and Widman [5.11]. The measurement structure is shown in Figure 5.3(a).

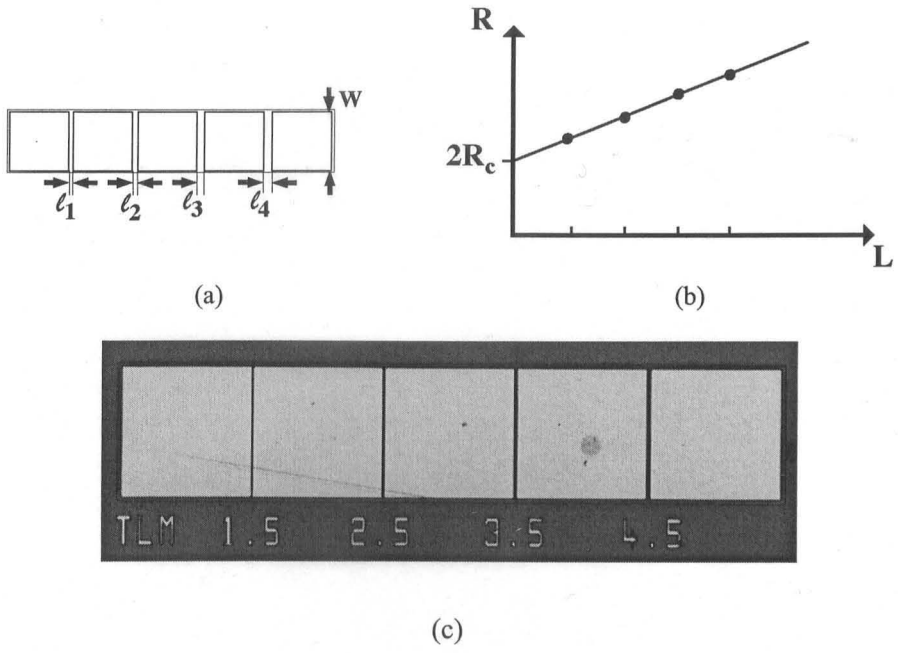


Figure 5.3 (a) Schematic of TLM test structure used. (b) Plot of measured resistance versus gap length. (c) Optical micrograph of TLM test structure.

The test structure consists of five ohmic contacts on the surface of the semiconductor. In this work, each contact was 150x150μm. The contacts are separated by an increasing distance, ℓ_1 to ℓ_4, the separations used being 1.5, 2.5, 3.5 and 4.5μm The resistance is measured by the four-point probe measurement. The resistance measured will be the sum of the contact resistance into the semiconductor from the metal and back out again, together with the resistance of semiconductor between the contacts, determined by the sheet resistance of the material. The resistance is plotted as a function of contact separation resulting in a straight line given by the equation

$$R = 2R_c + \frac{r_s l}{w}$$

Where r_s is the sheet resistance of the semiconductor. The value at which the line intercepts the y-axis is equal to twice the contact resistance. The accuracy of this method depends on the accuracy of the measured resistances and the accurate determination of the gap length.

	260°C	300°C	360°C
	Ni:Ge:Au (10:50:80nm)		
A1258	Non Ohmic	Non Ohmic	0.42
A1375	0.67	0.42	0.22
A1408	0.33	0.26	0.22
A1409	0.26	0.34	0.22
	Ge:Pd:Au (20:30:30nm)		
A1258	Non Ohmic	Non Ohmic	0.52
A1375	1.21	Non Ohmic	1.15
A1408	0.16	0.42	0.3
A1409	0.24	Non Ohmic	1.12
	Au:Ge:Au:Ni:Au (14:14:14:11:50nm)		
A1258	Non Ohmic	Non Ohmic	0.23
A1375	1.13	1.66	0.36
A1408	0.78	Non Ohmic	0.22
A1409	0.15	0.42	0.19

Table 5.2 Contact Resistance of the three contact metallisations on each wafer at three annealing temperatures.

Each ohmic recipe was tested on each cap layer at three different annealing temperatures, 260, 300, 360°C. Each measurement was averaged over at least 3 sites, these results are summarised in Table 5.2. It is seen that the Ni/Ge/Au metalisation only provided a low contact resistance when the sample was annealed at 360°C. At this temperature however the surface morphology was poor, as is shown in Figure 5.4. This is a known feature of the Ni/Ge/Au metallisation [5.7], although it was not expected at such a low temperature. The standard wafer, A1258 performed the worst, it required the highest annealing temperature before becoming ohmic. The layers with δ-doping all displayed a contact resistance of 0.22Ω.mm at an anneal temperature of 360°C.

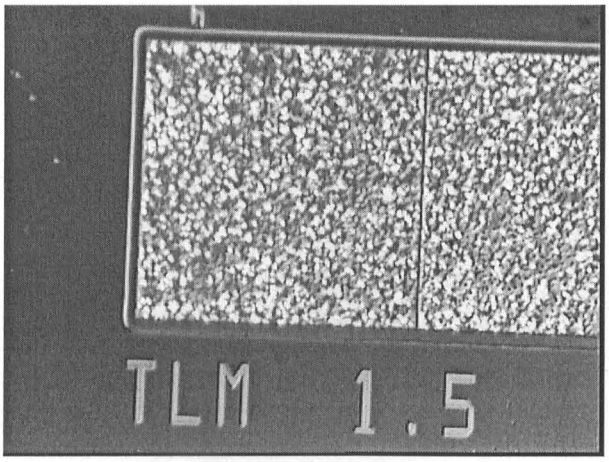


Figure 5.4 Optical Micrograph of poor surface morphology of Ni/Ge/Au annealed at 360°C.

The Ge/Pd/Au recipe showed no consistency in the results, it was ohmic on the structure containing δ doping when annealed at 260°C but not when annealed at 300°C. It was unknown what mechanism caused this temperature dependence.

A similar temperature dependence was observed with the Au/Ge/Au/Ni/Au recipe. The lowest resistance was measured for the lowest annealing temperature, the resistance then increased with temperature but reduced again at the maximum annealing temperature. It was thought that if this trend continued a lower annealing temperature would provide a lower contact resistance. The quickest way of testing this hypothesis was to test a device that had been metallised but not yet annealed. This was done and it was found that on wafer A1408 and A1409 the contacts were ohmic and provided a contact resistance of 0.13Ω.mm, the lowest of all the conditions measured. Following this result the other recipes were measured without annealing but no other recipes displayed this behaviour. The reason for this difference of the annealed contact performance is unclear. The metals at the surface of the Ge, Au, Ni all have a similar work function (5.0, 5.1, 5.15 eV respectively), however it should also be noted that a 100nm layer of Au was evaporated on to A1408 and was found to be ohmic displaying a contact resistance of 0.15 Ω.mm without annealing. This suggests that the low resistance, non-annealed behaviour is a particular property of the lower gold layer, perhaps an initial diffusion into the semiconductor during evaporation.

5.2.3 Further Development of Non-Annealed Ohmic contact

Two series of devices were fabricated on wafers A1408 and A1409 using the Au/Ge/Au/Ni/Au metallisation described above. These devices showed good RF performance and will be discussed in section 5.5. However from the DC characteristics it was noted that the low field resistance of the device was approximately three times larger than expected given the measured contact and sheet resistance. The reason for this difference was found when the layer structure of A1409 was modelled using the 1D Poisson Solver developed by Greg Snider [5.12]. This program allows the conduction band diagram and charge density to be calculated from the epilayer structures. The resultant charge density versus depth plot is shown in Figure 5.5. It is seen that the majority of the carriers are located in the cap region, around three times as many as located in the channel. This accumulation of electrons is what gives the contacts the low resistance when measured with the standard TLM method, conduction takes place in the cap region as well as the channel. Of-course in a HEMT it is the resistance between the contacts and the channel that is important, therefore a new TLM structure was designed to find this resistance.

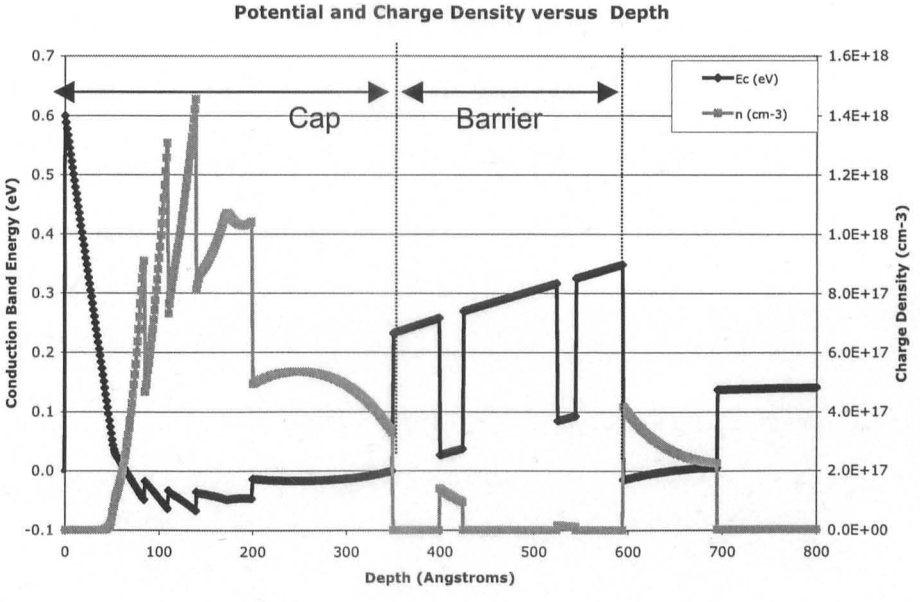


Figure 5.5 The Conduction Band Energy and Charge Density versus depth into the epilayer.

The new TLM structure was based on the standard structure and is shown schematically in Figure 5.6. The structure uses the five 150μm^2 pads, the separation of the pads has been increased to 3,4,5,and 6μm. Within each gap a rectangle is written with increasing width. The cap is then removed by selective wet etching. The resistance is measured before and after the etch, from these measurements the value of the contact resistance can be found. The value determined in this manner was 0.5Ω.mm, more than three times higher than the resistance measured by the standard method, which explains the observed increase in series resistance.

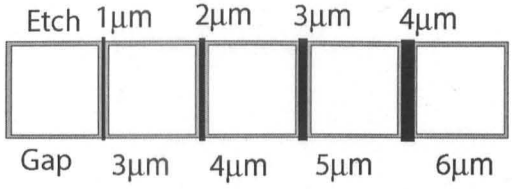


Figure 5.6 Schematic of revised TLM test structure including the etched regions between the ohmic pads.

This result shows that the measurement of the contact resistance using the standard method underestimates the resistance between the metal and the channel layer. The resistance of the non-annealed contact is limited by the vertical conduction through the epilayer rather than conductance of the metal-semiconductor interface. If the non-annealed contact is to be improved, conduction through the barrier layer must be considered. This work was carried forward in the department by Dave Moran [5.13].

5.3 Development of a Selective Gate Recess Etch

As mentioned previously, the gate level is the most critical in HEMT fabrication. It is essential that a good Schottky contact be formed i.e. the reverse leakage current is low. It was seen in section 2.2.3 that this is achieved by depositing the gate on a un-doped, high bandgap layer. This means that the highly doped, low bandgap cap layer must be removed from the region where the gate is deposited. This is done by etching the cap layer and stopping on the higher bandgap material (AlGaAs or InAlAs for GaAs pHEMT, InP LM/pHEMT and mmHEMT

respectively). A number of methods for the etching of the cap layers that have been used, these include selective dry etching [5.14,5.15] and non-selective wet etch [5.16, 5.17] There are a number of requirements for a gate recess etch that must be considered when developing a gate process.

The first of these issues is the most obvious one, the depth of the recess etch. It has been shown that the behaviour of a number of HEMT figures of merit are strongly dependent on the gate-channel separation, including the threshold voltage, V_{th}, and transconductance, g_m. (Equations 2.22 and 2.40 respectively) For repeatable device performance the etch depth must be carefully controlled. For this reason, selective etching is usually favoured as this gives greater control of the depth and is less susceptible to process shifts such as etch rate variations. The second feature of the gate recess is the recess offset. In order that the gate metal does not make contact with the cap layer the etch trench is made slightly wider than the gate metal as shown in Figure 5.7

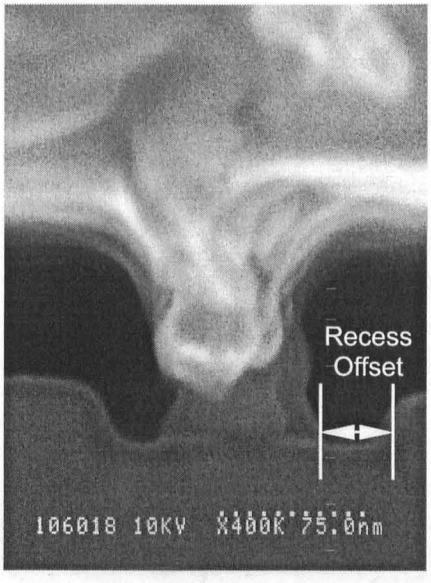


Figure 5.7 Cross sectional SEM micro graph of a 70nm T-shaped gate with a recess offset of 30nm.

The separation from the gate metal to the cap layer is referred to as the recess offset. The recess offset is an important parameter in the fabrication of a HEMT, as the output conductance, g_{ds}, feedback capacitance, C_{gd}, and breakdown voltage depend on the geometry of the recess. It has

been observed that as the offset is increased the cut-off frequency of the device decreases [5.18] and this is more marked in short gate length devices [5.19]. Therefore for high speed devices it is desirable to have a "tight" gate recess around the gate foot; however, the shorter the recess, the larger the electric fields in the drain side of the gate and the lower the breakdown voltage. For power devices that require a high breakdown voltage are large recess is used while high speed devices have small recesses. There are secondary factors that must also be taken in consideration, such as the damage that the etch may cause to the electrical transport properties of the 2DEG, of greater relevance to dry etch processes. In all cases, quality of the surface after etch is important as any oxides or other residues may affect the operation of the device.

At the start of the project, the gate etch used within the Ultra-Fast Systems group in Glasgow was dependent on the material system used. For InP and metamorphic GaAs HEMTs a non-selective orthophosphoric acid based wet-etch was used and a selective dry etch for GaAs pHEMTs. Both A1408 and A1409, the wafers that had displayed the lowest contact resistance, contain indium in the cap. The indium concentration is low compared to that of InP based HEMTs (20% rather than 52%) so the orthophosphoric based etch was unsuitable. Likewise, the dry etch process for GaAs pHEMTs was unsuitable as it didn't remove the indium in the cap layer and so another process had to be developed.

The etch to be developed had to be low damage, result in a clean surface and ideally be selective. In addition to this, the relative etch rates of each layer of the cap and etch stop layer had to be considered. It had to etch GaAs and InGaAs with In concentrations ranging from 0 to 20% at roughly the same rate to avoid overhanging by the etched cap. It must also have a much lower etch rate of $Al_{0.3}Ga_{0.7}As$ compared to InGaAs to allow a lateral offset to be formed before the etch stop layer is consumed.

A literature search showed that a number of papers referred to using a succinic acid based selective etch for the gate recessing of metamorphic devices [5.20, 5.21]. This etch was described in greater detail by Fourre et al. [5.22], who demonstrated the selective etching of various compositions on InGaAs over InAlAs and compared different etch solutions. It was suggested in this paper that the selectivity of the etch is related to the Al content of the etch stop layer. If this hypothesis was correct, such a method could be used for the etching of GaAs and InGaAs over AlGaAs. This theory was the basis of the development of the etch.

Work had previously been carried out within the group using a succinic acid based etch for the selective etching of metamorphic HEMTs [5.23]. This was taken to be the starting point for the development of the etch. The initial etch solution used was 20g of succinic acid powder dissolved in 100ml water, 10ml ammonia and 16ml hydrogen peroxide. This solution is then balanced to pH 5.5 with the addition of more ammonia. The chemistry of the etch is similar to other systems; hydrogen peroxide oxidises the surface, and the acid within the solution dissolves the oxide from the surface. In the case of this etch the selectivity is achieved because at a pH of 5.5 the acid is unable to dissolve the aluminium oxides that are formed when the $Al_{0.3}Ga_{0.7}As$ layer is reached, hence the etch stops.

Tests were carried out using the above recipe with 120nm T-gates in a PMMA/P(MMA-MAA)/PMMA resist stack as the etch mask on the wafer A1408. After lithography and development, the samples were de-oxidised for 30 seconds using a solution of 1:4 hydrochloric acid and water for 30 seconds before being recess etched for 30 seconds. The resultant etch profile is shown in Figure 5.8(a), which demonstrated that the cap has been etched. In addition the etch has stopped on the AlGaAs etch stop layer. A recess offset has been formed, of length of around 120nm rather than the desired length of 30nm. To reduce the lateral offset the etch solution was diluted by halving the quantities of all active ingredients - the results of this are shown in Figure 5.8(b)&(c).

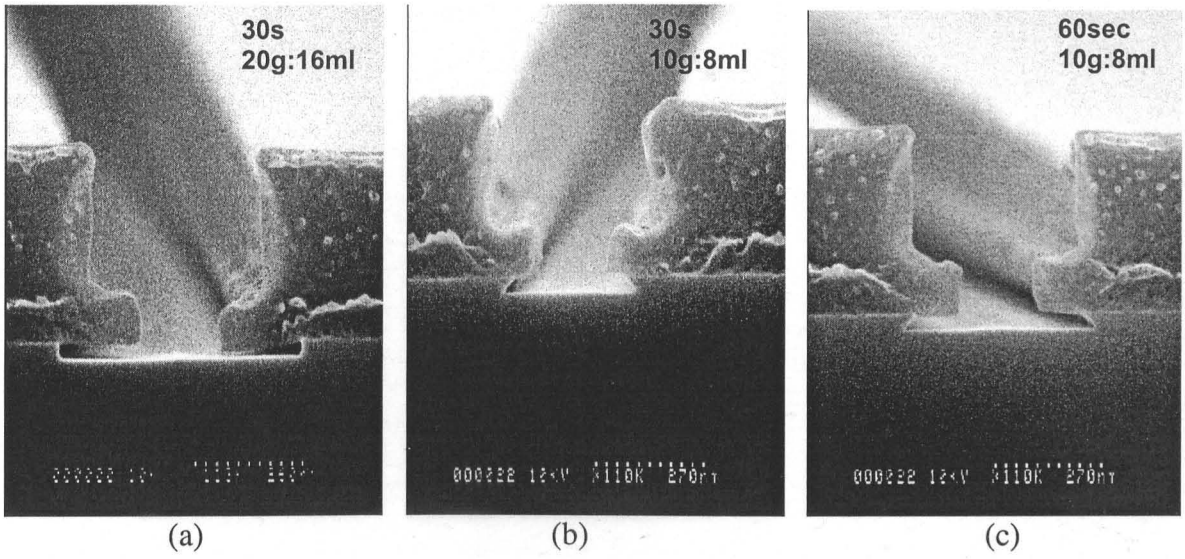


Figure 5.8 Results of succinic acid based selective etch on wafer A1408 (a) Initial solution 20g SA:100ml H_20: 16ml H_20_2 for 30seconds (b) Diluted Solution 10g SA:100ml H_20: 8ml H_20_2 etched for 30sec. (c) Diluted Solution etched for 60sec this shows the control of offset with etch time.

These SEM images show that dilution makes the etch much more controllable and that the recess offset length can be controlled by altering the etch time. Further studies on wafers A1408 and A1409 revealed that the optimum time for a 30nm offset is 25 seconds.

From SEM inspection using a Hitachi S900 High Resolution Scanning Electron Microscope the etch depth was found to be 30nm with an uncertainty of ±5nm, which agrees well with the cap thickness of 32.5nm given by the MBE growth sheets. In order to be assured of the selectivity of the each the following experiment was carried out.

A pattern containing squares and lines ranging in size from 1 to 10μm were written in PMMA resist on a GaAs/$Al_{0.3}Ga_{0.7}As$ superlattice with a thickness of the GaAs surface layer of 10nm and a 25nm $Al_{0.3}Ga_{0.7}As$ layer below. This structure was used because $Al_{0.3}Ga_{0.7}As$ is the material used for in the etch stop layer of A1408 and A1409. Samples were etched for various times ranging form 10 seconds to 2 minutes. The resist was stripped then the depth measured using a Vecco Dimension Atomic Force Microscope (AFM). Figure 5.9 shows the etch depth versus time.

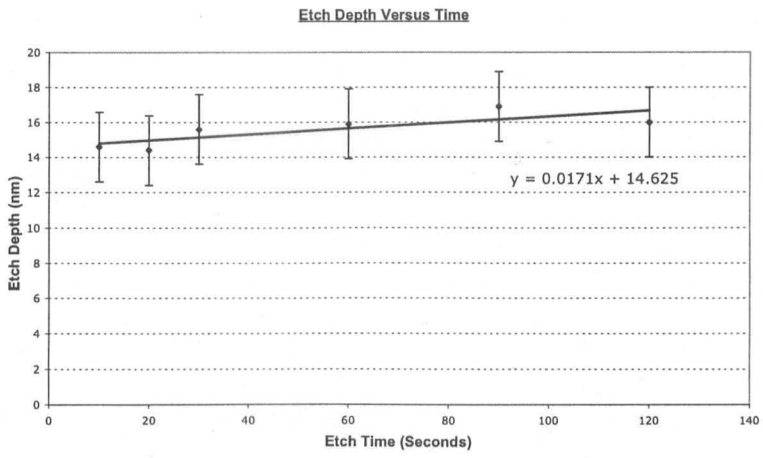


Figure 5.9 Measured etch depth versus time, on GaAs/AlGaAs superlattice. The etch rate of $A_{0.3}Ga_{0.7}As$ is found to be 0.017nm/sec.

This shows that the upper GaAs layer is removed within the first 10s, Thereafter, the etch rate slows dramatically, with less than 2nm etched in the next 2 minutes. This demonstrates the very high selective etching of GaAs over $Al_{0.3}Ga_{0.7}As$ using this etching system. The etch rate of the $Al_{0.3}Ga_{0.7}As$ layer was found to be 0.017nm/sec and therefore, in most cases, it can be taken to be negligible.

The etched surface was also studied by AFM to investigate the uniformity of the etching. Figure 5.10 is a 3D surface plot of the feed of a gate etched in succinic acid for 25 seconds. The etched surface has an RMS roughness of 1.6nm.

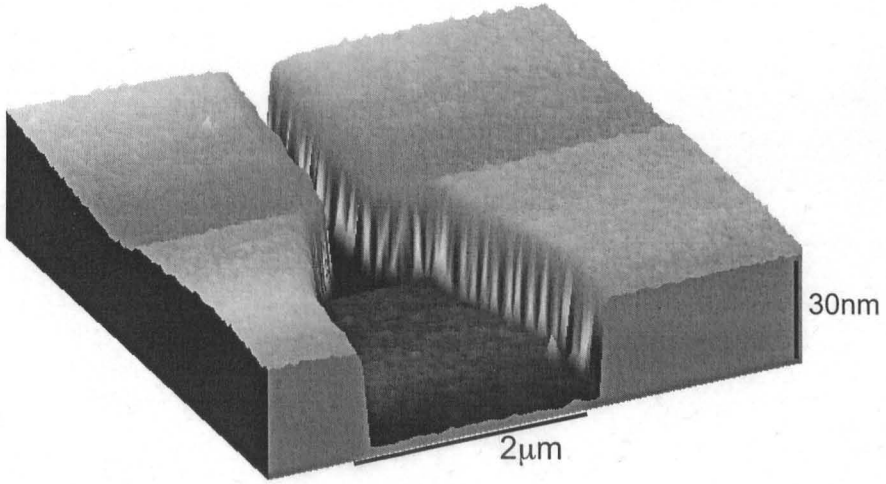


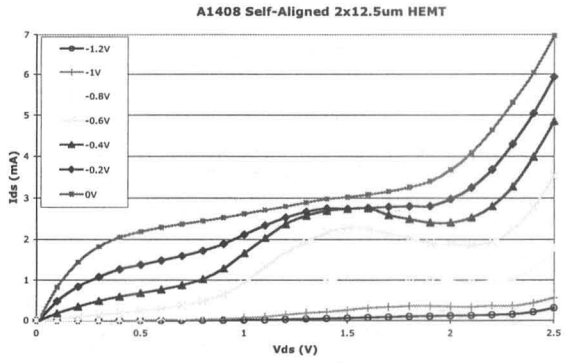
Figure 5.10 3D surface plot of gate head etched into A1409 showing smooth etched surface.

5.3.1 Succinic Acid etching in other material systems

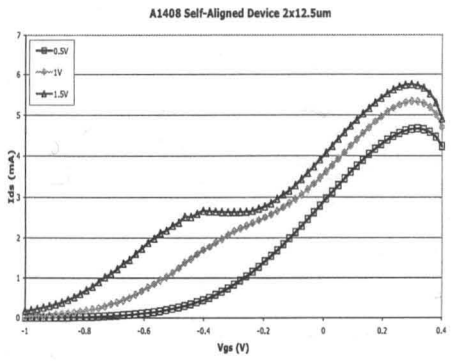
Although the etch was originally developed for the selective removal of an InGaAs cap in the non-annealed GaAs pHEMT system, it has shown itself to be a very versatile process. It has been used by the Ultra-Fast Systems group as the standard recess etch for GaAs pHEMTs, with a GaAs/AlGaAs cap/etch stop layer. Using this process the fastest GaAs pHEMT devices in the world have been fabricated that use a T-gate with a length of 50nm. [5.24] In addition it has also been used for metamorphic and lattice matched HEMTs using a $In_{0.53}Ga_{0.47}As/In_{0.52}Al_{0.48}/As$ cap/etch stop layers. [5.25]

5.3.2 Post Etch Treatments

As mentioned above, a number of sets of devices were fabricated on different materials using the succinic acid based etch. In the lattice matched and metamorphic material systems with indium concentration of over 40% a threshold voltage of –1.5V was required to fully deplete the channel compared with –0.5V when an orthophosphoric acid based etch was used. In the non-annealed GaAs pHEMT material the DC characteristics demonstrated an even more complicated behaviour. At low source drain bias the devices showed classical transistor behaviour, however as the drain bias increases to around 1V is becomes increasingly difficult to pinch off the device and modulate the channel current at certain gate biases. This is not the standard breakdown phenomena that is seen at a drain bias of more than 2V. This behaviour is shown in Figure 5.11(a) for a 2x12.5μm device fabricated on A1408 material although similar behaviour is also seen on A1409 material. The transfer characteristic of the same device shown in Figure 5.11(b) also demonstrates the effect.



(a)



(b)

Figure 5.11 DC Characteristics of 2x12.5μm Self Aligned HEMT on wafer A1408. (a) The output Characteristics (b) Transfer Characteristics.

It can be seen that the for higher drain bias a hump appears in the transfer curve of Figure 5.11(b), which has the effect of lowering the threshold voltage at the particular drain bias and reducing the transconductance. Two mechanisms have been proposed to describe the "kink effect" these have been high field charge traps in the buffer [5.26] or impact ionisation of the electrons in the channel [5.27]. Impact ionisation occurs when the electron have a greater energy than the bandgap, an electron may scatter with an lattice atom and provide a valence electron to escape into the conduction band leaving a hole, the electron flows to the gate where as the hole pass into the barrier layer and either collected by the gate or accumulate at the extrinsic source modifying the electric fields within the device. The collection of holes by the gate produces a negative gate current which Menozzi [5.28] notes can provide an indicator for impact ionisation "*A negative gate current (IG) component thus accompanies the presence of significant impact ionization. For good, non-leaky gates this component typically dominates the HEMT's IG at high drain bias and represents a useful and straightforward indicator whereby impact ionization can be detected, since under these conditions the IG-VGS curve takes on a typical bell shape*".Figure 5.12 shows the gate current of an affected device measured at a number of drain biases.

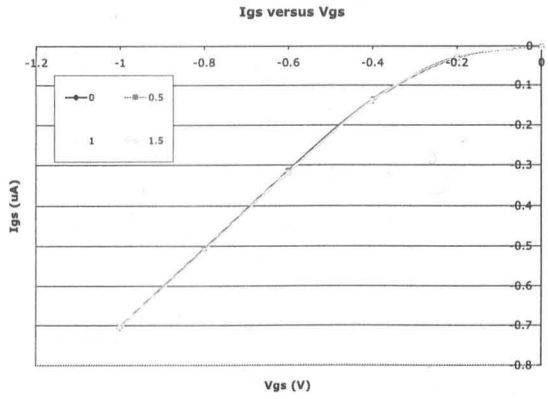


Figure 5.12 Gate Current versus Gate Voltage with applied drain bias from 0 to 1.5V

It is seen that there is no bell shape suggesting that the effect observed is not caused by impact ionisation.

The etch is selective because it does not remove the aluminium containing layers, this is because the succinic acid within the solution is unable to dissolve the aluminium rich oxides from the surface. It was thought that the recess leaves a large number of traps at the etch surface, initially these traps are unfilled, however at a high electric field ($2x10^6$V/m) the large number of electrons in the cap region have enough energy to pass into the AlGaAs etch stop layer, here they fill the traps, reducing the size of the depletion region, this lowers the potential of the channel allowing increased current flow.

Work had previously been carried out within the Ultrafast systems group using hydrofluoric acid to improve the DC performance of the GaAs pHEMT devices after dry etching [5.29]. This work used a similar $Al_{0.3}Ga0_{0.7}As$ etch stop layer to that in the non-annealed GaAs pHEMT. It was found that the gate characteristics of the dry-etched device improved if the sample was rinsed in hydrofluoric acid before gate metallisation. Although the etch chemistries were different it was decided to test the use of HF as a post etch for the succinic acid etch.

An experiment was devised to investigate the possible benefits of a post etch treatment including an HF rinse before metallisation. In addition to the HF rinse the standard de-oxidising clean of 4:1 H_20:HCl was tested with no post etch clean as control. A sample was prepared on

wafer A1607 (this is a later material with an identical design to A1408) with 120nm gate lithography, followed by a 25s succinic acid gate recess etch.

The sample contained four cells each with 80 devices, two of these cell were metallised without any treatment, one of the cells was dipped in 10:1 H_20:HF for 30sec with a H_20 rinse, while the final sample was rinsed in 4:1 H_20:HCl for 30sec with a H_20 rinse directly before the standard Ti:Pd:Au 15:15:180nm metallisation was deposited simultaneously on all samples. Lift-off was performed in warm acetone, the physical yield was then determined using an optical microscope. The sample which had been subjected to the HF treatment had the highest mechanical yield, with 75% of the devices having both gates present. The sample with HCl rinse had 59% mechanical yield. The sample without a post-etch treatment had only 16% mechanical yield. The physical yield measured in this way is a crude measurement of the gate process, however it is an important one, as transistors without a gate will not function.

The devices were completed with non-annealed ohmic contacts and RF bondpads. Next the transistor were characterised at DC using an Aglient 4155C semiconductor parameter analyser with Cascade Microtech V-band on-wafer probes to minimise oscillations issues. Figure 5.13 shows both I_{ds}/V_{ds} and I_{ds}/V_{gs} characteristics of the HCl and HF treated devices with a width of 2x50μm.

PAGE NUMBERING AS IN THE ORIGINAL THESIS

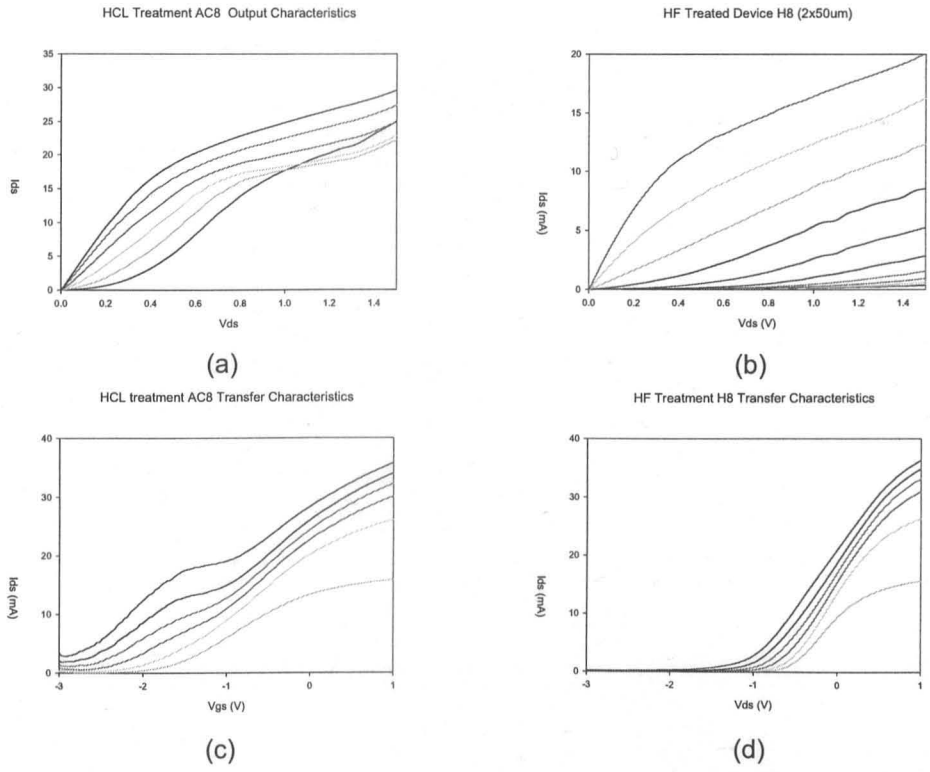


Figure 5.13 DC behavoiur of 2x50μm HEMTs (a) Output Characteristic of device treated with a 30sec HCl post etch clean. (b) Output Characteristic of device treated with 30sec HF post etch clean. (c) Transfer Characteristic of device treated with 30sec HCl post etch clean. (d) Transfer Characteristic of HF treated device.

The output characteristics were measured by sweeping V_{ds} from 0 to 1.5V and steping V_{gs} from –2V to 0V in steps of 0.2V, the transfer characteristics shown in (c)&(d) was obtained by sweeping V_{gs} from –3 to 1V and increasing V_{ds} from 0.25 to 1.5V in steps of 0.2V. It is seen that in the HCl post etch treatment the device does not pinch off at high drain biases and at bias above 0.75V a hump appears in the transfer characteristic and the threshold voltage shifts accordingly as had been observed for the untreated surface. This behaviour is not seen for the sample that had been treated with a HF rinse, the transfer characteristic shows that even with a drian bias of 1.5V the current drops off smoothly without any threshold voltage shift. Correspondingly the transconductance is much larger.

The mechanism causing the non-ideal behaviour of the HEMT fabricated with the original succinic acid etch is unclear; however it is thought to be directly related to the chemistry of the surface left by the etch process. The surface has been observed with the aid of AFM and SEM techniques and the etched surface look smooth. It has however been seen from the experiment

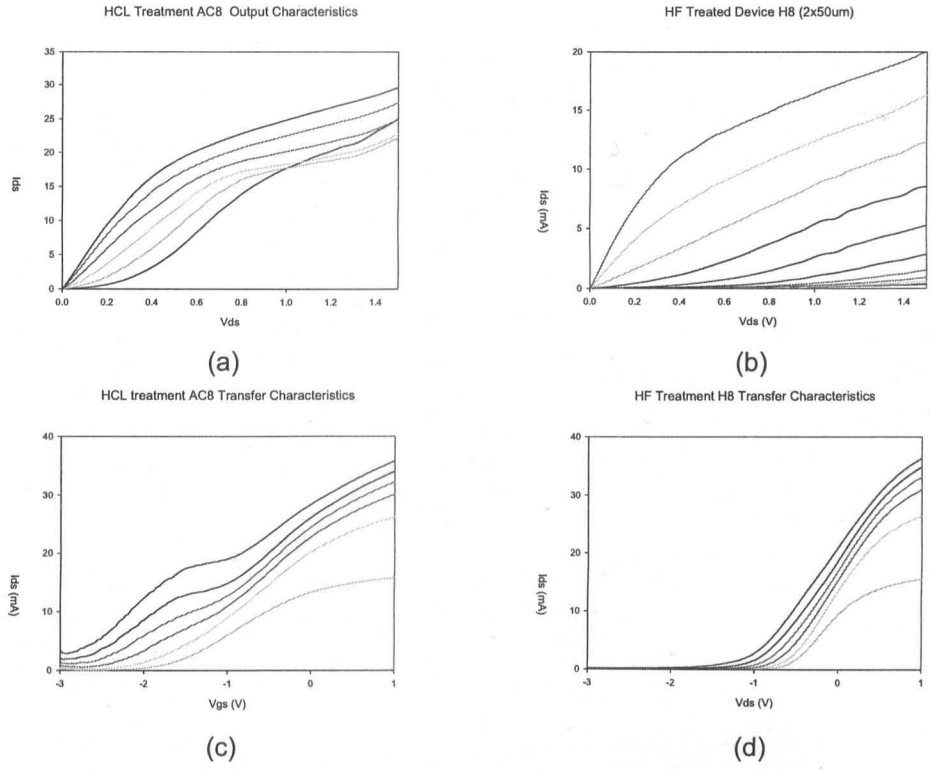


Figure 5.13 DC behavoiur of 2x50μm HEMTs (a) Output Characteristic of device treated with a 30sec HCl post etch clean. (b) Output Characteristic of device treated with 30sec HF post etch clean. (c) Transfer Characteristic of device treated with 30sec HCl post etch clean. (d) Transfer Characteristic of HF treated device.

The output characteristics were measured by sweeping V_{ds} from 0 to 1.5V and steping V_{gs} from –2V to 0V in steps of 0.2V, the transfer characteristics shown in (c)&(d) was obtained by sweeping V_{gs} from –3 to 1V and increasing V_{ds} from 0.25 to 1.5V in steps of 0.2V. It is seen that in the HCl post etch treatment the device does not pinch off at high drain biases and at bias above 0.75V a hump appears in the transfer characteristic and the threshold voltage shifts accordingly as had been observed for the untreated surface. This behaviour is not seen for the sample that had been treated with a HF rinse, the transfer characteristic shows that even with a drian bias of 1.5V the current drops off smoothly without any threshold voltage shift. Correspondingly the transconductance is much larger.

The mechanism causing the non-ideal behaviour of the HEMT fabricated with the original succinic acid etch is unclear, however it is thought to be directly related to the chemistry of the surface left by the etch process. The surface has been observed with the aid of AFM and SEM techniques and the etched surface look smooth. It has however been seen from the experiment

described above that the use of an HF rinse prior to metalisation is beneficial to device performance. The advantage is twofold, the adhesion of the gates to the surface is impoved, enhancing the device yield and additionally the rinse improves the electrical behaviour of the device.

A further refinement was made to the etch process, as it was noted that Van der Pauw test structures with a large etched area did not etch uniformly leaving a shallow pitted surface, as shown in Figure 5.14.

Figure 5.14 SEM Image of corner of 1x1mm Van der Pauw sample etched for 25s. The pitted etch surface is seen.

It was reported in [5.30] that agitation of the etch solution reduces the surface roughness of the etch surface. Samples were etched using a magnetic stirrer to agitate the solution during etching, this was seen to improve the etching. Although strictly unnecessary to the etching of the 120nm T-gates, agitation was incorporated into the etch process as it was anticipated that with the reduction of gate length below 100nm, the etch may be limited by the diffusion of etch components into the gate region.

5.4 Realisation of Self-Aligned HEMTs

Having developed the non-annealed ohmic contact and the selective wet etch of InGaAs and GaAs over AlGaAs it was possible to realise a GaAs pHEMT using self-aligned ohmic contacts. These devices were originally based on the standard Ultra-Fast Systems 120nm GaAs pHEMT process described in Appendix A. This was used as the basis because the 120nm T-gate process is a very stable, repeatable process, which provides a benchmark for further development. The self-aligned process deviated from the standard process in a number of ways, namely the sequence of lithography steps, the use of succinic acid as the recess etch (although this has now become standard) and finally the ohmic contact design.

In the standard process described in section 4.4.1 the ohmic level is written directly after the marker layer, the sample is then annealed and subsequently isolated by mesa etching. However since the ohmic level in the self-aligned device is designed to be deposited after the gate level the isolation level must be written immediately after the marker layer. This has the disadvantage of forcing the isolation to be monitored by measured depth rather than the favoured electrical monitoring. This limitation has been overcome however since the self-aligned process is an intrinsically low temperature process. It is therefore possible to metallise the electron beam alignment markers using the Au/Ge/Au/Ni/Au scheme and still maintain high quality markers. This allows the isolation to be carried out using the electrical method and provides information from test structures at a much earlier stage in the process flow. The design of the ohmic level is modified in the self-aligned process, as the source-drain gap need not be specifically defined.

Devices were fabricated on a number of substrates, all with an InGaAs cap layer suitable for non-annealed ohmic contacts. The initial devices were fabricated on wafer A1408, as it had consistently demonstrated the lowest contact resistance of the non-annealed wafers tested.

The fabrication process follows the flow described above. The duration of the gate recess etch was 25 seconds, which has been seen to give the desired recess offset. The ohmic contact recipe was 14/14/14/11/50nm Au/Ge/Au/Ni/Au with a overall thickness of 103nm. A SEM micrograph cross-section of the test structures fabricated along with the device wafer (Figure 5.15) clearly shows that the height of the stalk is greater than the ohmic contact thickness and that there is separation of the gate and the ohmic contacts. The devices were completed with 400nm thick bondpads.

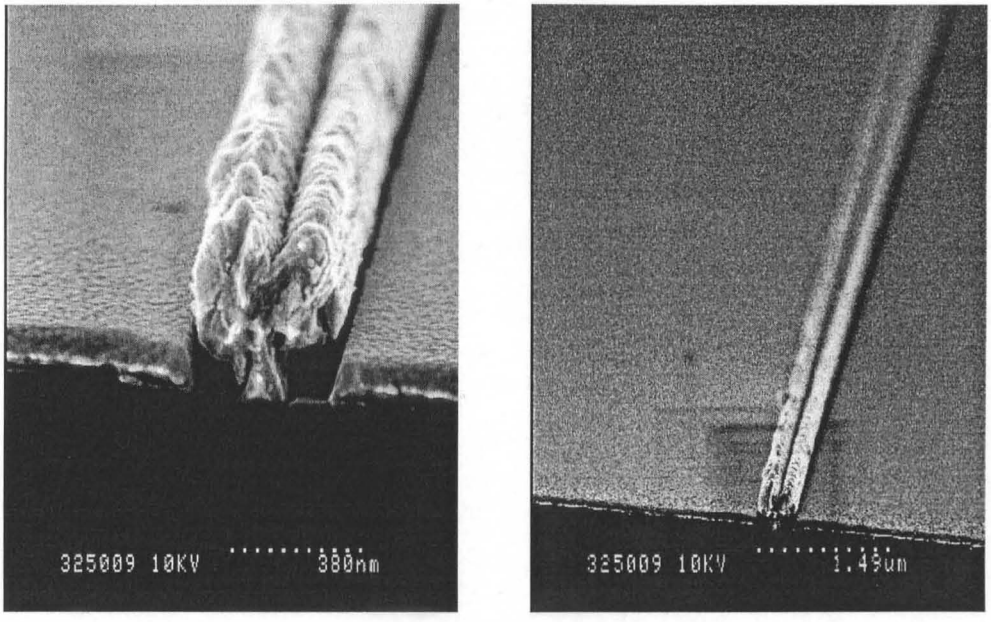


Figure 5.15 Cross section SEM micrographs of the Self Aligned Ohmic contacts. The close up shows the splitting of the ohmic metal over the gate and the ohmic thickness in comparison with the gate stalk height. The lower magnification image shows the gate-ohmic gap across the width of the gate.

The device cell contains a number of test structures to characterise the material and process during fabrication. One of these monitors, the Van der Pauw structure [5.31], allows the mobility and the 2DEG carrier concentration to be measured. The mobility of the material was found to be 2868cm^2/V-s and the sheet density was 1.17x10^{13}cm^{-2} with the cap in place, which were slightly low for GaAs pHEMT material.

The DC behaviour of devices was measured on wafer using Cascade Microtech V-band probes and a HP4145 semiconductor parameter analyser. The output characteristics, (I_{ds}/V_{ds}) and the transfer characteristics were measured for each device, from these measurements important parameters such as transconductance and threshold voltage were determined. As stated

previously, the threshold voltage of a device is the voltage that needs to be applied to the gate to fully deplete the electrons within the channel. However in practice there is typically a small leakage current, which depends on the width of the device. The threshold voltage is therefore taken to be some small percentage of the saturation current, in this case 2% [5.32]. The mean value of the threshold voltage was found to be –0.99V with a standard deviation of 0.18V. The slightly large standard deviation is caused by the variation in the threshold voltage with drain bias caused by surface states. This shows that despite the problems with the etch discussed in section 5.2.2 the etch is uniform across the sample.

The output and transfer characteristics of a 2x100μm device is given in Figure 5.16, the dependence on the transconductance with applied gate voltage is shown with the transfer plot. The device shown is part of the same batch as the results shown in Figure 5.10 and displays the same characteristic hump in the output characteristics.

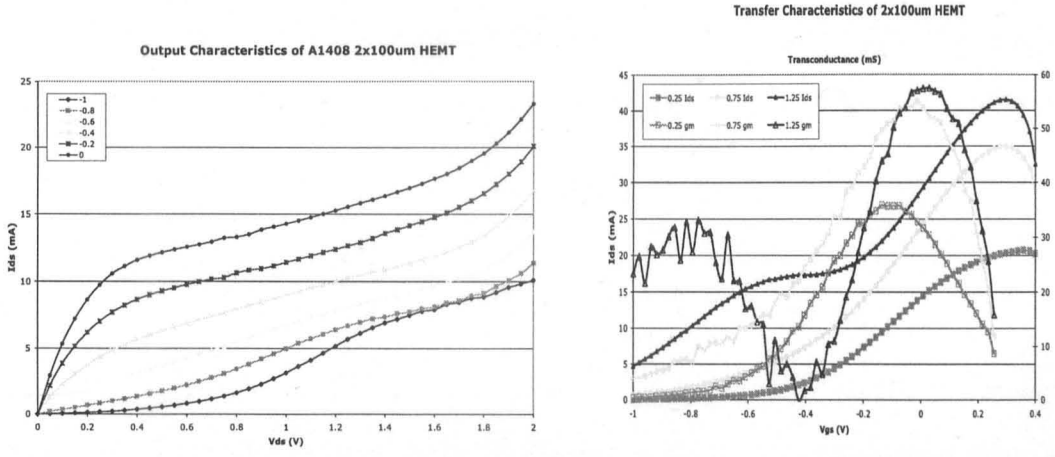


Figure 5.16 DC behaviour of a 2x100μm device fabricated on A1408 (a) Output Characteristics (b) Drain current and transconductance versus gate bias. Devices did not receive an HF dip.

It is noted that the saturation current of the devices is low, with a typical current density of around 100mA/mm, this compares poorly with the standard GaAs pHEMT that was fabricated several years ago, which displayed a current density of around 500mA/mm, the difference between the 2 devices, Figure 5.17 shows the two devices with similar bias conditions.

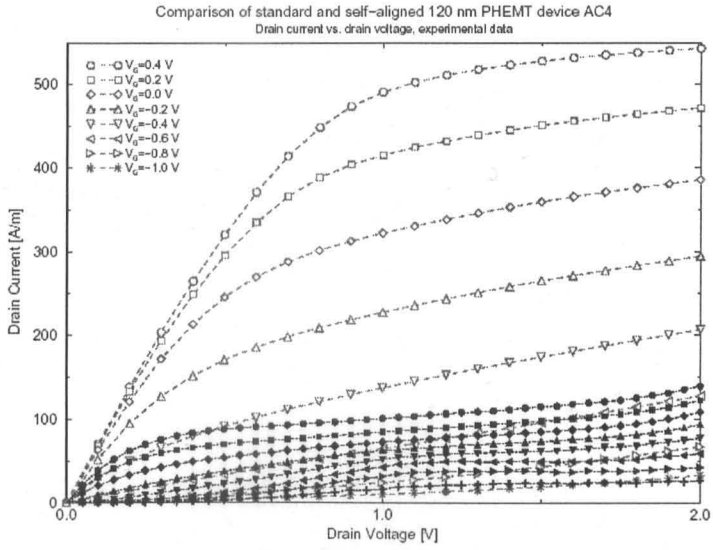


Figure 5.17 Comparision of the normalised drain currentr versus drain voltage of the Self-Aligned and standard GaAs pHEMT.

It is seen that the current density on the self-aligned device is around 5 times lower than the pHEMT devices; correspondingly the transconductance of the device is also lower. However in addition to the low transconductance the output conductance is also much lower for the self-aligned device. The large difference between the standard pHEMT data and the self-aligned results is thought to be due to material growth in particular the concentration on the active δ doping. The two wafers have nominally similar designs however these results have not been able to be repeated in recent years.

5.5 RF Characterisation

The RF characterisation of all devices was carried out using the following equipment, Anritsu 360B Vector Network Analyser (VNA) switchable to cover V-band (0.04-60GHz) or W band (67-110GHz) frequency ranges. The devices are probed on-wafer using V or W band Ground-Signal-Ground-Source probes from Cascade Microtech or Pico-probes respectively. The devices were biased using an Agilent HP4155C semiconductor parameter analyser connected via bias-tees within the VNA test set.

The VNA measures the magnitude and phase of the S-parameters of the device under test (DUT). However since this is measuring the response of the device to signals as opposed to the

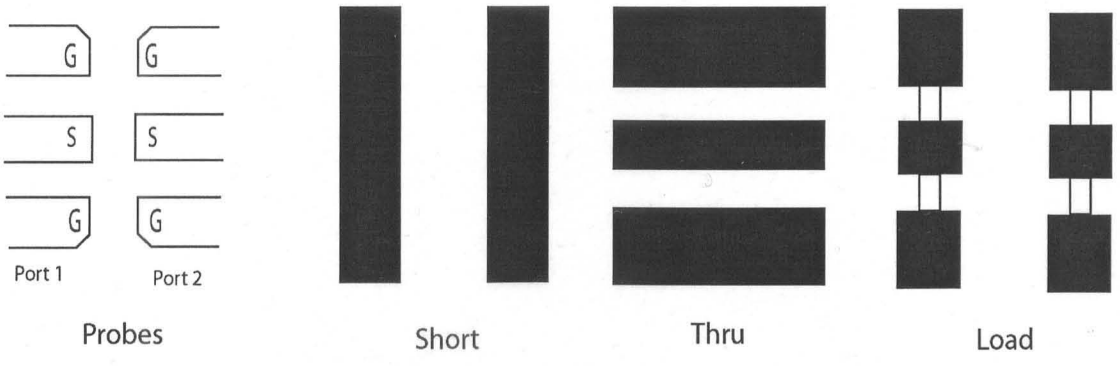
signals themselves care must be taken to determine where exactly the DUT stops and the measurement systems begins. In on-wafer measurement this boundary is referred to as the reference plane. The position of this reference plane is determined by the calibration of the VNA. The VNA is calibrated by measuring a series of standard devices with known behaviour, which are then used to build a 12 term error model [5.33] that relates the measured response to the known behaviour of the standard devices. This model is then subtracted from all subsequent measurements to give the correct measurements. This calibration removes non-ideal effects of the probes and cables between the VNA and DUT as well as removing imperfections within the VNA. Since all of the effects from the tips of the probes to the VNA are removed by calibration the reference plane shifts to the tip of the probes. (This is the case for most standards, however there are exceptions which will be discussed later)

5.5.1 Calibrating the VNA

There are a number of different ways to calibrate a VNA, depends on the number and types of standards, substrates and algorithms used, this section will introduce these techniques and discuss why they are used before presenting the calibration method used throughout this work.

Calibration Standards

Calibrations are usually carried out with some permutation of four calibration standards, Open, Short, Thru, Load. The designs of these standards are given in Figure 5.18 along with a schematic of the Ground-Signal-Ground probes configuration. The standards are formed by the deposition of gold patterns onto the substrate.



(a) (b) (c) (d)

Figure 5.18 Illustration of the Ground-Signal-Ground Probes and the designs of the calibration standards.

The Open standard is achieved by simply lifting the probes from the substrate. The short standard is a vertical bar of gold that connects all three terminals of the probes as shown in Figure 5.18(a). The Thru standard is a short length of 50Ω transmission line, connected the two sets of probes as shown in (b). The load standard is comprised of a NiCr resistor between the terminals of each probe. The resistance is carefully trimmed to be within 0.3% of the required dc value (50Ω). The calibration standards are normally measured on an Impedance Standard Substrate (ISS), this is an alumina substrate 1mm thick and with dielectric constant of 9.9. The advantage of the ISS is that since it has well defined standards, measurements calibrated using an ISS can be easily compared. Sometimes it is useful to have the calibration standards on the same substrate as the DUT to reduce uncertainties due to the difference in permittivity between alumina and semiconductor such as GaAs. If this is done it is also possible to include the CPW transmission lines as used in the HEMT into the calibration standards. Figure 5.19 shows a composite optical micrograph of CPW calibration standards fabricated on a GaAs substrate. This enables the effects of the CPW lines to be included in the calibration and therefore shifts the reference plane from the tips of the probes to the end of the transmission line reducing uncertainties in de-embedding the device and feed-lines.

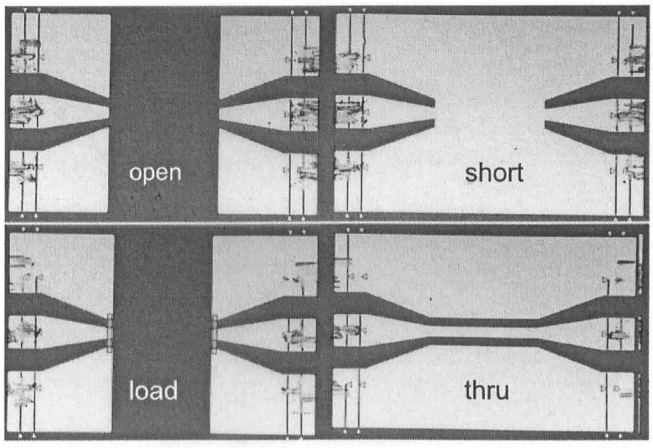


Figure 5.19 Composite Optical Micrograph of CPW calibration standards fabricated on a GaAs wafer.

Calibration Algorithms

A number of calibration algorithms are available for calibration of a VNA, each has it own relative merits. In this section two calibration strategies, Short-Open-Load-Thru, (SOLT) and Line-Reflect-Reflect-Match (LRRM) are introduced.

The SOLT method, a widely used common form of calibration, consists of measuring Short, Open and Load calibration standards at both ports of the VNA, together with a Thru standard connecting the ports to complete the calibration. [5.34] Using this method all four standards must be accurately known [5.35]. For on-wafer calibrations, the response of the standards is dependent on the exact placement of the probes [5.36], which brings an uncertainty to the measurement from calibration to calibration. The LRRM calibration uses the same standards as the SOLT however it is not as dependent on the Open and Short that are used as reflection standards. The Thru standard needs to be specified with both the thru delay and the loss and the dc resistance of one of the Load standards. The LRRM technique has been shown to be less sensitive to probe placement [5.33].

The calibration method used throughout this work was the LRRM method on ISS substrates for maximum repeatability.

The measurements are carried out using a manual probe positioning and automated bias and data acquisition software which measures the device throughout a user set bias range which enables the optimum operating point to be found and allows for the non-linear modelling of the devices.

5.6 Small Signal Modelling and De-embedded S-Parameters

The CPW bond pads will affect the measured performance of the HEMT at RF frequencies. It is therefore necessary to remove their effects, the process of de-embedding. This can be achieved by removing the CPW lines during calibration, however a small length of line tends to remain that can effect the measurements. The other method is more time consuming but is ultimately more insightful, this uses a small signal model as discussed in section 2.3.5 with the addition of lengths of transmission line at port 1 (Gate) and port 2 (Drain) to model the CPW bondpads. This model is found in the following way.

The initial model is found using a program called Lysander, this was written within the Ultrafast Systems group by Mike Taylor. Lysander takes the S-parameter data and fits a 18 parameter lumped element model over the complete bias range. This model is used as the basis for a model created in Microwave Office which is then manually fine tuned to match the small signal model element values to the magnitude and phase of all four S-parameters. The model is considered to be valid when it matches the measured S-parameters with physically realistic element values. The transmission lines located at either port of the device to model the CPW lines are then removed from the model, to leave the de-embedded model. It is possible to use this model to extrapolate the behaviour of the device at frequencies higher than the measurement range. Figures of Merit such as f_t and f_{max} can be found by extrapolating H21 and MAG to unity. Extreme care needs to be taken when producing the model to check that it is both physically correct and numerically realistic otherwise the results produced will be invalid.

5.6 RF Measurement of Self-Aligned Devices

Using the method detailed above the self-aligned devices fabricated on wafer A1408 were characterised in the frequency range of 0.04 to 60GHz. The optimum performance was achieved at a bias point of 1.25V on the drain and –0.2V on the gate. The de-embedded plots of H21 and MAG are shown in Figure 5.20. If these plots are extrapolated to unity gain for each case values of 150GHz and 180GHz are obtained for f_t and f_{max} respectively. This compares very favourably with the results from the standard GaAs pHEMT which had an f_t of 120GHz and f_{max} of 180GHz. This difference in performance at DC and RF performance suggests that the poor DC performance is caused by trapping of electrons. The time constant associated with these traps will be much longer than the period of the RF signals hence the trapping effect will not be noticed at high frequencies.

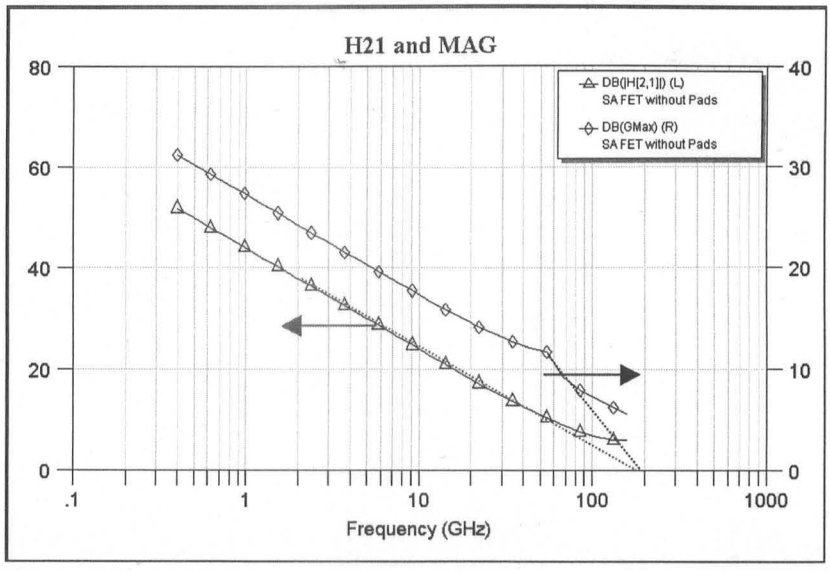


Figure 5.20 De-embedded plots of |H21| and MAG/MSG for a self-aligned GaAs pHEMT.

This work was the first demonstration of a self-aligned process at Glasgow University. It showed that it was possible to fabricate devices using the standard 120nm T-gate process. However, the results had not been conclusive that the RF performance enhancement was due to the self-aligned design rather than material issues. A new batch of devices were fabricated that had both standard and self-aligned device designs in the same cell in order to determine experimentally the advantages of the self aligned design. The new cell was to be fabricated on

wafer, A1607, which had the same design as A1408. The new cell design contained 40 standard and 40 self-aligned devices ranging in width from 25-200μm. The design of the transistors was identical with the exception of the ohmic layer. The mobility and the sheet density with the cap layer intact was found as before, to be 3150cm^2/Vs and 1.18x10^{13}cm^{-2} respectively. The mobility is 10% higher than A1408 with a similar sheet density. The ohmic metal was deposited which had a thickness of 100nm, this caused a number of shorts lowering the electrical yield to 20%. However there were enough operating devices to provide statistically meaningful comparisons to be made between the self-aligned and standard devices.

The DC characteristics of self-aligned and standard devices are shown in Figure 5.21. These HEMTs are two finger devices with a combined gate width of 50μm they are located close to each other in the device cell.

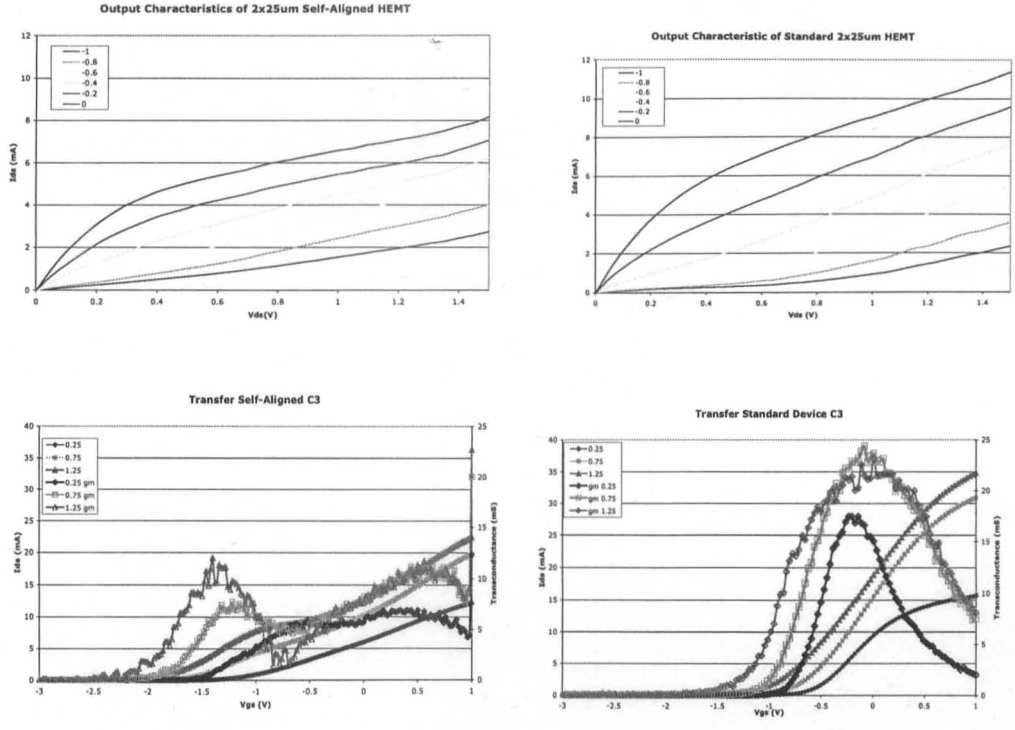


Figure 5.21 DC characteristics of standard and self-aligned 2x25μm devices fabricated on wafer A1607.

It is seen from the output characteristics of the devices that the standard device has a larger current flow compared with the self-aligned device. The self-aligned also requires a greater threshold voltage to pinch off the device, at higher drain bias it becomes increasingly hard to

fully pinch the device off. This behaviour resembles breakdown as the current increases despite the applied gate bias. The transfer characteristics show the same features. The threshold voltage of the self aligned device has shifted by around –0.5V. The standard device has a single peak in the transconductance at around V_{gs}=0V for all drain biases, the peak transconductance was 460mS/mm achieved with a gate drain bias of 1.25V. The current control in the self-aligned device is much more complex. At low drain bias the behaviour is similar to the behaviour of the standard device, albeit with a much reduced current. As the drain bias is increased to 0.75V a hump appears in the I_{ds}/V_{gs} curve, leading to a second peak forming in the transconductance curve. This effect continues as the drain bias increases; with a drain bias of 1.25V the secondary transconductance peak is larger than the original peak with a value of 250mS/mm at a drain bias of –1.5V.

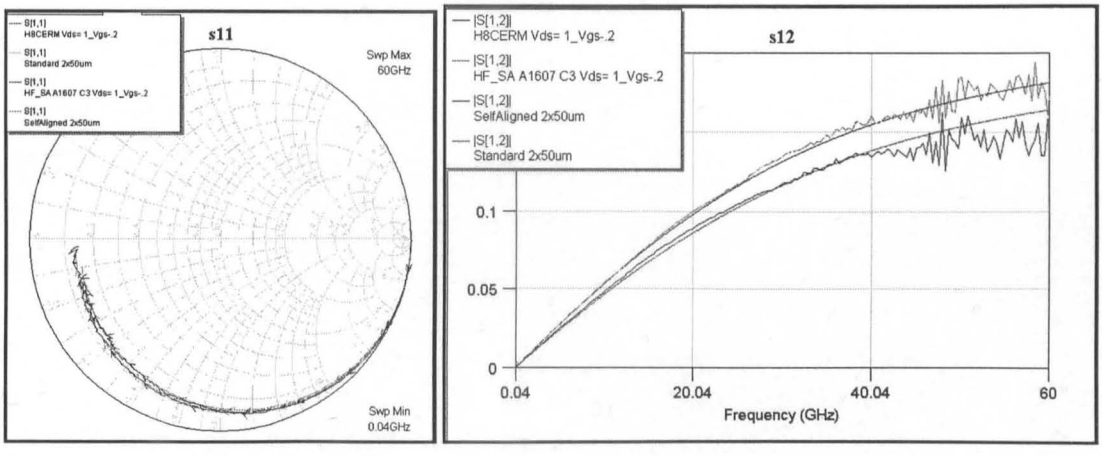
It is noted that this is similar behaviour to that observed in the devices fabricated on A1408. It was seen that the use of the HF treatment reduced this behaviour in standard HEMT designs. This supports the earlier argument that this behaviour was caused by the surface states and that these are reduced using HF. However with the higher fields in the self-aligned devices the problem resurfaces.

The devices were tested between 0.04-60GHz and a small signal model was produced from the S-parameters over a range of bias voltages as described previously. More information on the behaviour of each device at high frequencies can be extracted from the equivalent circuit models. The model parameters for two devices are shown in Table 5.3, both devices are 2x50μm wide, the standard design has a 1.5μm wide source-drain gap.

	Self-Aligned Device	Standard Device
Cgs (fF)	31	31
Gm (mS)	32	37.4
Cgd (fF)	9.11	7.61
Gds (mS)	7.6	8.2
Tau (pS)	0.46	0.46
Cds (fF)	11	11
Cpdg (fF)	2.65	2.65
Cpg	1.4	1.4
Cpd (fF)	13.5	13.5
Rs (Ω)	0.8	1.45
Ls (pH)	1	1
Rg (Ω)	2	3.3
Lg(pH)	1.1	1.11
Rd (Ω)	0.5	0.5
Ld (pH)	1	1
Rin (Ω)	1.73	2.5
Rgd	43.3	33.4

Table 5.3 Table showing the small signal model values of a 2x50μm Standard and Self-Aligned pHEMTs.

The two models were achieved in the following way, a model was produced that matched the S-parameters of the standard design. This model was then fitted to the self-aligned by adjusting key elements of the circuit such as the transconductance, g_m, output conductance, g_{ds}, and the parasitic resistances, R_s, R_d and R_g. This method allows the relative differences between the models to be studied rather than creating two independent models which may individually fit the data but do not allow valid comparisons to be made.



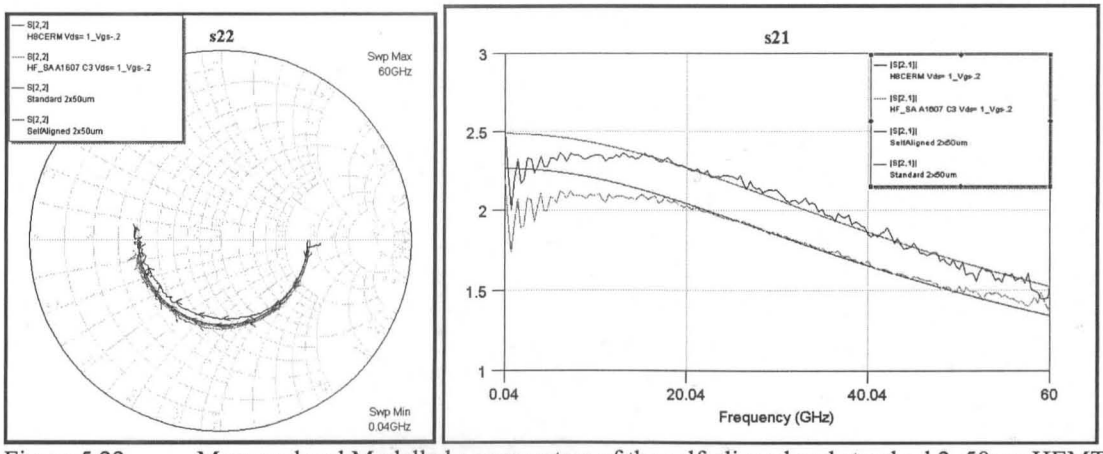


Figure 5.22 Measured and Modelled s-parameters of the self-aligned and standard 2x50μm HEMTs fabricated on A1607.

The modelled and measured S-parameters for the self-aligned and standard device are compared in Figure 5.22. The key difference between the two sets of S-parameters is the difference in the input transmission coefficient, S21, and the output reflection coefficient, S22. These suggest that the both the transconductance and the output conductance need to be lowered in addition to a reduction of R_s and R_d. These adjustments are physically valid, it was seen from the DC data that the self-aligned devices demonstrated a lower g_m. The reduction of the gate-drain separation will increase the electric field in the gate region and so the output conductance will decrease. The reduction of the source-drain gap will reduce the access resistances R_s and R_d. It was also seen that the turnover of the magnitude of S12 was less in the self-aligned device, to achieve this behaviour it was necessary to increase the gate-drain capacitance, this may be caused by the increased depletion region on the drain side of the gate. Due to the difficulties of extracting the source and drain resistances using this method these values should be considered as qualitive rather than quantitive, allowing for a comparison between the two designs.

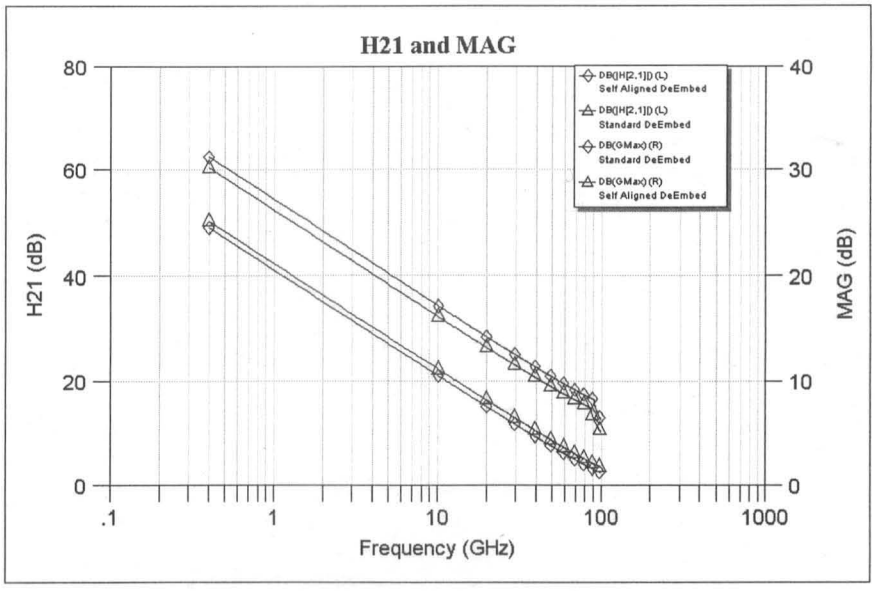


Figure 5.23 De-embedded H21 and MSG/MAG of the 2x10μm standard and self-aligned HEMTs fabricated on A1607.

The plot of the de-embedded H21 and MAG versus frequency is shown in Figure 5.23 for two 2x50μm devices biased at the optimum bias point for maximum gain. This graph shows that overall the standard pHEMT displays slightly better performance. For instance by consideration of the short circuit current at 10GHz, extrapolation at 20dB/decade this translates into a higher cut-off frequency of 160Ghz compared with 120GHz for the self-aligned device. f_{max} of the self-aligned devices is also lower than that of the standard design, with a value of 160GHz compared with around 180GHz for the standard device. The slightly better performance of the standard device is attributed to the larger transconductance in the device that was seen from the DC data.

From these results it was concluded that at present using the GaAs pHEMT structure there was no evidence of the expected performance benefits of the self-aligned geometry. It was seen that the self-aligned devices demonstrated lower parasitic resistance as expected, however, the overall performance was badly affected by the lower transconductance of the self aligned devices caused by the surface states in the etched region and the high electric fields in the self-aligned device. However in the course of the development of the self-aligned pHEMT two key technologies have been developed, the selective succinic acid based wet gate recess etch. This

has been proved to be very successful and has been become the standard process for the gate recess etching of GaAs pHEMT, metamorphic, and lattice matched HEMTs within the University of Glasgow. The second process developed was the non-annealed ohmic contact using a delta-doped InGaAs cap layer. This has two benefits, firstly it is an intrinsically low temperature process, which has been suggested to reduce damage to the growth layer and provides better performance [5.37]. This explains the enhanced RF devices performance of the standard devices fabricated using the non-annealed contacts compared to previous pHEMT devices. The second key benefit of the non-annealed ohmic contact is the freedom given to the process flow. This enables the gate to be written at any stage, even first. This freedom has allowed novel devices to be fabricated such as the 120nm pHEMT fabricated by Nano-Imprint Lithography [5.38,5.39].

5.8 References

[5.1] S. Babiker, et al., "*Complete Monte Carlo RF Analysis of "Real" Short-Channel Compound FET's*", IEE Trans. Elect. Dev., **45**, 1998, p1650

[5.2] U.K. Mishra et al., "*Novel high performance self-aligned 0.1μm long T-gate AlInAs/GaInAs High Electron Mobility Transistors*", IEDM Technical Digest, 1989, p101

[5.3] L.D. Nguyen et al., "*650Å Self-Aligned Gate Pseudomorphic $Al_{0.48}In_{0.52}As/Ga_{0.2}In_{0.8}As$ High Electron Mobility Transistors*" IEEE Elec. Dev. Lett., **13**, 1992, p143

[5.4] L.D. Nyuyen, et al., "*50nm Self-Aligned Pseudomorphic AlInAs/GaInAs High Electron Mobility Transistors*", IEEE Trans. Elec. Dev., **39**, 1992, p2007.

[5.5] N. Braslau et al. Solid State Physics, 10, 1967

[5.6] E.D. Marshall et al., "*Nonalloyed ohmic contacts to n-GaAs by solid-phase epitaxy of Ge*", Applied Physics Letters, **73**, 1998, p1670

[5.7] A.K. Rai et al. "*Alloying behavior of gold, Au-Ge and Au-Ge-Ni on GaAs* ", Thin Solid Films, **114**, 1984, p379

[5.8] T.C. Shen et al., "*Recent Developments in ohmic contacts for III-V compound semiconductors*", J. Vac. Sci. Technol., **B10**, 1992, p2113

[5.9] W. Shockley, "*Research and Investigation of Inverse Epitaxial UHF power Transistor,*", Final Technical Report, No. Al-TDR-64-207, AFAL (AFSC), W-PAFB, 1964

[5.10] H. H. Berger, "*Models for contacts to planar devices*", Solid State Electronics, **15**, 1972, p145

[5.11] H. Murrmann and D. Widmann, "*Messung des übergangswiderstandes zwischen metall und diffusionsschicht in Si-planarelementen*" Solid State Electronics, **12**, 1969, p879

[5.12] Greg Snider 1D Poisson Solver Program Website, http://www.nd.edu/~gsnider/

[5.13] D.A.J. Moran et al. "*Self-aligned T-gate InP HEMT realisation through double delta doping and a non-annealed ohmic process*" Microelectronic and Nanoelectronic Engineering 2003, Cambridge, September 2003

[5.14] N.I. Cameron et al., "*Gate recess engineering of pseudomorphic $In_{0.3}GaAs/GaAs$ HEMTs*", Electronics Letters, **32**, 1996, p770

[5.15] H.C. Kao et al. "*Selective gate recess RIE etching by CHF_3+BCl_3 in InAlAS/InGaAs HEMTs*", Proceeding of IPRM, 1997, p137.

[5.16] D.L. Edgar et al. "*94 and 150GHz Coplanar waveguide MMIC amplifiers realized using InP Technology*", IEEE MTT-S Digest, 1999, p247

[5.17] D.L. Edgar et al. *"millimeter-wave performance of InAlAs/InGaAs HEMTs using a UVIII/PMMA bilayer for 70 nm T-gate fabrication"*, European Microwave Conference 2002, Milan, Italy, October 2002

[5.18] J.C. Huang et al. "*The effect of channel dimensions on the millimeter-wave power performance of a pseudomorphic HEMT*", Proceedings of GaAs IC Symposium, 1993, 1993, p177

[5.19] K. Shinohara et al, "*Importance of Gate-Recess Structure to the Cutoff Frequency of Ultra-High-Speed InGaAs/InAlAs HEMTs*", Proceedings of IPRM 2002, 2002, p451

[5.20] C.S. Whelan et al. "*High breakdown Voltage InAlGaAs/$In_{0.32}Ga_{0.68}As$ Metamorphic HEMT for Microwave and mm-wave Power Applications*", IEEE MTT-S, 1999, p1187.

[5.21] H. Rohdin et al. "*A 0.1μm MHEMT millimetre-wave IC technology designed for manufacturability*", Solid State Electronics, **43**, 1999, p1645

[5.22] H. Fourre et al. "*Selective wet etching of lattice matched InGaAs/InAlAs on InP and metamorphic InGaAs/InAlAs on GaAs using succinic acid.hydrogen peroxide solution*", J. Vac. Sci. Technol. **B 15**, 1996, p3400

[5.23] N. Cameron, "*Compound Semiconductor Processes and Devices*", Internal Ultra-fast Systems Document.

[5.24] X. Cao et al., "*mm-wave Performance of 50nm T-Gate AlGaAs/InGaAs Pseudomorphic High Electron Mobility Transistors with f_t of 200GHz*", Proceeding of GAAS2003, 2003, p13

[5.25] D.A.J. Moran et al., "*Self-aligned 0.12 μm T-gate In.53Ga.47As/In.52Al.48As HEMT technology utilising a non-annealed ohmic contact strategy*" Proceedings of ESSDERC, Estoril, September 2003

[5.26] T. Zimmer et al. "*Kink effect in HEMT structures: A trap-related semi-quantitative model and an empirical approach for spice simulation*", Solid-State Electronics, Vol 35, p1543, 1992

[5.27] M. Somerville et al., "*Direct Correlation Between Impact Ionisation and the Kink Effect in InAlAs/InGaAs HEMTs*", IEEE Electron Devices, Vol 17, p473, 1996

[5.28] R. Menozzi, "Hot Electron Effects and degradation of GaAs and InP HEMTs for microwave and millimetre wave applications", Semiconduct, Sci. Tech., Vol 13, p1053, 1998.

[5.29] K. Elgaid et al. "*Optimisation of DC and RF performance of GaAs HEMT-based Schottky diodes*", Electronics Letters, 35, 19, 1999, p1678

[5.30] H. Rabah et al. "*Study of GAAS chemical etching in a mixture of hydrogen peroxide/succinc acid ammonia. Thiourea effect on surface roughness and on the presence of surface states after etching*", Applied Surface Science, **171**, 2001, p34

[5.31] J.H. Davies, "*The Physics of Low-Dimensional Semiconductors*", Cambridge, Cambridge University Press, 1998, p217

[5.32] R. Williams, "Modern GaAs Processing Methods", Boston, Artech House, 1990 p345

[5.33] A. Davidson et al., "*Achieving greater on-wafer S-parameter accuracy with the LRM calibration technique*", Cascade Microtech Technical Note.

[5.34] "*On-wafer vector network analyser calibration and measurements*" Cascade Microtech Application Note

[5.35] A.M.E. Safwat and L. Hayden, "*Sensitivity Analysis of calibration standards for SOLT and LRRM*", Cascade Microtech Technical Note, 2001

[5.36] K.Jones et al. "*Where are my reference planes*?", IEEE ARFTG Digest, 1987

[5.37] Y. Yamashita et al. "*High f_t 50nm Gate InAlAs/InGaAs High Electron Mobility Transistors Lattice Matched to InP substrates*", JJAP, **39**, 2000, p L838

[5.38] Y. Chen et al. "*Fabrication of high electron mobility transistors with T-gates by nanoimprint lithography*", J. Vac. Sci. Technol. B, 20(6), 2002, p2887.

[5.39] Y. Chen et al. "*High electron mobility transistors fabricated by nanoimprint lithography*", Microelectronic Engineering, 67-68, 2003, p189.

6.0 Development of Sub-100nm T-gate HEMTs

The main goal in the development of III-V HEMTs is to improve the RF performance of the devices, normally quantified by the f_t and f_{max}. There are a number of ways in which the performance of a HEMT can be enhanced. One of these methods is increasing the indium concentration of the channel. This has been seen to give performance benefits and has led to lattice matched and pseudomorphic HEMTs grown on InP. The second method used is the reduction of parasitics in the device. This can be achieved by either reducing the ohmic contact resistance or reducing the source-drain gap, so reducing the access resistance through the material. Alternatively the access resistance can be reduced by using a self-aligned design. There are limits to these methods, which will be discussed in the following section where a third method will be introduced; gate length scaling. The rest of the chapter will detail the work carried out in scaling the gate length below 100nm.

6.1 Gate Length Scaling for High Performance HEMTs.

Increasing the channel indium concentration and the reduction of parasitic resistances are both important ways of improving the performance of a device, however each has its limits. There is obviously a limit to the indium concentration of the channel, but there are other problems, which must be addressed before this limit reached. As the indium concentration is increased the channel layer thickness must be reduced to avoid dislocations due to lattice mismatch. High indium concentrations can be achieved using InP substrates but this is not ideal, they are more fragile than GaAs and are not available in large wafer sizes. There is a limit to how much the parasitic resistances can be reduced, after this the intrinsic performance must be improved to provide enhanced performance. The remaining method for the improvement of RF performance is the reduction of gate length. It was seen in section 2.3.2 that the intrinsic transconductance and hence f_t and f_{max} is strongly dependent on the gate length. The reduction in gate length improves the RF performance of a device in other ways, it is seen in equation 2.48 that f_t is inversely proportionally to the capacitances C_{gd} and C_{gs}. These arise from the capacitance between the gate electrode and the channel, a reduction of the gate length will reduce the area of the capacitor, reducing the gate capacitance and therefore improving f_t and f_{max}. As the gate length is reduced the length of the channel becomes comparable with the mean free path of the electrons. This means that electrons have a much better chance of passing through the gate region without scattering, leading to higher electron velocity in the channel. This was predicted by Kalna et al. [6.1] for the case of scaled GaAs pHEMTs. This work used Monte Carlo simulations of fully scaled GaAs pHEMT devices with a gate length ranging from 30 to 120nm. It was the found than the peak electron velocity increased as the gate length decreased, resulting in an increase in the intrinsic transconductance of the devices. The advantages of gate length scaling have also been demonstrated experimentally. Devices with a gate length of 25nm currently hold the record f_t of 562GHz for an InP pHEMT with an $In_{0,8}Ga_{0.2}As$ channel. This

performance enhancement motivated the development of the sub-100nm T-gate technology described in this thesis.

It was decided to develop two separate processes for the fabrication of ultra short gate length devices. One process would concentrate on the fabrication of sub-100nm gate length devices, optimised for flexibility and ease of fabrication. The second process would be designed to fabricate gates with as small a gate length as possible. It was decided to introduce this two pronged approach as it was realised that to produce sub-50nm devices extra fabrication steps would be needed that were not needed for gate lengths greater than 50nm.

6.2 Development of Sub-100nm Gate Process using Co-Polymer

At the start of the development of the sub-100nm gate length process there was already a 70nm technology in use within the group, based on a relatively complex process using PMMA and UVIII resists separated by a thin layer of aluminium [6.2]. UVIII is a chemically amplified resist, and is light sensitive so care needs to be taken during processing to avoid white light. The process also requires a Post Exposure Bake (PEB) to activate the UVIII resist and the time delay between the PEB and development can alter the resist profile.[6.3] The use of the aluminium between the two resists was included to prevent cross-linking at the interface; aluminium was chosen because it is removed by the developer of UVIII. Aluminium however is also etched very quickly in HF so it was not compatible with the post etch clean described in section 5.2.2.

In comparison with the UVIII process, the P(MMA-MAA) process used for 120nm T-gates is relatively simple, requiring three layers of resists, with an overnight bake. There is no PEB and so no limitation on the time between the exposure and development. It would be ideal if a process was developed that had all the benefits of the co-polymer process and was able to produce gate lengths below 100nm.

6.2.1 Introduction to 120nm co-polymer process

The basis for the development of a sub-100nm process using a P(MMA-MAA) resist stack was the standard 120nm co-polymer process that had been used at Glasgow University for a number of years. This process was repeatable and therefore provided an ideal starting point for development of a new gate process.

The process uses a tri-layer resist stack, the base layer is 4% ELV2041 PMMA and typically has a thickness of 110nm. The next layer of the resist is a layer of 8% co-polymers of Methyl-methacrylate and Methacrylic Acid P(MMA/MAA) with a thickness of around 300nm. The final layer is a thin layer of ELV2010 PMMA, typically 100nm thick. The co-polymer and 2010PMMA layers are much more sensitive than the 2041PMMA. This difference in sensitivity allows the T-shaped profile to be produced when the correct dose pattern is used.

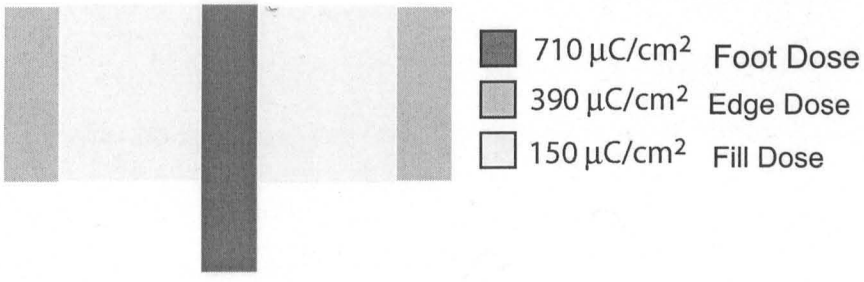


Figure 6.1 Illustration of the dose pattern used for the realisation of 120nm co-polymer gate.

The dose pattern used is shown in Figure 6.1. Three doses are used, labeled the foot, edge and fill doses. The foot dose produces the gate footprint with a designed width of 40nm, however due to lithographic issues such as the proximity effect and forward scattering of the beam the developed size is 120nm. The foot dose is much higher than the other doses, typically 710μC/cm^{-2} is needed to develop out the relatively insensitive ELV2041 layer. The edges of the gate are defined using two 40nm wide rectangles at the outer edges of the gate, with a dose of 390μC/cm^{-2} -slightly higher than the dose that would normal be used to clear out 2010PMMA because of the relatively small area and a reduced contribution from the proximity effect at the outer edges. The dose is chosen so that the pattern is defined in the co-poly and 2010PMMA layers but is not high enough to expose the 2041PMMA layer. The fill region is comprised of

two rectangles either side of the gate foot, each with a width of 150nm. The fill region uses a dose of 150μC/cm^{-2} to ensure the shoulder of the T-shaped profile is well defined. The development process is as follows; the sample is developed in MIBK:IPA in a ratio of 2.5:1 for 30 seconds rinsed in IPA then blown dry with N_2. The temperature of the developer was 23±0.5°C.

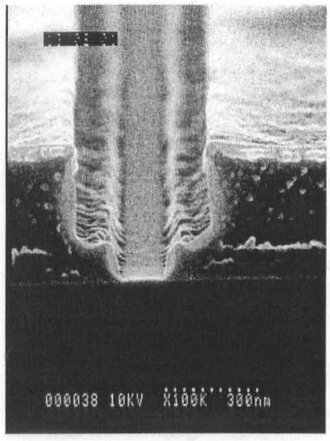


Figure 6.2 SEM cross section of 120nm T-shaped resist profile in P(MMA-MAA) Tri-layer.

The resultant resist profile is shown in Figure 6.2. The sample has been tilted slightly enabling the trench that defines the gate foot to be clearly seen. The total resist thickness is around 500nm.

6.2.2 *Modifications to 120nm Process*

In order to realise sub-70nm gate length structures, the 120nm T-gate process was modified to reduce the overall thickness of the resist stack. As shown in section 4.3.2, the electrons experience forward scattering as they pass through the resist. Therefore to achieve smaller feature sizes it is important to reduce the forward scattering of the electrons by thinning the overall resist stack. The thickness of the bottom layer of resist that defines the foot influences the development time of the sample, as the developer has to be able to reach the bottom of the resist profile. If the resist trench is too deep and too narrow the diffusion process of the developer may not be enough to fully remove the exposed resist. Initially the 120nm T-gate resist stack was used for sub-100nm T-gate development, the only difference being that the resist was spun on a planar substrate as opposed to being between a source drain gap as it had

been observed that the resist stack within a source-drain gap is approximately 25% thicker than on planar substrate. Using this resist structure it was possible to produce 70nm T-gate structures on a planar substrate, as will be discussed in section 6.2.4.

The gates of standard devices of course are not written on a planar substrate, they are written between source and drain ohmic contacts. To achieve similar lithography the resist thickness within the source-drain gap must be the same as on the planar substrate. This was achieved by reducing the concentrations of the resists used. The resist thickness can be estimated by the following equation [6.4] where K is a constant that depends on the resist used.

$$Thickness = \frac{K(\%concentration)^2}{\sqrt{SpinSpeed}}$$

The concentration of the 2041 PMMA layer was reduced from 4% to 2.5%, however this made the resist slightly too thin within a source-drain gap so the spin speed was reduced from 5000rpm to 3000rpm. It is seen from the above equation that the resist thickness is much less dependent on the spin speed than concentration so it can be used for fine tuning. Similarly the co-polymer concentration was reduced from 8.5% to 7% and the spin speed reduced from 5000 to 3000rpm. The 2010 PMMA layer concentration was changed from 2.5% to 1.5% with no change in spin speed. This was done because it was observed that the resist thickness increase in the source drain gap was mainly caused by this layer.

The development process was identical to the 120nm process as it worked and did not require modification.

In addition to the reduction of resist thickness for the sub-100nm process, the lithography was carried out using a electron energy of 100keV. As discussed in section 4.3.2 this was to reduce the effect of both forward and backscattering. As is shown in Figure 6.3, while a T-gate structure with a footwidth around 70nm can be achieved using 50kV accelerating voltage, the

profile is not ideal. It was concluded that the 50keV process would be harder to scale to smaller feature sizes, therefore all subsequent work was carried out using 100kV.

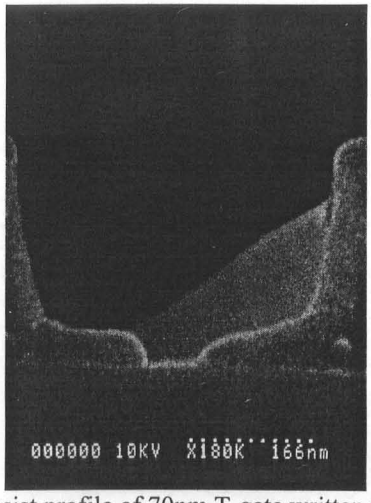


Figure 6.3 SEM Image of Resist profile of 70nm T-gate written using 50keV.

6.2.3 *Dose Strategy*

The main issue with the development of the co-polymer process is the relative dose required by each area of the gate. This is because the sensitivity ratio of the PMMA/Co-Polymer resist stack is low, with a ratio of 2.7 compared to 12.5 for the PMMA/UVIII system [6.5]. The small sensitivity ratio means that the different dose areas cannot be considered as independent; they will influence each other due to a combination of the proximity effect and forward scattering.

The first consideration is the foot of the gate. To produce a narrow gate foot with a well defined shoulder, the distribution of the exposure energy in this region must be large and narrow. To have more control of the gate foot length the contribution to the foot dose from the other dose areas such as the fill and edge must be small as the dose of the surrounding areas will cause the overall energy distribution of the gate foot to widen increasing the gate footprint size.

The first of the design modifications was to narrow the high dose foot geometry. In the 120nm process this had a designed width of 40nm resulting in a feature size of 120nm post development. In the modified design this was reduced to 10nm in order to narrow the dose

profile as much as possible. To reduce this width any further, the resolution of the pattern would have to be increased so the block size would be reduced leading to potential stitch errors along the gate. Besides, the smallest spot size available on the Lieca EBPG5 is 12nm so a smaller design would not produce a smaller line.

The various doses are applied to the different areas in the following way; the gates are designed in a CAD software package such as Wavemaker with each different dose region designed on a different layer. The pattern is then fractured by CATS. During the fractionation, the different layers are each assigned a dose multiplier, these multipliers are defined in a simple text file called a CCFA file. An example of a simple CCFA file is given below

```
sort by ccfa
3     ! Dose 0, Dose of 300% for layer 6 (fill)
5.3   ! Dose 1, Dose of 530% for layer 7(edge)
15    ! Dose 2, Dose of 1500% for layer 18 (Central Line)
layer 6
0
layer 7
1
layer 18
2
```

The first section of the file lists the dose multipliers that will be used, in this case for a 70nm T-gate the multipliers are 3, 5.3 and 15. The second part of the file assigns each of the multipliers to a particular layer. The dose that each level will receive will be the base dose defined in BELLE multiplied by the factor set by the CCFA file.

6.2.4 Dose Testing

The dose to produce the desired resist profile is found by performing a dose test. The test pattern is repeated a number of times. The first cell is written at the lowest dose, the dose of each subsequent cell is multiplied by a factor, k. The dose of each cell is given by

$$ith\_Dose = LowestDose \times (k)^i$$

Where i is the number of the cell. To cover a particular dose range it is better to have a larger number of cells, this allows the increment between doses to be smaller. It is possible for most dose tests to be examined from above with either an optical microscope or a SEM depending on the size of the pattern. This allows a large number of doses to be quickly examined. However for the development of a T-gate process it is not just the line width that is of interest but also the profile of the developed resist. To do this the sample needs to be cross-sectioned and viewed in a high resolution microscope such as the Hitachi S900 SEM. This microscope was used because it has the highest resolution of all the SEMs in the department, due to a TEM style stage, that allows a very small working distance between the lens and sample. The disadvantage of this stage is the small sample size that it requires. The maximum sample size is approximately 5mm long and 2.5mm wide. Since the size of the gates are very small, (300nm wide in 500nm thick resist) it requires a high magnification before the gates can be seen, this high magnification makes navigation between different doses much more difficult than for a planar view and means that relatively few doses can be observed (typically six doses). The exact dose is found by an iterative process each time using a smaller dose range.

A challenge particular to the development of the co-polymer gate process is finding the correct multiplying factors for the different dose regions. At one base dose the foot may be developed out but the edge region is over exposed. An example of this is shown in Figure 6.4, at a base dose of 150 the foot is well defined, however the fill region is under exposed. When the fill region is correctly exposed the edge region is over exposed.

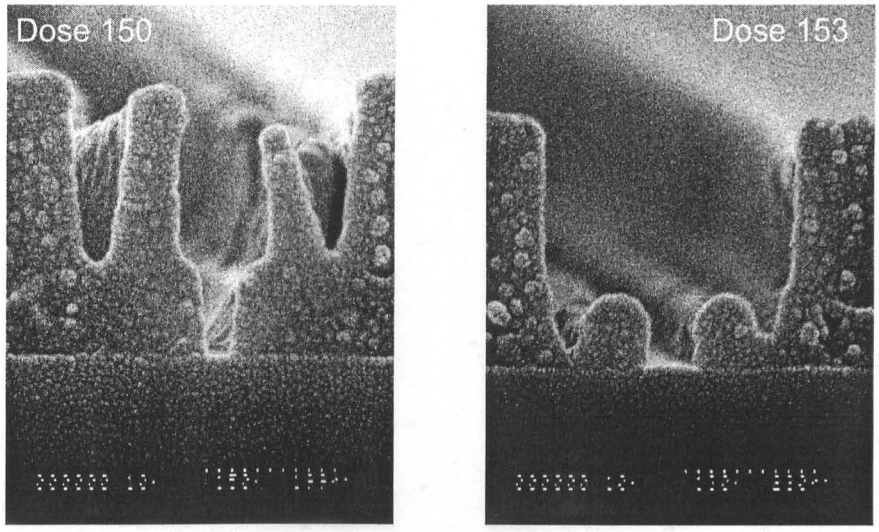


Figure 6.4 Sample of a dose test. The ccfa multipliers are not correct, this results in an underexposed fill region when the foot and edge is developed (Dose 150) and an over exposed edge region when the fill is developed out (Dose153)

All the doses need to be studied and the best base dose for each area found. A new CCFA is then written that produces the correct dose for each layer at a particular base dose. This CCFA file is then tested because, as was seen previously, the different regions are interdependent. The process is repeated until the best ratio is found.

Using the technique described above, the base dose and the relative dose ratios for each region were found that would produce a T-shaped gate with a gate length of 70nm. Using a base dose of 116, was found to give a gate foot length of 70nm. The relative doses were 150% for the fill layer, 350% for the edge dose and 1500% for the central line. In order that the foot length of the gate did not change during inspection due to deformation of the 2041 PMMA layer a thin layer of metal, (15nm Ti, 30nm Au) was deposited after development. Figure 6.5 shows the gate profile for a base dose of $116\mu C/cm^{-2}$. It was seen that smaller foot lengths were achieved at lower doses but it was not clear if the profile was continuous along the width of the gate.

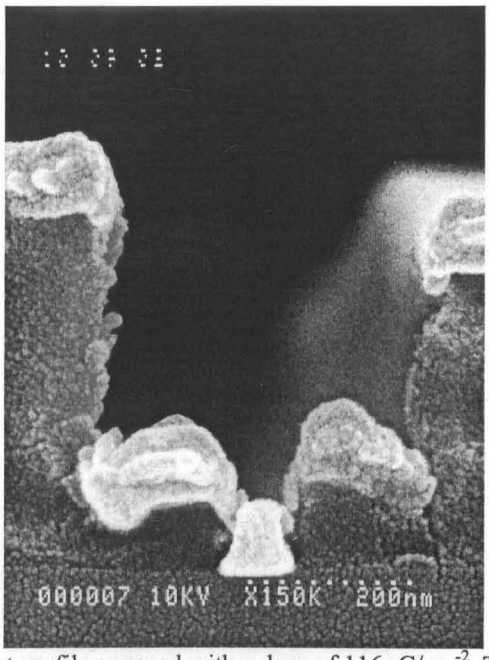


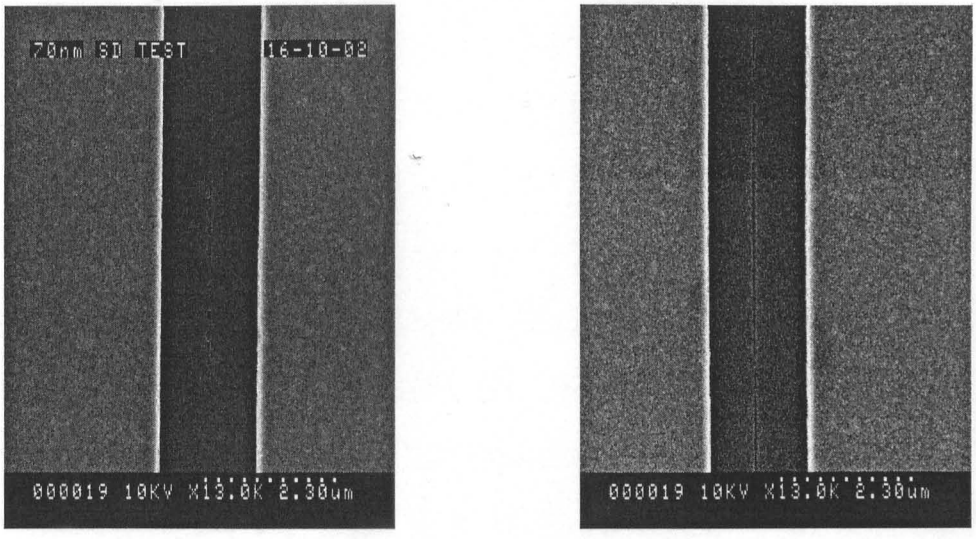
Figure 6.5 SEM of resist profile exposed with a dose of 116μC/cm^{-2}. The sample has been metallised with a thin layer of metal to protect the resist profile. The foot width is 70nm.

A further test was performed to assess process validity. It was not possible from the cross section to determine if all the PMMA had been removed, even when the sample was tilted. As it had been reported that the development of PMMA can leave a very thin (<1nm) residue on the surface [6.6]. In order to see if all the PMMA had been removed, an additional step was included. After development and ashing, the sample was wet etched using either succinic acid based selective etch or ammonia/hydrogen-peroxide non-selective wet etch. Ideally the selective etch would be used as this is used in the fabrication of real devices. However the preparation of this etch is time consuming and requires epi-layer material to achieve the selectivity. By non-selective etching for 15s the correct etch dimensions were achieved. In this way, inspecting the etch trench continuity enabled an unambiguous determination of whether any resist residues remained after development.

6.2.5 Continuity Tests within a Source-Drain Gap

It was now possible to produce a 70nm T-gate using a PMMA/Co-polymer tri-layer on a planar substrate. However for the fabrication of most HEMTs the gate is written between source and drain ohmic contacts. The thickness of resists was modified to account for the effects of the

source-drain gap as detailed in section 6.2.2. A dose test was performed within the source-drain gap, designed to find the smallest dose which would provide a continuous gate. The sample was etched using a ammonia, hydrogen peroxide and water non selective etch for 15 seconds as described above, the resist was then removed and the sample was viewed in the S900 SEM. Figure 6.6 shows the planar view of two different gate stripes, one, was written with a base dose of 112μC/cm^{-2}, the other with a dose of 116μC/cm^{-2}. It is seen that the recess trench formed using the gate pattern written with the lower dose is not continuous. Increasing the dose increases the continuity of the gate recess trench. The base dose was increased marginally to 118μC/cm^{-2} envisaged to be sufficient to provide a continuous gate.



(a) (b)

Figure 6.6 Planar view of etched gate within a source-drain gap (a) Sample written with a dose of 112, there are large gaps between the etched areas (b) Dose 116, the gate is nearly complete along the length of the gate, however there are some discontinuities.

The next stage of development was to transfer the process from small test pieces onto real device samples. The devices were to be fabricated on wafer A1602, a lattice matched HEMT structure grown on InP. The full layer structure is given in Appendix C, the thickness of the barrier is 10nm and the spacer thickness is 4nm giving a total gate to channel separation of 14nm. To avoid short channel effects the gate to channel separation must be less then a third of the gate length. The 2DEG however does not form at the interface but rather ~5nm into the well,

increasing the effective gate channel separation further. This means that the material is well scaled for 70nm gate length devices, with a gate-2DEG separation of 19nm.

6.2 Fabrication of 70nm HEMTs on wafer A1602

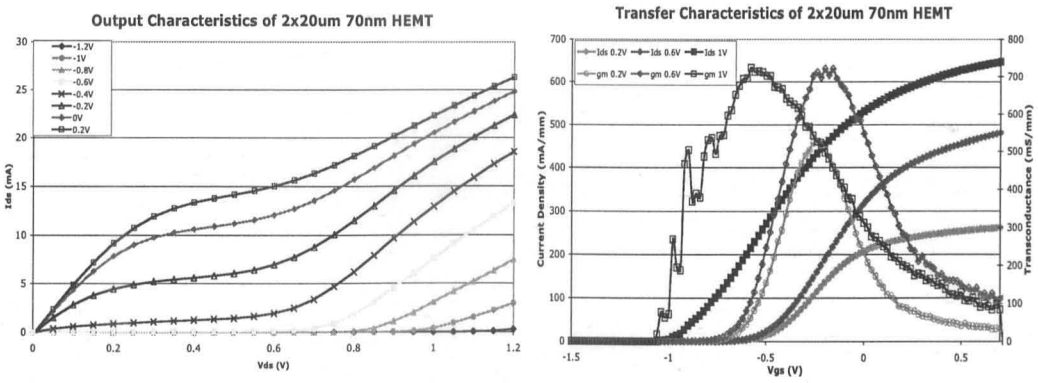
The fabrication of devices on A1602 followed the standard InP process flow given in Appendix D with some modifications. The marker layer and ohmic contacts were combined in the first level of lithography, the two patterns were fractured separately, the ohmic contacts were fractured using proximity correction to provide better defined source-drain gaps. The pattern was metallised using a Ni:Ge:Au metallisation with thicknesses of 10:50:80nm. This gave a total thickness of 140nm. This is a standard metallisation but still relatively thin, well suited to writing gates within the source-drain gap. The samples were annealed at 240°C for 30s then 280°C for 60 seconds. The surface morphology was inspected and found to be good. The next stage was the isolation level, and because the ohmic contacts were formed it was possible to monitor the isolation etch electrically. The sample was etched for a total of 50 seconds in 1:1:100ml orthophosphoric acid :hydrogen-peroxide :water. The current was measured between the isolation test structure, starting at 39mA and falling to 66nA. This was considered to be sufficiently isolated, the resist was removed and the etch depth measured using AFM, was found to be 44nm. From the layer structure it is seen that this is half way through the channel. The ohmic contact resistance was measured and was found to be 0.05Ω.mm using the TLM method. The unetched mobility and sheet concentration was measured using the Van de Pauw method, and found to be 5020cm^2/v-s and 1.21x$10^{13}$$cm^{-3}$ respectively as expected for lattice matched InP material. The next stage was the gate level. The resist stack that was used is given in Table 6.1

Resist	Concentration	Spin Speed
2041 PMMA	2.5%	3000rpm
Co-Polymer	7.5%	3000rpm
2010 PMMA	1.5%	5000rpm

Table 6.1 Resist Stack used for 70nm T-gates within a Source-Drain Gap

The gates were written, developed, then ashed for 40 seconds to remove any remaining resist from the gate foot. The gate recess was formed using the succinic acid selective etch for 45 seconds, as it was found this resulted in a recess offset of around 25nm in this material. The devices were rinsed in 10:1 HF for 30s immediately before being placed in the Plassys for metallisation, the metal used was Ti:Pd:Au 15nm:15nm:200nm. Lift-off was performed and the devices were completed with the addition of 400nm thick CPW bondpads.

The devices were then characterised using the equipment and methods described previously, the yield of the devices was quite high with around 60% of the devices working at DC, very encouraging for a first time process and shows the strength of the gate technology. The average threshold voltage was –1.1V with a standard deviation of 0.2V. This is a rather low threshold voltage. This is due to a slow breakdown behaviour, which caused the threshold voltage to shift as the drain bias is increased. The uniformity of the etch was good. However there was some notable exceptions and a number of devices displayed a higher threshold voltage (typically –0.6V) and also displayed the highest value for the transconductance of 1000mS/mm. The average value however was 675mS/mm, rather low for lattice matched devices, particularly with a gate length of 70nm. Figure 6.7 shows the DC characteristics of a typical 2x20μm device and with a pinch off voltage of –1.1V.



(a) (b)

Figure 6.7 DC characteristics of 2x20μm 70nm HEMT fabricated on A1602. (a) Output Characterstics of device measured with range of V_{ds} of 0-1.2V and V_{gs} from 0.2 to –1.4V. (b) Transfer Characteristics showing I_{ds} and g_m for V_{ds} = 0.2,0.6 and 1V.

It is seen from the low field region of the output characteristic that the series resistance is low, it is also seen that breakdown starts at a low drain bias, at around 0.5V for V_{gs}=-0.4V. The transfer characteristic shows that the threshold voltage of this device is –0.55V for a drain bias of 0.6V but shifts to –1.1V at a drain bias of 1.2V. This behaviour is caused by the early breakdown of the device, due to a combination of the material system, $In_{0.52}Ga_{0.48}As/In_{0.53}Al_{0.47}As$ which is known to display a lower breakdown voltage than GaAs pHEMTs. The length of the gate recess also plays an important role in the breakdown voltage of a device. The shorter the gate recess etch the lower the breakdown. It is therefore likely that the gate recess of these devices is small.

The devices were characterised between 0.04-60GHz using the VNA described previously. Using the collected S-parameters a small signal model was fitted as described in section 5.5. Figure 6.8 shows both the modelled and measured S-parameters, it is seen that the fit of the model is good from a frequency range of 20 to 60GHz.

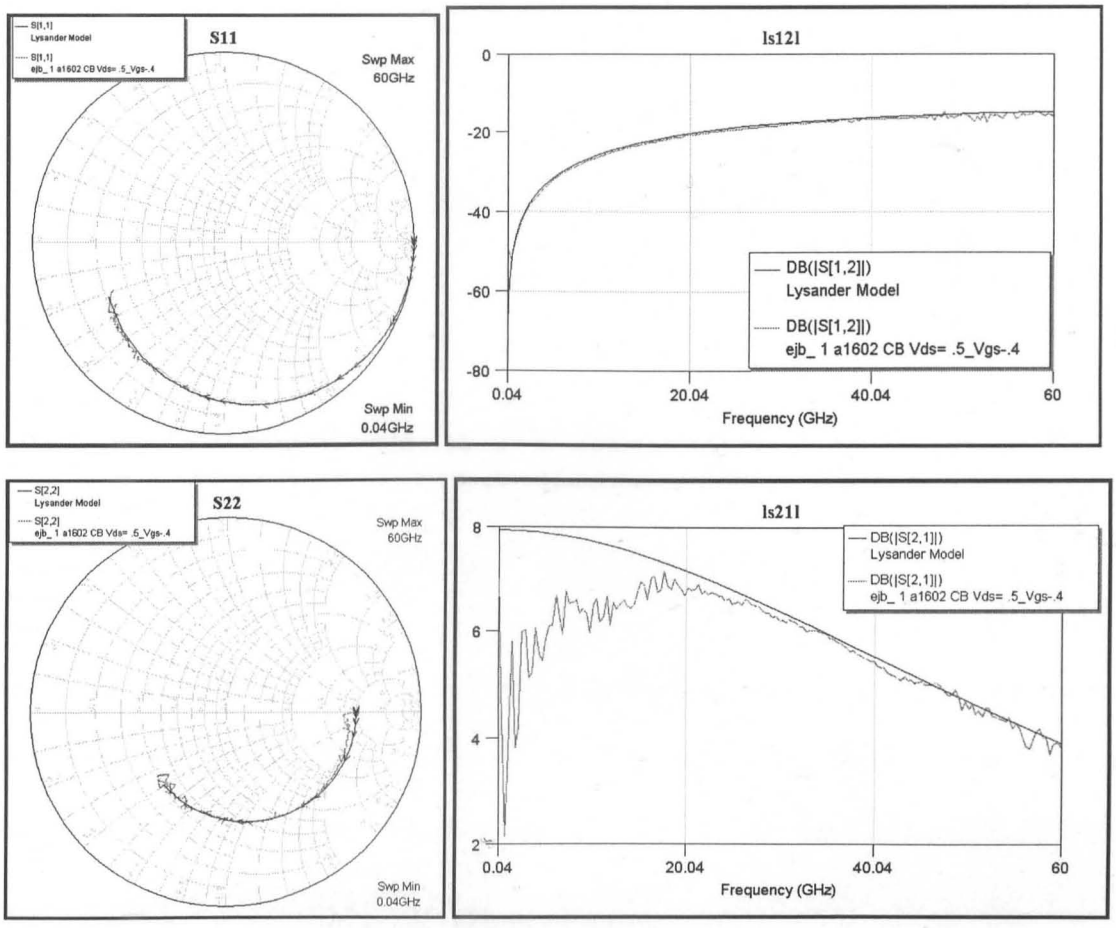


Figure 6.8 2x20μm 70nm HEMT Lumped Element Model versus measured s-parameters between 0.04 and 60 GHz.

The effects of the CPW lines were then removed from this model to give the de-embedded model, Figure 6.9 shows the resultant short circuit current gain, h21, and the Maximum Available Gain, MAG. The extrapolated cut-off frequency was around 210GHz and a maximum frequency of oscillation , f_{max}, of 200GHz.. This is slightly disappointing for the type of device; for a 70nm device on lattice matched InP, a figure of 250GHz for f_t is expected. This was in part because it was necessary to use a bias point of V_{ds}= 0.5V V_{gs} = -0.2V, to avoid the breakdown region.

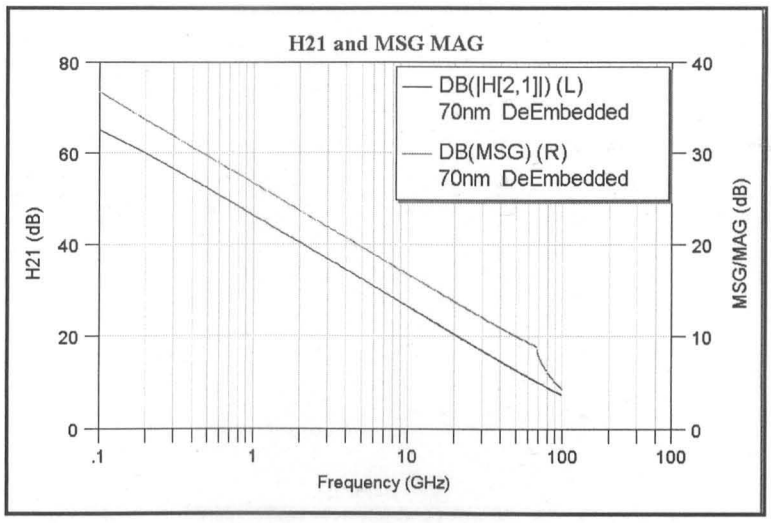


Figure 6.9 H21 and MAG/MSG for the de-embedded model of A1602 70nm lattice matched pHEMT.

To obtain more information on the devices, particularly in the gate region, the devices were inspected in the S900 SEM. Preparing device samples for cross sectional analysis is not straightforward. The device cell is designed with lithography issues in mind. As discussed in section 4.2.3, the finite size of blocks that can be written in the beamwriter and is reduced as the resolution becomes smaller. To avoid stitching problems the cell is designed so that the gate are within the block boundary at a resolution of 0.01 which has an area of 160x160μm. The width of the devices on this sample were mostly 2x20μm, which improves the yield as none of the gates cross a block boundary and so do not suffer from stitching problems. This combined with the low density of the gates means that to view a cross section of the T-gate the accuracy the cleaving needs to be very high. New methods of analysis of active devices will be considered in Chapter 7. Despite these difficulties the image shown in Figure 6.10 was obtained.

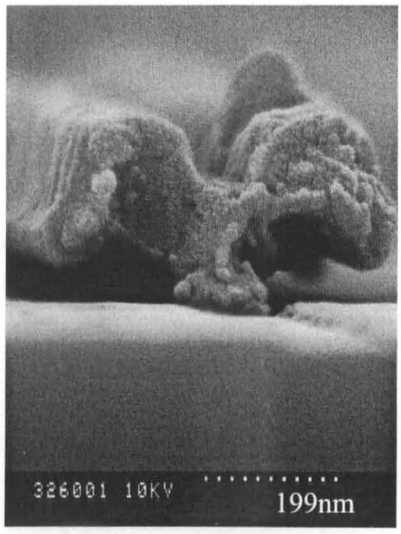


Figure 6.10 SEM cross section of "70nm" device fabricated on A1602

It is seen from Figure 6.10 that the gate lithography of the device was poor. Both the edge and foot region are over exposed. This leads to a larger foot width, around 90nm rather than the expected 70nm. This is the reason that the performance of the device was poor compared to other 70nm devices. The edge region has also been overexposed, bringing the edges of the T-gate very close to the substrate, this will have the effect of increasing the parasitic gate capacitance. The recess offset of the device was, as suspected from the breakdown behaviour of the device was very small, with a length of around 10nm.

Work was carried out to investigate the cause of the dramatic dose change in what had been until now a stable and repeatable gate process. When the beamwriter writes a pattern it produces a log file that contains information of the job. This log file includes information of the alignment of the sample, the writing frequencies that were used, from these the actual doses written can be found and most importantly in this case information, on the height of the sample. It was found that the sample had not been completely flat in the holder, it had a slant of -1μm/mm, this is a very shallow slope but as discussed in the next section this can have a great effect on the resultant resist profile.

6.4 Central Alignment Marker

The measured sample height is very important when writing small features due to the way in which the beam is focussed. The beam is initially focussed on a metal on GaAs marker on the sample holder. However, since many different thicknesses of substrate are used, this height will not be the same as the sample height. This is corrected by measuring the height difference using the laser height monitor that was shown in Figure 4.5. The beam is then focussed on the sample surface using the measured height difference and a look up table that gives the change in focus for a particular height difference. This method normally works well. However there are some issues that can affect the focus on the surface of the sample. The change of focus depends purely of the measured height of the sample. There are a number of factors that can influence the measured height of the sample. As the height is measured using a laser, the material that the laser is aimed at can alter the height measured, so if the laser is aimed at a metallised region the height measured will be incorrect, a number of values are taken at different positions on the wafer to reduce this uncertainty. The slope of the sample will also affect the way that the laser is reflected and so introduces errors in the height measurement. A slope of 1μm/mm of the sample will introduce an error of around 1μm in the measured height and a corresponding increase in the spot size of around 6nm. The defocusing effect of an error in the measured height is illustrated in Figure 6.11

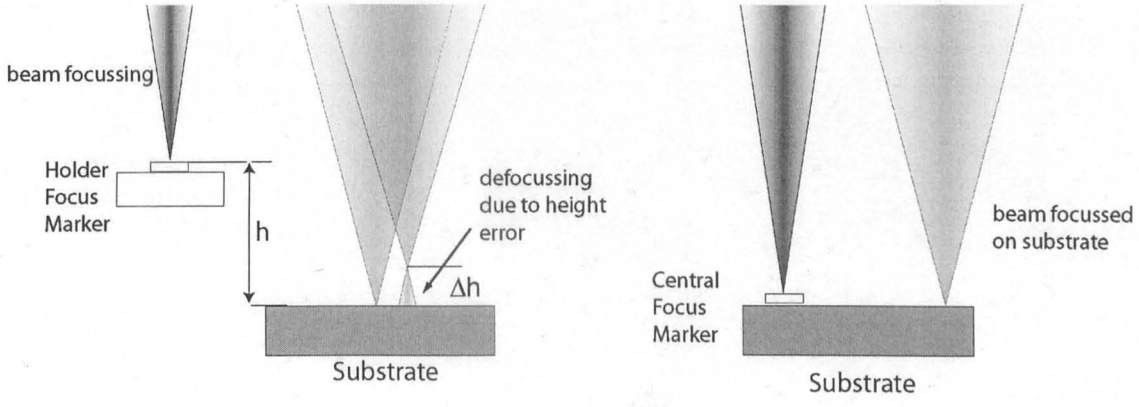


Figure 6.11 Illustration of the focussing problems that can arise from incorrect height measurements introduced by the tilt of the sample.

The problem was solved by Stephen Thoms who developed a method that allows the beamwriter to focus the beam on the sample rather than holder and therefore removes the dependence on the accurate measurement of the height of the sample. This is carried out in the following manner; during the marker level a central focus marker is written and metallised on the sample. The design of this marker is shown in Figure 6.12, it consists of a 20x20μm square in the centre of a region of material greater than 1x1mm. (This large, empty area allows the beam writer software to find the marker automatically). The beamwriter is then instructed to use this marker to focus the beam rather than the marker on the holder. By using this method the beam is already focused on the sample and doesn't need to be refocused which can introduce errors due to incorrect height measurements.

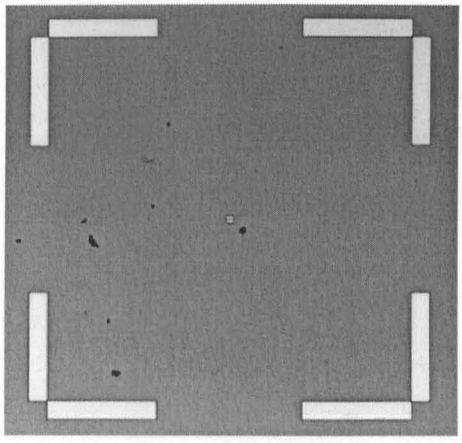


Figure 6.12 Optical Micrograph of Central Alignment Marker consisting of a 20x20μm metallised square within a 1x1mm square.

This method was tested and it was found the dose required to expose a 70nm gate foot was around 10% smaller than without the on-wafer calibration. This is due to the fact the focus of the beam on the sample surface is better, thus the energy distribution is narrower. This means that more of the exposure energy is located in the pattern region and less in the surrounding resist so a lower dose is needed to develop out the pattern. The narrow beam distribution also enables smaller features to be developed out with greater consistency.

6.5 Development of 50nm Co-polymer Gate Process

During this period a great deal of work was being carried out around the world aimed at increasing device performance by gate length scaling. Devices with sub-100nm gate lengths had already been demonstrated such as the 70nm devices fabricated at Glasgow using a UVIII/LOR/PMMA [6.7], 60nm metamorphic devices fabricated at Lille [6.8], and 25nm devices fabricated by Fujitsu [6.9].It was therefore decided that instead of fine tuning the 70nm gate length T-gate process using co-polymer to improve the gate profile, work would be carried out to investigate the lower limits achievable with the PMMA-Copolymer resist structure.

It had been seen on a number of occasions during dose tests for the 70nm process that it was possible to produce resist profiles with a gate length of 50nm. However these profiles were not repeatable and were not continuous along the width of the gate. If it was going to be possible to produce 50nm T gates the process would need to be modified further.

The modifications that were introduced to move from 70nm to 50nm were the same as used to scale from 120nm to 70nm. The thickness of the resist stack was reduced further, the spin speed of the base PMMA layer on a planar substrate was increased from 3000rpm to 5000rpm, this reduced the thickness from 80 to 65nm. The speed at which the co-polymer layer was spun was also increased from 3000rpm to 5000rpm. This was done for two reasons; the first is for improved lithography the thinner resist stack reduces the forward scattering slightly, as well as aiding development of the gate. The second reason is to improve the metallisation of the gate. The gate recess is performed before the gate metallisation and typically has a depth of 20 or 30nm depending on the epilayer used. The gate must be filled up from the etched surface to the top of the 2041 PMMA for the T-shape to be formed. As it is deposited, metal will stick to the sidewalls of the resist. This narrows the trench that defines the stalk of the gate. If the distance from the top of the PMMA layer to the substrate is much larger than the width of the trench, the gate stalk will close over and the gate foot will not be connected to the head of the gate. An

example of this is shown in Figure 6.13 this shows a 50nm T-gate that has been metallised with 15:15:160nm Ti:Pt:Au.

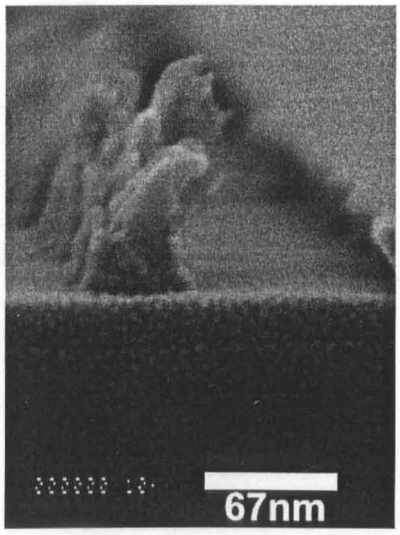


Figure 6.13 SEM Image Metallised 50nm T-gate structure where the thinning of the stalk lead to the gate lead separating from the foot.

It can be seen that the gate foot is well defined with a length of around 50nm, the metal narrows to form a point forming a pyramid structure. Either the gate head was never connected or the metal was too thin to withstand lift-off. By thinning down the 2041 PMMA layer that defines the gate stalk this effect is minimised.

A slightly different method was used to find the relative doses that would give the best T-gate process with the smallest foot width. Previously the relative ratios set by the ccfa file were constant and the base dose was varied, however since the correct doses for the edge and fill had been found for the 70nm gate, these were kept constant and the central dose was varied by writing each central dose as a separate layer. This method allowed the dose for the central line dose to by found quickly without the need for multiple iterations. An array of devices were written with this method using a dose range of 1350 to 1440 $\mu C/cm^{-2}$, a succinic acid based gate recess etch and an HF post etch clean was performed before the samples were metallised using Ti:Pt:Au 15:15:160nm and the gate lifted off. This test was designed to be as an accurate representation of device fabrication as possible. It not only gave information on the resist

profile, but the metallisation effects and the effect of the gate recess. The mechanical stability of the gates is also seen.

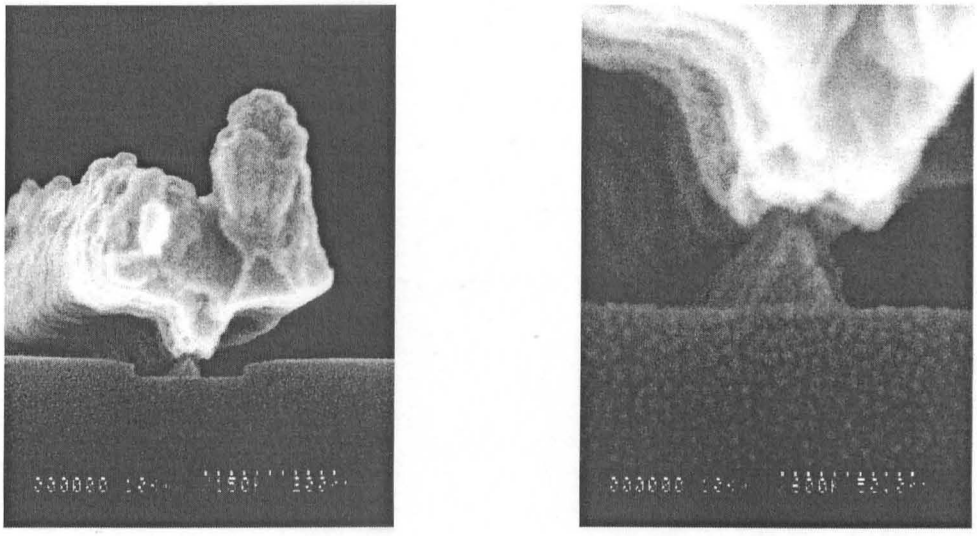


Figure 6.14 SEM cross section of metalised T-gate structure, exposure dose of 1350 for central line. (a) Magnification of x150k, showing the profile of the whole gate and recess etch. (b) Close-up of the foot region, magnification of x600k, the foot width is around 50nm.

It was found that a dose of 1350μC/cm^{-2} for the central line gives a metallised T-gate with a foot length of 50nm. Figure 6.14(a) shows the cross section of the gate with a magnification of x150k, this allows us to see the profile of the whole gate and the recess etch. In order to see the gate foot better Figure 6.14(b) show this region with a magnification of x600k. It is seen that the gate length is around the same size as the scale bar which reads 50nm. There is however some uncertainty over the exact dimensions of the device due the difficulty of focussing at this high magnification.

The process was transferred from the planar substrate on which it had been developed to within a source drain gap as in a device. This was achieved by increasing the spin speed in order to keep the resist thickness the same as that of the planar sample. To this effect the spin speed was increased from 5000 to 6500rpm for each layer. A cross section dose test was carried out using the central focus marker, it was seen that using the same dose as for the planar that the foot width increases to 55nm. The dose was therefore reduced slightly to achieve 50nm.

6.6 Fabrication of 50nm T-gate HEMTs

Devices were fabricated using this process on wafer A1881, a metamorphic GaAs wafer, the layer structure of which is given in Appendix E. The design of the layer structure is similar to that of a lattice matched HEMT on InP. This consists of a 20nm $In_{0.53}$ $Ga_{0.47}As$ cap with a doping concentration of $1x10^{19}cm^{-3}$, the barrier layer with a thickness of 8nm, a layer of delta-doping with a concentration of $6x10^{-12}cm^{-2}$, and a spacer layer with a thickness 4nm above a 20nm channel. It is seen that the gate-channel separation is only 12nm making the material suitable for 50nm gate length devices.

The devices were fabricated using the standard InP HEMT fabrication process used for the 70nm devices fabricated on A1602 up to the gate level. The gates were written using the 50nm process described above. The devices were completed by the evaporation of 400nm CPW waveguide feed-lines.

The devices were inspected by optical microscope, the yield was around 50% from inspection. However when the devices were characterised at DC is was seen that none of the devices showed gate control. To investigate this the samples were inspected in the Hitachi S4700 SEM, which has slightly lower resolution than the S900 but has the advantage of being able to view much larger samples. However, because of the T- shaped gate is was not possible to view the gate foot and etch for most devices. It was however possible to study the gates that had been damaged during lift off. Figure 6.15 shows the source-drain gap of one of these devices. In the centre of the image the gate recess is visible, it is seen that this is quite wide (around 200nm), within the recess the gate foot visible. The gate foot is not complete; it is formed by a number of separated dots of metal. This is why there was no gate control of the devices. The fact that the recess was continuous suggests that the sample was close to being developed out properly.

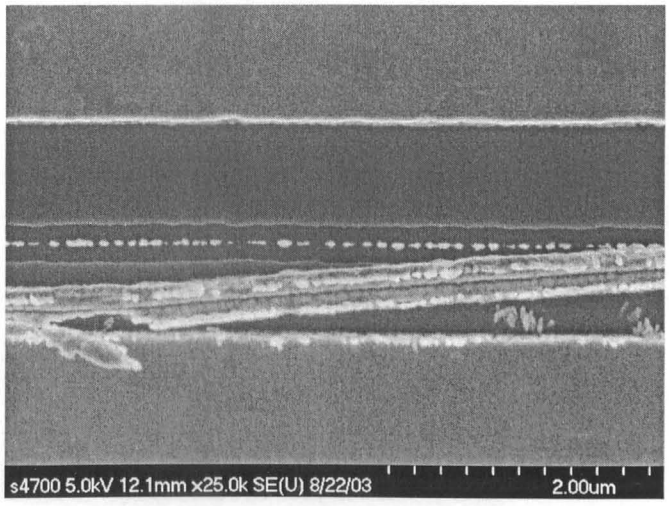


Figure 6.15 SEM image of failed 50nm gate written within a source-drain gap. The recess has formed but the gate foot is incomplete.

It was suspected that the reason for the poor lithography within the source-drain gap was caused by the resist thickness within the source-drain gap rather than a dose shift because other short gate length processes showed no change over this period.

The resist thickness within the source-drain gap was studied using AFM in the following way. The initial layer of the resist 2.5% 2041 PMMA spun at 6000rpm on real device sample. Several windows were opened up across the source drain gap. The area was then scanned using the AFM and from this, the thickness of resist within the gap could be found together with the thickness uniformity across the gap. A number of sites across the wafer were scanned an average thickness of resist was found to be 67nm, this is slightly thinner than the same resist spun at 5000rpm on a planar sample. The resist was stripped and the sample re-spun with a Bi-layer of 2.5% 2041PMMA spun at 5000rpm and 7.5% Copolymer spun at 5000rpm. Using the above technique, thickness was found to be 310nm, this is 20% higher than the same resist on a planar sample. The sample was then cleaned and spun with the tri-layer of 2.5%2041@5000rpm, 7.5%Co-polymer@5000rpm and 1.5% 2010 @ 6000rpm. The overall thickness was found to be 550nm compared with 460nm on a planar substrate. It is seen that the majority of the thickness difference come from the upper layer. This layer was originally introduced to the 120nm T-gate process to give an undercut profile to aid lift-off as described in

section 4.3.1. However because of the effects of metal evaporation that produces the pyramid gate shown in Figure 6.13 the gate head narrows as more metal is deposited. This narrowing gives the effect of an undercut if the thickness of the co-polymer layer is greater than the metal thickness. The total metal thickness used is 210nm compared to the thickness of the co-polymer layer 250nm so it is possible to remove the 2010 layer completely if the spin speed of the co-polymer layer is reduced slightly to give a marginally thicker layer.

Devices were fabricated using this bi-layer resist stack on a new wafer B888, this is also a metamorphic wafer with a similar layer structure to A1881 making it suitable for devices with a gate length of 50nm. The devices were written with a central dose of 1360μC/cm^{-2}. The devices were etched for 20 seconds using the succinic acid based gate etch with a 30 seconds HF dip prior to metallisation. The physical yield after lift-off was better than 80% and the devices were characterised at DC. It was seen that the gates on this batch of devices was much better than the previous set, giving an indication that the use of the bi-layer improves the gate lithography. However what was noted that the devices were working deep within enhancement mode, a large positive voltage had to be applied to the gate before current would flow through the device. This behaviour had been seen before on this type of material. The 70nm T-gate devices fabricated on A1881, had also shown this behaviour. It was thought that this was due to the etching of the gate recess which was formed using a highly selective etch, it was therefore unlikely that the etch depth was the source of the problem, but rather the recess width. Since the barrier layer was very thin (8nm) the majority of the electrons from the channel δ-doping layer filled the surface states rather than populate the channel. The wider the recess etch, the greater the number of surface states and the increased depopulation of the channel. It was not possible to investigate this effect using the 1-D Poisson solver that was normally used to find the electron density within the channel, since this was a 2-D problem, as it depended on the width of the etch not just the depth.

The device sample was cross-sectioned and viewed in the S900, to gather information on two things. The first was the gate length of the devices as the use of the bi-layer had not been tested prior to this sample so it was important to check the profile of the gate and to see that the length of the gate is indeed 50nm. It was also important to find the width of the gate recess offset. However because of the aforementioned problems of cleaving and viewing of device samples in the S900 it was not possible to obtain an image that showed the T-gate and gate etch at the edge of the sample so the gate length and the recess offset could be accurately measured. Figure 6.16 shows the resultant image, it is seen that the gate has broken away from the edge of the sample. This made it difficult to get an accurate value for the gate length of the device. It was seen however that the recess width at the edge of the sample was around 60nm. From the image below it is seen that the recess is very small, around 5nm. This gives an estimated gate length of 50nm.

Figure 6.16 SEM cross section of a 50nm device which operated deep within enhancement mode The gate recess is small, of the order of 5nm.

A further set of devices were fabricated on wafer A1881, the process flow used was the standard process for lattice matched InP HEMTs. The ohmic contacts were a Ni:Ge:Au recipe annealed at 240/280°C for 30/60 seconds. The measured ohmic contact resistance was 0.07Ω.mm. The mobility and the carrier concentration was found to be 5047 cm^2V-s and $9.95x10^{12}$ cm^{-2} respectively with the cap on. This suggests that this material is suitable for fabricating high speed devices.

The gate process used was the co-polymer bi-layer described above, the central dose was 1389μC/cm^{-2} and the central alignment strategy for improved focussing was employed. The gate recess was formed by etching for 17 seconds using the succinic acid based etch. This time was chosen to reduce the recess offset as much as possible while ensuring the gate was not in contact with the highly doped cap layer. Immediately before metallisation the sample was dipped in HF for 30 seconds. The metal used was 15nm Ti: 15nm Pt: 180nm Au. The sample was lifted off and completed with CPW bond-pads.

6.7 DC Characterisation of 50nm T-Gate Devices on A1881

The completed devices were characterised at DC using the methods and equipment detailed in section 5.4 The device yield was extremely high, 50 devices were measured, all but one of the devices had both gates present and showed good gate control. This shows the quality of the gate lithography and the gate recess etching of the devices. The output characteristics of a typical device is shown in Figure 6.17. This is a 2x20μm measured over a drain voltage range of 0 to 1.1V. Above 1.1V, the device begins to breakdown.

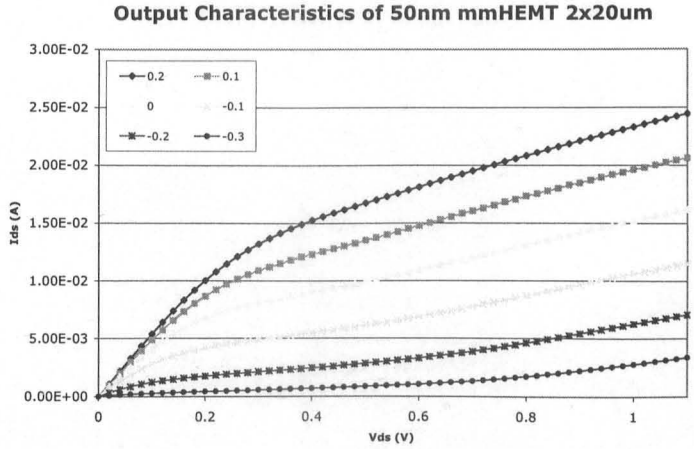


Figure 6.17 Output Characteristic of a 2x20μm 50nm metamorphic HEMT.

Breakdown is characterised by the rapidly increasing drain current and the reduction of gate control of the device. This low breakdown voltage can be explained by the very short recess

offset required to ensure that the devices would operate in depletion mode. The transfer characteristics of the device including the transconductance is shown in Figure 6.18 It can be seen that the transconductance of the device is very high, with a figure of 1200mS/mm achieved at an applied gate voltage of –0.05V. The current density is around 420mA/mm compared with typical values of 700mA/mm for a longer gate length lattice matched InP devices.

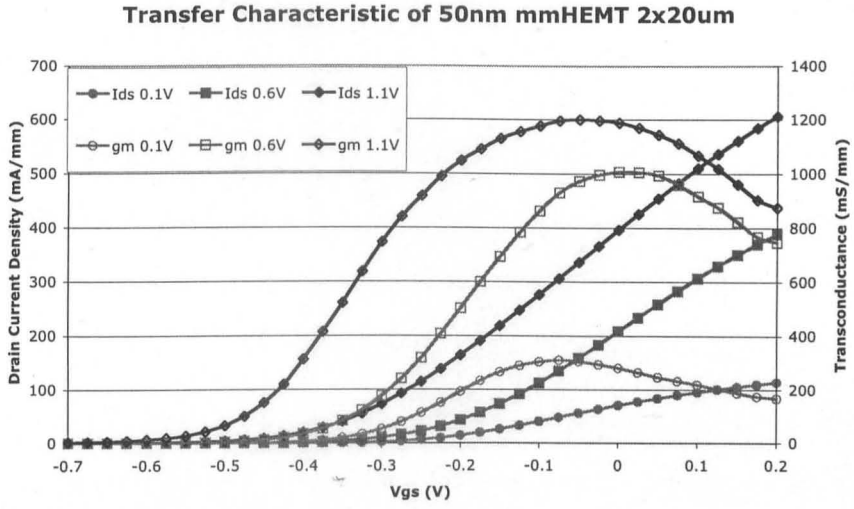


Figure 6.18 Transfer characteristics of 2x20μm 50nm HEMT fabricated on wafer A1881.

The design of this particular cell contained mostly 2x20μm devices, because the continuity of the gate is generally influenced by the width, smaller devices tend to be more complete and achieve better lift off, this design and the high yield allowed a number of identical devices to be measured. From this it is possible to find the spread of the DC figures of merit such as V_{th}, g_m, and I_{dss} which gives information on the uniformity of the gate lithography and the gate recess etch.

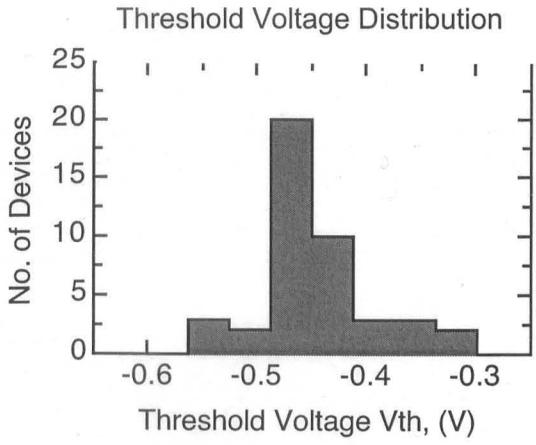
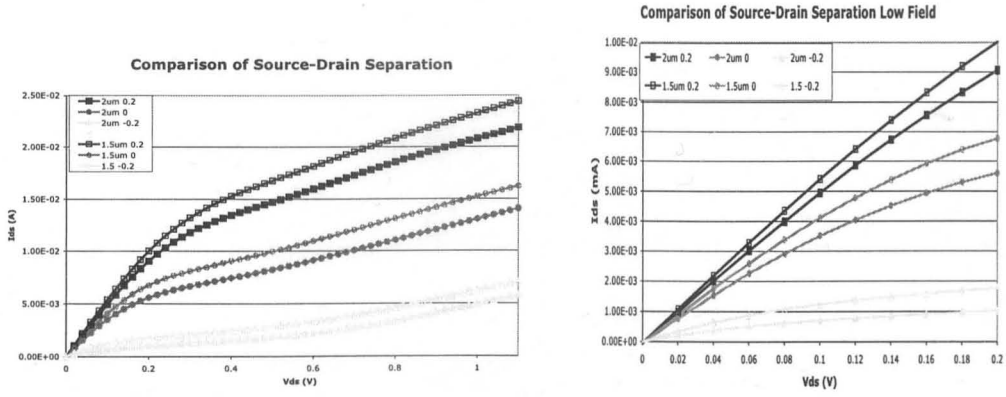


Figure 6.19 Histogram of Threshold voltage across 50nm Co-polymer cell.

Figure 6.19 shows the histogram of the measured threshold voltage of 43 devices measured across the cell. Threshold voltage was taken to be the value at which the drain current falls to 500μA for a drain bias of 1.1V. It is seen that out of the 43 devices measured, 30 devices have a threshold voltage of between –0.425V and –0.475V. The average value is -0.445V with a standard deviation of 0.005V. This is the lowest threshold voltage spread that has ever been seen at this gate length.

The unit cell contained two device geometries, these were identical with the exception of the width of the source-drain gap; 2μm and 1.5μm respectively. The smaller source-drain gap was designed to provide a lower source resistance, and therefore increase the extrinsic transconductance of the device. The inclusion of both designs within the cell allows for direct comparison between the DC and RF performance of each design. The peak transconductance of twenty devices of each design were measured. The average transconductance of the 1.5μm devices was 1169±83mS/mm compared with 1085±75mS/mm for the 2μm. These values show the importance of optimising the devices structure to minimise the parasitic resistances in addition to aggressive gate length scaling. Figure 6.20 shows the effect that the reduction of the access resistance has on the output characteristics.



(a) (b)

Figure 6.20 Comparison of two source-drain separations (a) Output characteristics of 2x20μm devices for three gate bias voltages (b) Low Field Behaviour of the same devices showing the decrease in series resistance with shorter source-drain gap.

It is seen that for a given bias condition the channel current is greater for the device with a 1.5μm separation. The low field behaviour is also shown; the 1.5μm device has a lower series resistance.

It is seen from the DC characteristics that the device process would be ideal for the fabrication of millimetre monolithic integrated circuits (M^3ICs) such as amplifiers because of the uniformity of both the threshold voltage and the transconductance. However it is the performance at RF frequencies that the devices are judged by. It is important that the theoretical benefits of the scaled gate length such as the reduced gate capacitances, are seen and that the devices out perform devices of a larger gate length fabricated on similar materials.

6.8 RF Characterisation of 50nm HEMT on A1881

The devices were measured at RF using the techniques and equipment detailed in section 5.5. The devices were measured with the drain bias in the range of 0.1V to 1.1V and a gate bias of 0.1V to –0.5V. This allows the performance of the device to be measured over a range of bias conditions while not damaging the device by operating in the breakdown regions. From the measured S-parameters an equivalent circuit model would be produced.

6.8.1 Cold FET Analysis of 50nm HEMTs

A new method was used to find the small signal model of these devices, as it is important when creating a model that the parasitic elements are accurately determined. With the standard modelling technique based on hot FET measurement (where a bias is applied across the source and drain) accurate determination of these components is difficult as the effects of the parasitic elements are second order. Diamant and Laviron [6.10] suggested that "cold FET" measurements of the S-parameters could be used to find the device parasitics because the equivalent circuit is much simpler. The intrinsic elements such as C_{gs}, C_{gs}, g_m and g_{ds} are bias dependent and need to be determined at the bias point of interest, however the parasitics are largely independent of bias. Hence values determined by "cold FET" measurements are valid when the biases are applied. The equivalent circuit can be further simplified by applying a positive gate bias generating a gate current. If the gate current density is in the regions on $5x10^{-7}$ -10^{-8} A/m^2 the capacitance of the gate become negligible and the Z_{11} parameter becomes real and the influence of the parasitic capacitances C_{pg} and C_{pd} become small. The extrinsic Z parameters then be written as

$$Z_{11} = R_s + R_g + \frac{R_c}{3} + \frac{nkT}{qI_g} + j\omega(L_s + L_g) \qquad (6.1)$$

$$Z_{12} = Z_{21} = R_s + R_c/2 + j\omega L_s \qquad (6.2)$$

$$Z_{22} = R_s + R_d + R_c + j\omega(L_s + L_d) \qquad (6.3)$$

From equations 6.1-6.3 it is straightforward to find values for the parasitic resistance and parasitic inductances from the real and imaginary parts of the impedance respectively.

The parasitic capacitances similarly can be found by reducing the conductivity of the channel, applying a voltage less than the threshold voltage fully pinches off the channel. The resultant small signal model is given in Figure 6.21 where C_b is the fringing capacitance due to the extended depletion region either side of the gate.

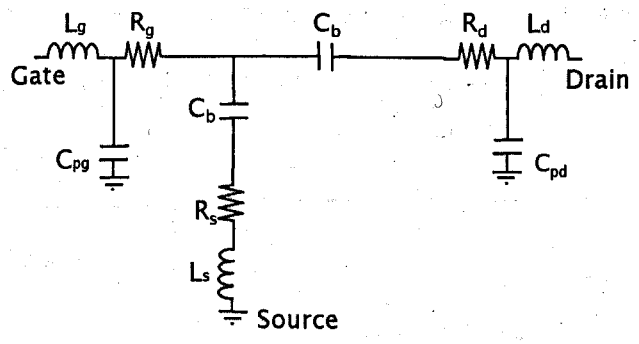


Figure 6.21 Equivalent circuit model of a FET measured with zero drain bais and gate voltage less than the pinch off voltage. After G. Dambrine et al. [6.11]

With the channel pinched off, up to a few of gigahertz the influence of the resistances and inductances has no effect, the imaginary part of the admittance parameters become.

$$\mathrm{Im}(Y_{11}) = j\omega(C_{pg} + 2 \cdot C_b) \qquad (6.4)$$

$$\mathrm{Im}(Y_{12}) = \mathrm{Im}(Y_{21}) = -j\omega C_b \qquad (6.5)$$

$$\mathrm{Im}(Y_{22}) = j\omega(C_b + C_{pd}) \qquad (6.6)$$

Using the above relations the parasitic capacitances can be determined.

This technique was used to determine the parasitic elements associated with a 2x40μm device with a source-drain gap of 2μm. The device was measured at two bias points, cold FET pinched off, (V_{ds} = 0V, V_{gs} = -1V) and cold FET forward biased (V_{ds} = 0V, V_{gs} = +1V). There was a problem with the technique, however, in order to match the simple model to the measured data transmission lines to represent the CPW bondpads had to be included. The behaviour of the imaginary part of the admittance was more sensitive to small changes in line length than changes in the inductance. This requires a high degree of accuracy in probe positioning to avoid errors in the inductance value extracted. The extracted values were then set as fixed values in the Lysander extraction program, using a number of biases over a range of V_{ds} =0.1V to 1.1V and V_{gs}= 0.1V to –0.2V. The resultant values were then imported into Microwave Office where the model was fine tuned. Figure 6.22 shows the resultant modelled S-parameters along with the measured S-parameters. It was found that the behaviour of the transistor was not ideal, for a

number of reasons. There was a "kink" in the magnitude of both S11 and S21, this made the transistor difficult to model over the full frequency range. The model follows the trend of all the parameters but had been optimised to be more accurate at frequencies above 40GHz.

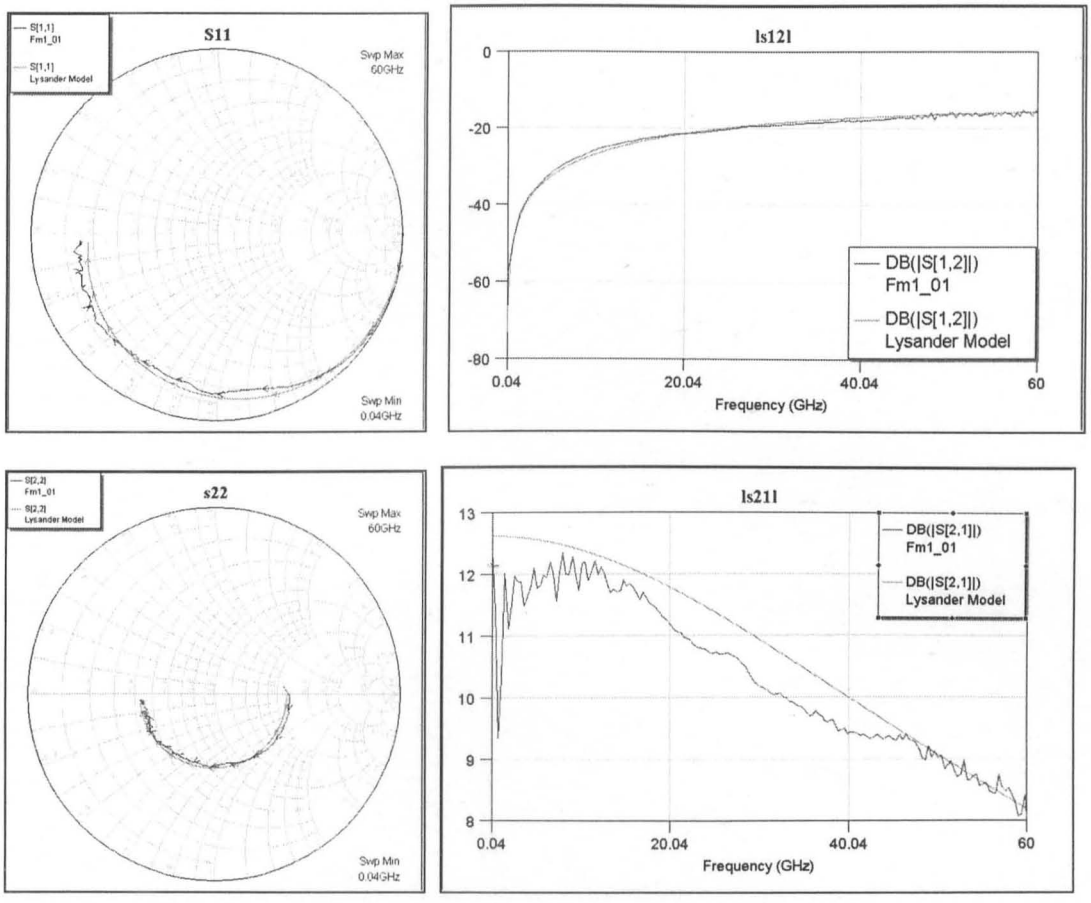


Figure 6.22 Measured and Modelled s-parameters of a 2x30μm 50nm HEMT measured between 0.04 and 60GHz.

The magnitude of S21 was the hardest parameter to fit, it is seen that there is a bend in the plot between 20 and 45GHz, which could not be replicated in the small signal model. The transconductance and the output conductance were found by plotting the real parts of Y21 and Y22 respectively, the fit of both of these parameters was good.

The model was de-embedded by removing the transmission lines from the model, the resulting |H21| and MAG are shown in Figure 6.23. The f_t can be found by extrapolating |H21| at 20dB/decade to unity gain, the extrapolation of f_{max} is not as straightforward. The modelled data

must be extended up to a frequency of 160GHz, at which point the device becomes unconditionally stable. The MAG is then extrapolated at 20dB/decade from this point to unity. Using the above method values of 330GHz and 250GHz were found for f_t and f_{max} respectively.

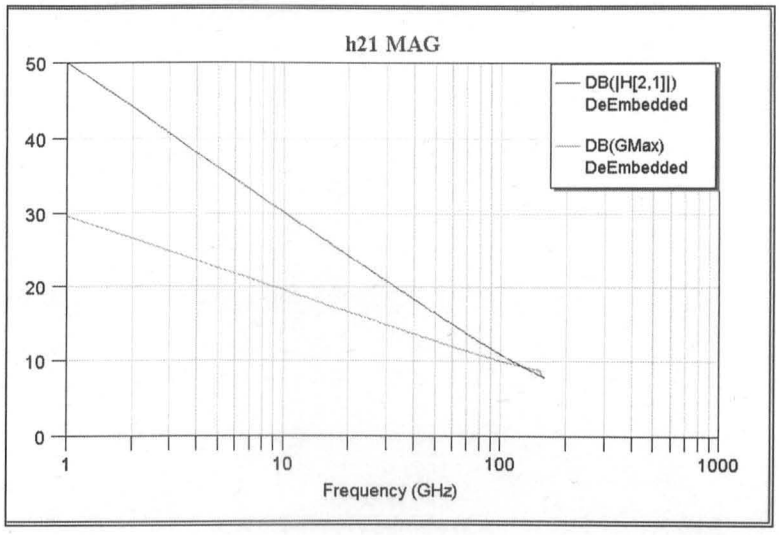


Figure 6.23 H21 and MSG/MAG of de-embedded model based on 2x30μm 50nm HEMT fabricated on A1881.

Parameter Value	50nm De-Embedded Model
g_m /(mS)	85.6
g_{ds}/(mS)	12.1
C_{gs} /(fF)	8.8
C_{gd} /(fF)	4.7
C_{ds} /(fF)	11.7
R_{gs} /(Ω)	18000
R_{gd} /(Ω)	134
R_s /(Ω)	3.33
R_d /(Ω)	5
R_g /(Ω)	3
L_s /(pH)	0.3
L_d /(pH)	1.5
L_g /(pH)	3.4
C_{pd} /(fF)	6.9
C_{pg} /(fF)	11.1
C_{pgd} /(fF)	6.2

Table 6.2 Values of the Lumped Element Circuit of 2x30μm 50nm HEMT on A1881

The main points to note in the lumped element values are the high transconductance around (1400mS/mm) and the low values of the intrinsic gate capacitances, C_{gs} and C_{gd} as expected from the reduction of the gate length. The parasitic capacitances, in comparison are relatively large. It is thought that this arises from the thin resist that forms the gate stalk leading to

increased capacitance between the gate and the cap layer. The combination of a high transconductance and low gate capacitance leads to a high figures for f_t and f_{max}.

The two 2x20μm devices with a source drain gap of 1.5μm and 2μm shown in Figure 6.20 were characterised between 0.04 and 60GHz. Figure 6.24 shows the S-parameters of both devices, it is seen form this that the parameters S11 and S12 are almost identical. However the is a clear difference between the forward transmission coefficient, S21, and reverse reflection coefficient, S22 of both devices. This suggests the different geometry causes changes in the drain side of the active region.

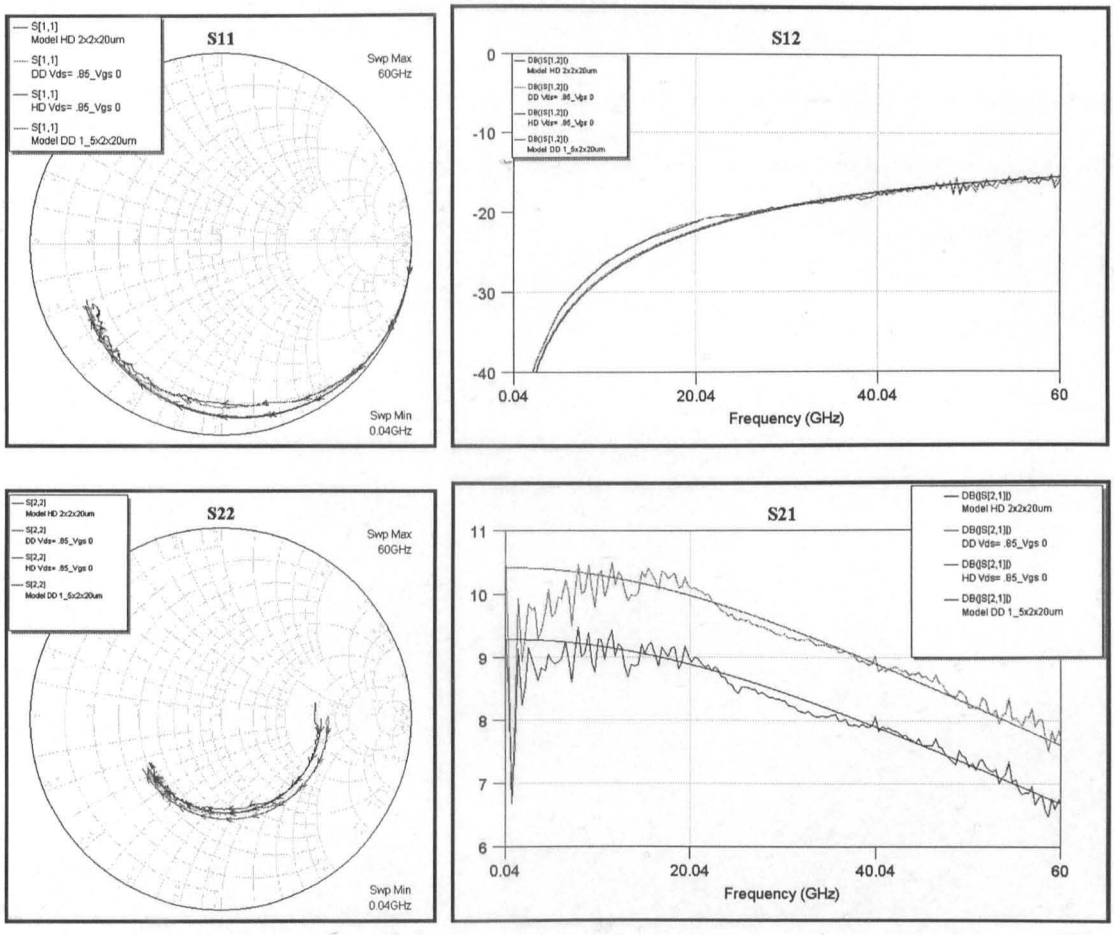


Figure 6.24 Measured and Modelled s-parameters of 2x20μm 50nm HEMTs with a source-drain separation of 1.5μm (Device DD) and 2μm (Device HD) respectively.

The two models were produced in a similar manner to the comparison of the standard and self-aligned pHEMT devices in section 5.6. A model was produced that fitted accurately the S-

parameter of the devices with the 2μm source-drain separation using the method described above. This model fitted the data well with a number of exceptions; it was not possible to fit S12 throughout the full frequency range, as there is a noticeable change in gradient which is not possible to account for in a model such as this. The parameter S11 also showed some irregularities, there is a dip in the magnitude that is not possible to model, as shown in the Smith Chart.

This model was replicated and modified to match the 1.5μm data by varying a limited number of terms, g_m, g_{ds}, R_s, and R_d. The values of the two models are compared in Table 6.3. To account for the differences in S22 the output conductance needed to be lowered as was the drain resistance. This change in output conductance caused a large change in the magnitude of S21, however, by decreasing the source resistance a close fit was obtained. It was however necessary to slightly increase the intrinsic transconductance of the device (~2%) to achieve a better match, this change is thought to be due to a difference in the intrinsic device rather than a change caused by reduced resistance which should be accounted for by R_s and R_d.

Parameter Value	50nm 1.5x2x20μm	50nm 2x2x20μm
g_m /(mS)	61.3	59.5
g_{ds} /(mS)	8.04	9.7
C_{gs} /(fF)	4.5	4.5
C_{gd} /(fF)	5.02	5.02
C_{ds} /(fF)	11.7	11.7
R_{gs} /(Ω)	30k	30k
R_{gd} /(Ω)	148	148
R_s /(Ω)	5.7	7
R_d /(Ω)	5.72	7.1
R_g /(Ω)	3	3
L_s /(pH)	0.13	0.13
L_d /(pH)	3	3
L_g /(pH)	3	3
C_{pd} /(fF)	4.6	4.6
C_{pg} /(fF)	6.43	6.43
C_{pgd} /(fF)	3.27	3.27

Table 6.3 Lumped Element circuit values of 2x20μm devices with a 1.5 and 2μm source-drain separation respectively.

The effect of the reduction of the parasitic resistances can be seen in Figure 6.25, this shows the short circuit current gain, h21, and the maximum unilateral gain, U, as a function of frequency. The device with the reduced source drain separation has 2dB greater |H21| at 20 GHz, this translates in to higher cut-off frequency, f_t, of around 330GHz compared to 310GHz for the device with a 2μm source-drain gap. The unilateral gain is also increased with the reduction of parasitic resistances, this was plotted in this case rather than the usual MSG/MAG because it doesn't require modelling to very high frequencies to find MAG and thus extrapolate f_{max}. Extrapolating U to unity gain gives an f_{max} of 230 and 260GHz for the 2μm and 1.5μm devices respectively.

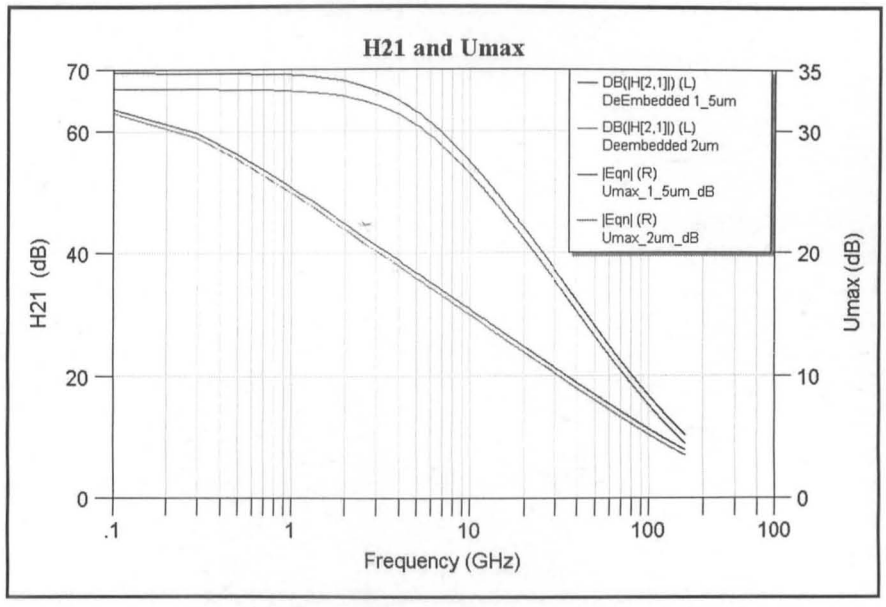


Figure 6.25 Short circuit current gain and the maximum unilateral transducer, U, gain.

6.9 Conclusions

It has been seen from this chapter that it is possible to use the PMMA/Co-polymer resist scheme to fabricate T-gate with a gate length as small as 50nm. These devices demonstrated high yield and excellent uniformity. The DC performance showed that the gate was uniform along the width of the device, with a pinch-off voltage of –0.45V. The transconductance of the devices was 1085 and 1169 mS/mm for a source-drain separation of 2 and 1.5μm respectively. This

demonstrates the importance of the parasitic elements when fabricating high speed HEMTs using an ultra-small gate length.

The RF performance of the devices was excellent, with a cut-off frequency of 330GHz for both a 2x30μm device with a source-drain separation of 2μm and a 2x20μm with a source-drain separation of 1.5μm. This suggests that with the reduction of the access resistance due to a reduction of the source drain gap, the high frequency performance of this technology can be improved further. The maximum frequency of oscillation of these devices was 250GHz and 260GHz respectively. In future this could be improved by the reduction of the resistance of the gate, for example by increasing gate metal thickness.

The combination of high speed performance and the highly uniform DC characteristics mean that this technology is very well suited to be used as a active component of M^3IC's operating deep into the millimetre wave regime.

6.10 References.

[6.1] K. Kalna et al. ""*Scaling of pseudomorphic high electron mobility transistors to decanano dimensions*" Solid State Electronics, **46**, 2001, p631

[6.2] Y. Chen, D. Macintyre and S. Thoms, "*Electron beam lithography process for T- and -shaped gate fabrication using chemically amplified DUV resists and PMMA*", J. Vac. Sci. Technol. **B 17**,1999 p2507

[6.3] L. Grella et al. "*Nanolithography performances of ultraviolet III chemically amplified positive resist*" "", J. Vac. Sci. Technol. **B 15,** 1997, p2569

[6.4] S. Thoms, "*Electron Beam Lithography Course Notes*", Lesson 7

[6.5] Y. Chen et al., "*Effect of Resist Sensitivity Ratio on T-gate fabrication*", J. Vac. Sci. Technol., **B19**, 2001, p2494

[6.6] I. Maximov et al., ""*Investigation of polymethylmethacrylate resist residues using photoelectron microscopy*", J. Vac. Sci. Tech., **B20**, 2002, p1139

[6.7] D.L. Edgar et al., "*millimeter-wave performance of InAlAs/InGaAs HEMTs using a UVIII/PMMA bilayer for 70 nm T-gate fabrication*", Proc. of European Microwave Conference 2002, Milan, October 2002

[6.8] S. Bollaert et al., "$0.06\mu m$ *gate length metamorphic* $In_{0.52}Al_{0.48}As/In_{0.53}Ga_{0.47}As$ *HEMTs on GaAs with high* f_t *and* f_{max}" Proc. IPRM 2001, 2001, p192

[6.9] A.Endoh et al. "*InP-based High Electron Mobility Transistors with a very Short Gate Channel Distance*", JJAP, **42**, 2003, p2214

[6.10] F. Diamant and M. Laviron, "*Measurement of the extrinsic series elements of a microwave MESFET under zero current condition*" Proc. 12th European Microwave Conference, 1982, p451.

[6.11] G. Dambrine et al. "*A New method for Determining the FET Small-Signal Equivalent Circuit* ", IEEE Trans. Microwave Theory and Techniques, **36**, 1988, p1151.

7.0 Transmission Electron Microscope (TEM) Analysis of HEMT Devices

Chapter six showed the importance of analytical techniques such as cross sectional SEM in the development of ultra-small gate length transistors. It is invaluable to have a method of studying the profile of fabricated devices. Important parameters that determine the behaviour of the transistor such as gate length, recess depth and recess offset can all be determined from the device cross section. There are limitations to the standard method of cross sectional analysis by SEM, these will be discussed in the following section. The benefits of the cross-sectional TEM samples prepared using focused ion beam techniques will be explored and the development of a technique of preparation will be detailed.

7. 1 Cross-sectional Analysis by SEM

Cross-sectional analysis by SEM is a quick and powerful technique for the investigation of the geometry of fabricated devices. However there are some limitations to this technique. The first is purely a practical issue, the difficulty of preparing the sample so that the T-gate is sectioned at the edge of the sample. The standard method of preparation of cross sectional sample is to use a diamond tip scribe equipped with a micrometer controlled stage and low magnification microscope to cleave the sample at the point of interest. The average width of a short gate length device is 20μm so to be able to cross section the gate the cleave must be accurate to within ±10μm, which is not easy to achieve on the equipment available. The second effect is even more serious, sub-100nm T-gates are very fragile with very little metal in contact with the substrate this leads to a tendency for the gates to lift-off or break at a distance away from the sample edge. These two effects lead to a very low yield of samples that are suitable for examination. This problem may appear trivial but it is difficult to overcome, the position of the cleave can be improved with practice, however the behaviour of the gate is essentially random.

It is possible to produced test pieces designed to be examined in this way. These test samples have a much higher density of gates, many of them overlapping to increase the probability of a useful profile. The drawback of this method is that it provides information on the test piece rather than the actual sample never mind the actual device of interest so this method is of limited use.

There is a more fundamental problem; it is concerned with the ultimate resolution available using SEM. The microscope that is used in this work is a Hitachi S900 SEM, this has a very small stage, which allows a very small working distance, it can be used to view sample up to a magnification of x800k. However at this magnification the setup of the microscope becomes crucial and becomes increasingly difficult to achieve good images. Usually there is always

some imperfection in either the focus or the stigmation of the beam, which makes it difficult to determine lengths with an uncertainty of better than ±3nm. This is acceptable when looking at devices with a gate length of 50nm, however it makes a great deal of difference when the gate length is 25nm.

An ideal technique of analysis would provide very high resolution and provide very accurate control of the position of the cross section. This can be achieved by TEM analysis of samples prepared using Focus Ion Beam (FIB) milling. The use of TEM provides atomic resolution and so it is possible to measure even the smallest features with a high degree of accuracy. FIB milling allows placement of a cross section sample to within ±50nm, this allows not only individual devices to be selected but even specific areas of the device. The size of the sample is very small therefore several samples can be made along the width of the gate. The process is non-destructive to the sample as a whole so other devices on the sample can be tested at a later date.

7.1 Sample Preparation by Focused Ion Beam Milling

A FIB system operates in much the same was as a SEM with the exception that ions are used instead of electrons for imaging. The ions are field extracted from a liquid metal ions source that consists of a tungsten needle wetted by a liquid metal, typically gallium. When an electric field is applied ($> 10^8 V.cm^{-1}$) a cone with a radius of 10nm is formed and from this the ions are extracted.[7.1] The extracted ions are then accelerated with an accelerating voltage of between 5 and 30keV. The ions are focused on the sample, swept across the sample and either the electrons or ions emitted from the sample can be used to image the sample much like a SEM. However in this work the beam of ions is used to sputter material away from the incident area. In this way the system can be used as a milling machine. The FIB system that was used in this work was the FIE FIB200TEM. This uses a gallium source and is capable of producing a beam

of accelerated ions with a spot size of 7nm. This very small spot allows very thin samples (less than 100nm) to be milled.

7.2 Methods for TEM sample preparation

There are two methods for the preparation of cross sectional samples for TEM, the Trench method and the Lift-Out Method.

7.2.1 Trench Method

In this technique a thin slice of material (typically 2mm long and 30 to 100μm thick) containing the area of interest is made by either dicing or polishing. The sample is mounted on a TEM sample grid for milling. A platinum bar with a thickness of around 1μm is deposited on the sample by cracking a platinum containing gas using the ion beam. The purpose of this bar is to protect the surface during milling. The sample is thinned using large beam currents to mill trenches on either side of the area of interest, these trenches are around 30μm wide and 30μm deep, using this large beam current the sample is thinned down to approximately 500nm. The sample is thinned down to around 100nm using progressively small beam currents. During this step the sample is tilted 1° into the beam (Figure 7.1) as each face is milled. This avoids material getting sputtered back on to the face of the sample and minimises the damage to the faces of the sample. The sample can then be removed and studied using the TEM.

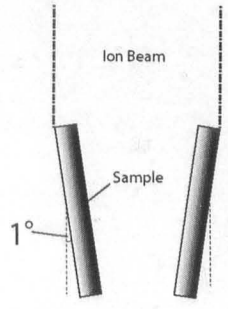
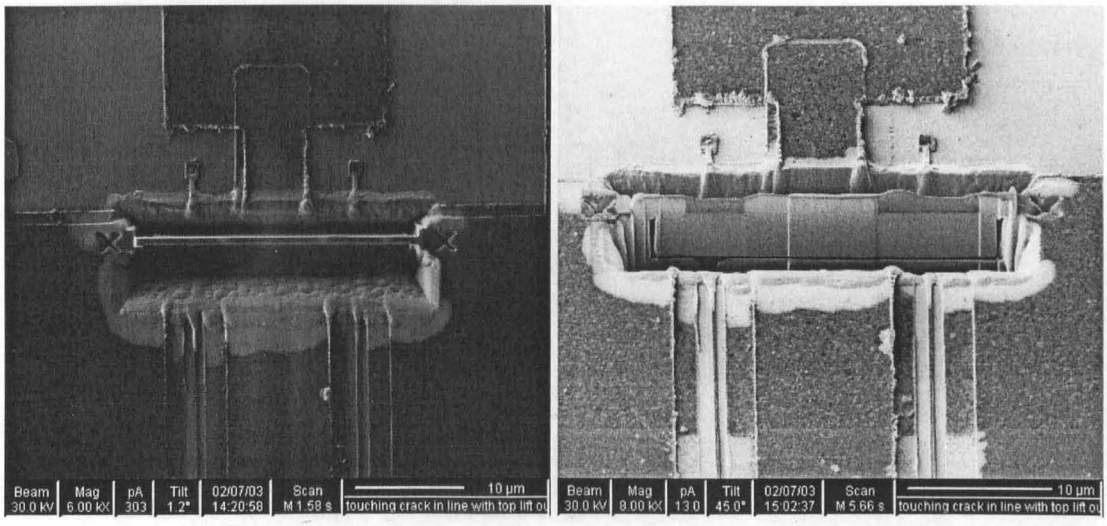


Figure 7.1 Tilting of sample during final stages of thinning to reduce damage.

This process has the advantage that transfer of the sample from the FIB tool to the TEM is straightforward as the sample is already mounted on the TEM sample holder before milling. The major disadvantage with this method is the preparation that is required before the sample can be milled. The dicing and polishing of the sample before milling is time consuming and destroys many devices on the sample. In addition, the time it takes to mill the sample strongly depends on the thickness of the sample which can lead to long milling times.

7.2.2 *The Lift-out Technique*

The second technique of sample preparation is the Lift-out method. In this technique the wafer is placed in the FIB milling tool intact, the sample is thinned in situ and lifted out. The first step is to select the region to be cross sectioned, two markers are then milled either side of this region, this act like the markers used in the beamwriter, it allows the tool to find and align to the correct region. A platinum bar is deposited in the same way as for the trench method. Two trenches are milled front and back of the area of interest, these are step like, getting deeper closer to the sample as shown in Figure 7.2 (a). The thickness of the sample at this stage is approximately 500nm.



(a) (b)

PAGE NUMBERING AS IN THE ORIGINAL THESIS

7.0 Transmission Electron Microscope (TEM) Analysis of HEMT Devices

Chapter six showed the importance of analytical techniques such as cross sectional SEM in the development of ultra-small gate length transistors. It is invaluable to have a method of studying the profile of fabricated devices. Important parameters that determine the behaviour of the transistor such as gate length, recess depth and recess offset can all be determined from the device cross section. There are limitations to the standard method of cross sectional analysis by SEM, these will be discussed in the following section. The benefits of the cross-sectional TEM samples prepared using focused ion beam techniques will be explored and the development of a technique of preparation will be detailed.

7. 1 Cross-sectional Analysis by SEM

Cross-sectional analysis by SEM is a quick and powerful technique for the investigation of the geometry of fabricated devices. However there are some limitations to this technique. The first is purely a practical issue, the difficulty of preparing the sample so that the T-gate is sectioned at the edge of the sample. The standard method of preparation of cross sectional sample is to use a diamond tip scribe equipped with a micrometer controlled stage and low magnification microscope to cleave the sample at the point of interest. The average width of a short gate length device is 20μm so to be able to cross section the gate the cleave must be accurate to within ±10μm, which is not easy to achieve on the equipment available. The second effect is even more serious, sub-100nm T-gates are very fragile with very little metal in contact with the substrate this leads to a tendency for the gates to lift-off or break at a distance away from the sample edge. These two effects lead to a very low yield of samples that are suitable for examination. This problem may appear trivial but it is difficult to overcome, the position of the cleave can be improved with practice, however the behaviour of the gate is essentially random.

It is possible to produced test pieces designed to be examined in this way. These test samples have a much higher density of gates, many of them overlapping to increase the probability of a useful profile. The drawback of this method is that it provides information on the test piece rather than the actual sample never mind the actual device of interest so this method is of limited use.

There is a more fundamental problem; it is concerned with the ultimate resolution available using SEM. The microscope that is used in this work is a Hitachi S900 SEM, this has a very small stage, which allows a very small working distance, it can be used to view sample up to a magnification of x800k. However at this magnification the setup of the microscope becomes crucial and becomes increasingly difficult to achieve good images. Usually there is always

some imperfection in either the focus or the stigmation of the beam, which makes it difficult to determine lengths with an uncertainty of better than ±3nm. This is acceptable when looking at devices with a gate length of 50nm, however it makes a great deal of difference when the gate length is 25nm.

The ideal technique of analysis will provide very high resolution and provide very accurate control of the position of the cross section. This can be achieved by TEM analysis of samples prepared using Focus Ion Beam (FIB) milling. The use of TEM provides atomic resolution and so it is possible to measure even the smallest features with a high degree of accuracy. FIB milling allows placement of a cross section sample to within ±50nm, this allows not only individual devices to be selected but even specific areas of the device. The size of the sample is very small therefore several samples can be made along the width of the gate. The process is non-destructive to the sample as a whole-so other devices on the sample can be tested at a later date.

7.1 Sample Preparation by Focused Ion Beam Milling

A FIB system operates in much the same was as a SEM with the exception that ions are used instead of electrons for imaging. The ions are field extracted from a liquid metal ions source that consists of a tungsten needle wetted by a liquid metal, typically gallium. When an electric field is applied ($> 10^8 V.cm^{-1}$) a cone with a radius of 10nm is formed, from this the ions are extracted.[7.1] The extracted ions are then accelerated with an accelerating voltage of between 5 and 30keV. The ions are focused on the sample, swept across the sample and either the electrons or ions emitted from the sample can be used to image the sample much like a SEM. However in this work the beam of ions is used to sputter material away from the incident area. In this way the system can be used as milling machine. The FIB system that was used in this work was the FIE FIB200TEM. This uses a gallium source and is capable of producing a beam

of accelerated ions with a spot size of 7nm. This very small spot allows very thin samples (less than 100nm) to be milled.

7.2 Methods for TEM sample preparation

There are two methods for the preparation of cross sectional samples for TEM, the Trench method and the Lift-Out Method.

7.2.1 Trench Method

In this technique a thin slice of material (typically 2mm long and 30 to 100μm thick) containing the area of interest is made by either dicing or polishing. The sample is mounted on a TEM sample grid for milling. A platinum bar with a thickness of around 1μm is deposited on the sample by cracking platinum compound gas using the ion beam. The purpose of this bar is to protect the surface during milling. The sample is thinned using large beam currents to mill trenches on either side of the area of interest, these trenches are around 30μm wide and 30μm deep, using this large beam current the sample is thinned down to approximately 500nm. The sample is thinned down to around 100nm using progressively small beam currents. During this step the sample is tilted 1° into the beam (Figure 7.1) as each face is milled. This avoids material getting sputtered back on to the face of the sample and minimises the damage to the faces of the sample. The sample can then be removed and studied using the TEM.

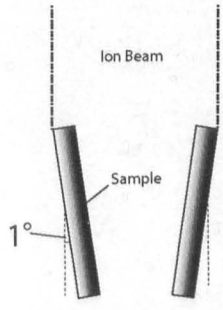
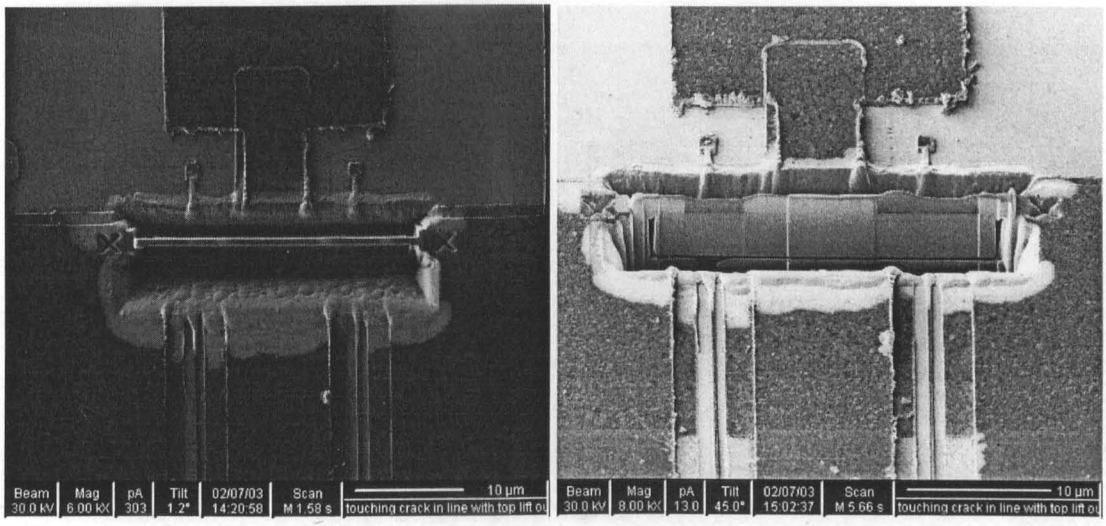


Figure 7.1 Tilting of sample during final stages of thinning to reduce damage.

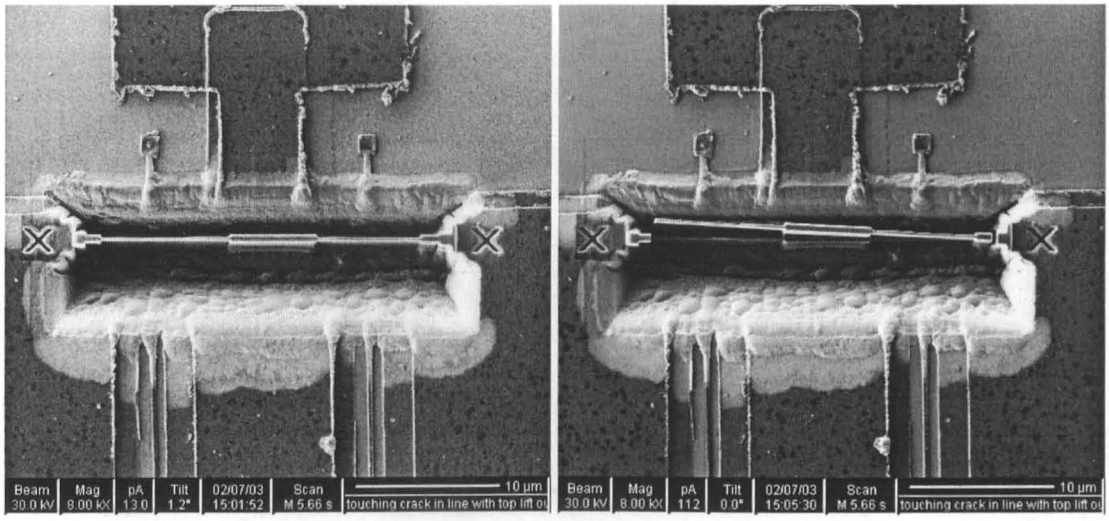
This process has the advantage that transfer of the sample from the FIB tool to the TEM is straightforward as the sample is already mounted on the TEM sample holder before milling. The major disadvantage with this method is the preparation that is required before the sample can be milled. The dicing and polishing of the sample before milling is time consuming and destroys many devices on the sample. In addition, the time it takes to mill the sample strongly depends on the thickness of the sample which can lead to long milling times.

7.2.2 *The Lift-out Technique*

The second technique of sample preparation is the Lift-out method. In this technique the wafer is placed in the FIB milling tool intact, the sample is thinned in situ and lifted out. The first step is to select the region to be cross sectioned, two markers are then milled either side of this region, this act like the markers used in the beamwriter, it allows the tool to find and align to the correct region. A platinum bar is deposited in the same way as for the trench method. Two trenches are milled front and back of the area of interest, these are step like, getting deeper closer to the sample as shown in Figure 7.2 (a). The thickness of the sample at this stage is approximately 500nm.



(a) (b)



(c) (d)

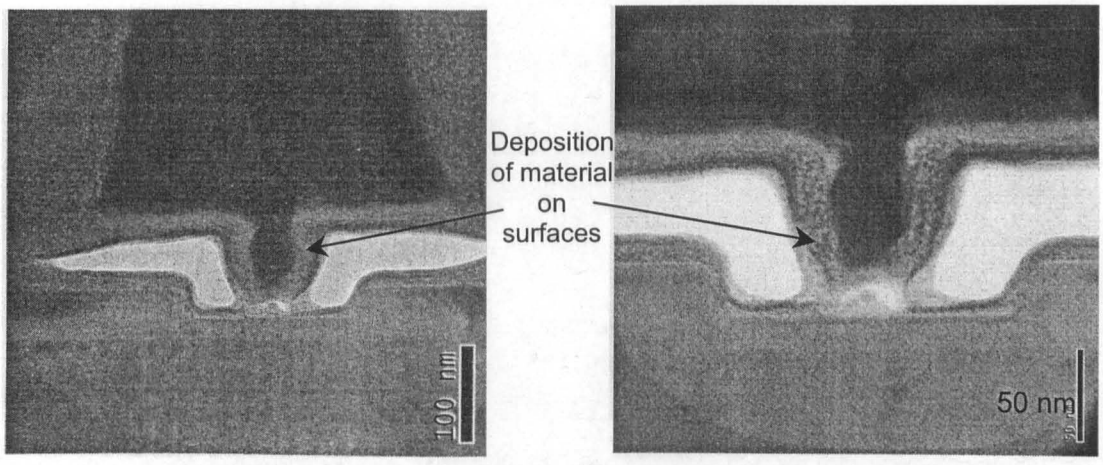
Figure 7.2 FIB images of TEM sample at various stages of fabrication.

The next stage of preparation is to tilt the sample 45° and cut down the sides and across the bottom of the sample leaving the sample held in place by only the small area at the surface as is clearly seen in Figure 7.2(b). Milling single pixel lines either side of the sample using progressively smaller beam currents are then used to thin the sample down further. During this milling the sample is tilted in the same way as the trench method to prevent damage to the surface. The sample is thinned down as necessary, although for high resolution TEM images the thickness should be less than 100nm. The plan view of a thinned down sample is shown in Figure 7.2(c), it is seen that the sample has been thinned only in the region of the gate, (this sample includes two gates) this reduces the milling time and increases the strength of the sample. The final stage of the milling is to fully free the sample from the substrate as seen in Figure 7.2(d). The sample must now be removed from the substrate and placed on the thin membrane of TEM grid. This is done using a micromanipulator equipped with a long working distance optical microscope. The probe is positioned in the shallow end of the trench and slowly moved towards the same until the sample is electro-statically attracted to the probe. The sample is then carefully transferred on to the TEM membrane. This is the most difficult stage of the process, the sample is very small (approx. 25x5x0.1µm) and is very fragile, however with care the success rate is very high.

The obvious disadvantage of this technique is the transfer from the substrate to the membrane. However it is quick, requires no time consuming preparation of the sample prior to milling and allows multiple cross-sections of a single gate. It was for these reasons that this method was chosen to prepare TEM samples of HEMT devices.

7.3 Sample preparation by the Lift-Out method

In the first instance the TEM sample was prepared exactly as above. The width was chosen to be 20μm, this allows both gates of the transistor to be included in the cross section. The platinum bar with a thickness of 1μm was deposited using the standard procedure. The samples were milled to a thickness of 100nm and then transferred onto the TEM grid. The samples were studied using the FEI TECNAI F20 Transmission Electron Microscope. This microscope is capable of operating in standard TEM, STEM and EDX modes. Sam McFadzean of the Department of Physics operated the microscope and all the following images were taken with his help. This specialist help was needed because a TEM is much more complex to operate than a SEM and requires a great deal of training and experience. Figure 7.3 shows the results of the first attempt at sample preparation.



(a) (b)

Figure 7.3 TEM Image of 50nm HEMT (a) Image showing dark gate head surrounded by material. (b) Close up of same device, the material also coats the substrate and into the gate recess.

The images obtained provided a great deal of information on the device geometry with a greater confidence than before. The depth of the recess was found to 29nm, this agrees very well with the material layer structure, the recess width was found to be 150nm and the slope of the etched side was found to be 63°, these figures were fed back to the device modelling group to enable them create more accurate models based on the measured device geometries. There were however a number of problems with the process as it stood; the first was that material that had been deposited, not only on the gate but also on the substrate and within the gate recess, making determining the exact geometry harder. The second problem is more important. It can be seen from Figure 7.3(b) that the gate had been lifted up slightly from the etch trench. Figure 7.4 demonstrates this effect further; this shows the other gate within the sample, the gate had been completely lifted off and is encased in platinum around 200nm from its original position.

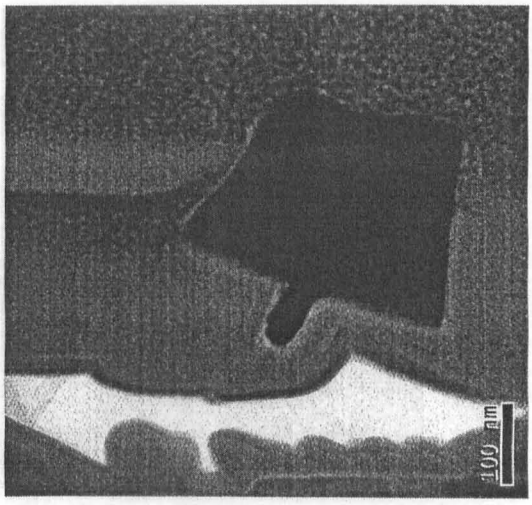


Figure 7.4 TEM Image of T-gate that has been lifted off the surface and is encased in the platinum protection layer.

It is noted that the gate had not only lifted up but had shifted horizontally as well. Other samples were studied and a similar behaviour was noted, of the two gates on a sample one gate would be fully intact while the other would be lifted off. The gate that was lifted off was always nearest the needle that injects the gas for the platinum deposition. The direction that the gates are shifted also corresponded to the direction of platinum deposition.

It was seen that for HEMT structures, because of the fragile, T-shaped gates some protection would be required to prevent damage from the deposition of the platinum protective layer. It was thought that if the gates were encased in a dielectric layer such as SiO_2, the gates would be suitably protected from the platinum deposition.

7.3.2 Sample Preparation using SiO_2

The use of SiO_2 as a protection layer was employed in the following way. The same device sample used previously had a 300nm film of SiO_2 deposited by PECVD. This is a low damage deposition process and should leave the device unaffected. The device was milled in the same way as before, with the exception that because of the thick layer of dielectric the sample suffered from charging effects, this meant that the charge neutraliser had to be used. This was switched off for the fine thinning stage. The sample was milled to around 100nm, although it was found that one site was milled slightly thinner than the other. Figure 7.5 shows the profile of the resultant sample, it is seen that the gate is firmly in place, supported by the light grey SiO_2 that surrounds the gate. The platinum layer can just be seen in the upper left hand corner as a lighter shade of grey. The dark areas either side of the gate are ohmic contacts, this being a 50nm self-aligned device.

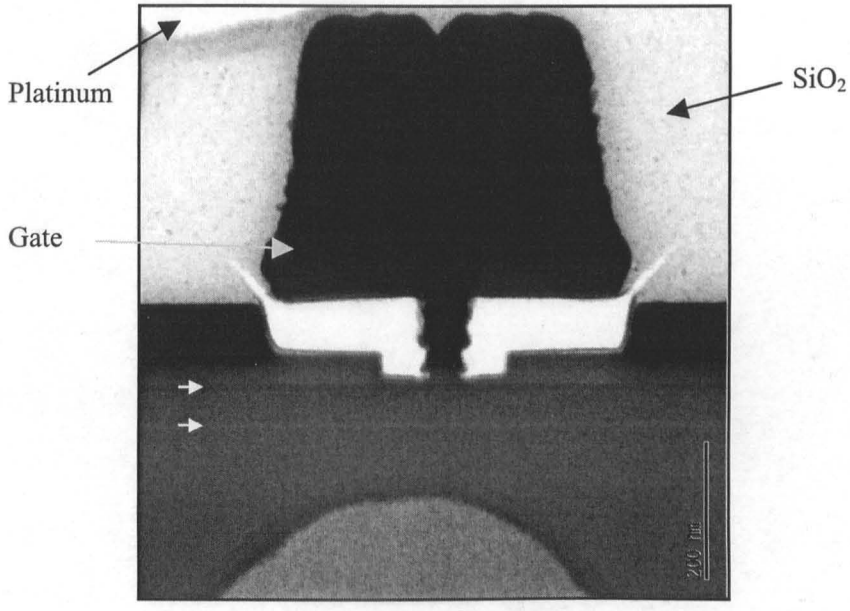


Figure 7.5 TEM image of site 1 of the sample prepared from a 50nm Self-Aligned HEMT including the SiO_2 layer protection layer.

The use of TEM rather than SEM gives much more information than just the simple device geometry. From Figure 7.5, is seen that it is possible to see the different layers that make up the layers structure (highlighted by the white arrows), it is also possible to see the diffusion of the ohmic contact into the cap layer. The second site on the wafer had been milled slightly thinner. The thinner sample was better for imaging however the recess region of the gate had been damage during milling.

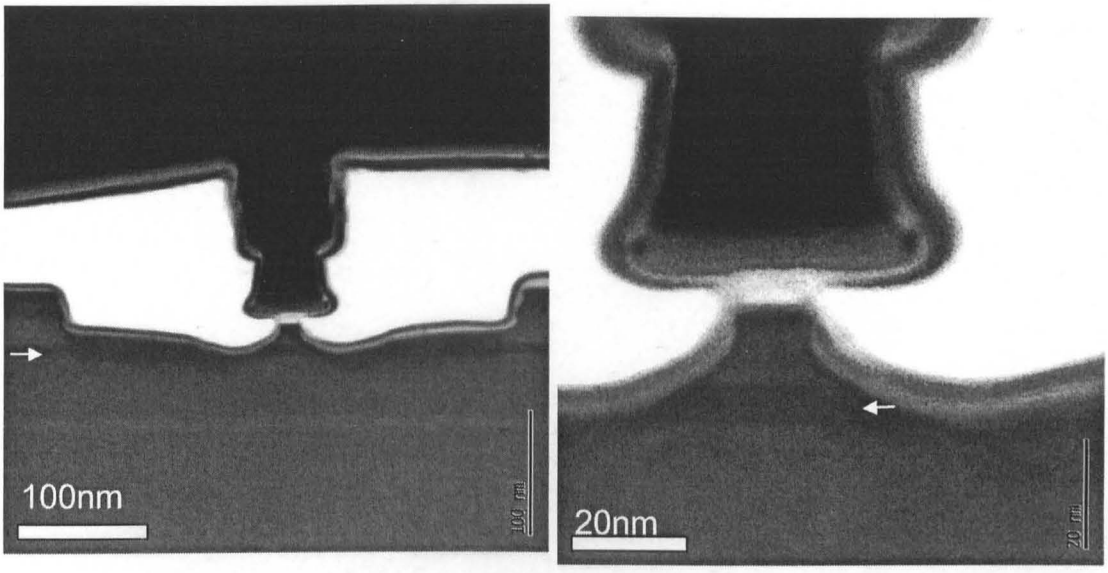


Figure 7.6 TEM image of site 2 of sample protected with SiO_2, the sample is thinner and has suffered damage due to milling. (a) Gate recess region, the epi-layer has been etched away to below the channel layer. (b) Gate foot sitting on a pedestal of epi-material.

Figure 7.6 (a)&(b) clearly demonstrate that material within the recess, particularly in the area close to the foot has been sputtered away to the bottom of the channel. (marked by white arrow) This damage is believed to be caused by the ions experiencing deflection as they leave the dense metal into the air gap underneath the gate. This deflection means the ions do not graze the recess region with a glancing blow as before but hit directly and sputter away material. It is also noted from these images that there is a thin pale line around the gate and the recess region. It was originally thought that this might be some deposition however by adjusting the focus it is possible to resolve the edges suggesting that this is an artefact of the rounded edge of the sample, again possibly caused by the air gap. By eliminating the air-gap between the gate and the recess it may be possible to reduce the damage at the surface of the recess. This would

require a dielectric that was able to fill the region underneath the gate head. This will be discussed in section 7.3.3.

STEM and EDX of SiO_2 protected sample.

Even greater resolution can be achieved by operating the microscope in STEM mode, In this mode, a very fine beam of electrons form a probe. This probe is then scanned across the sample much like the beam of an SEM. Using this method the image shown in Figure 7.7 was taken. This is a close up view of the epi-layers of the material, at the edge of the ohmic contacts. It is possible to see all the layers, including the etch stop, the GaAs surface layer below the etch stop and even the thin GaAs layer that contains the Si delta doping layer (thin pale line).

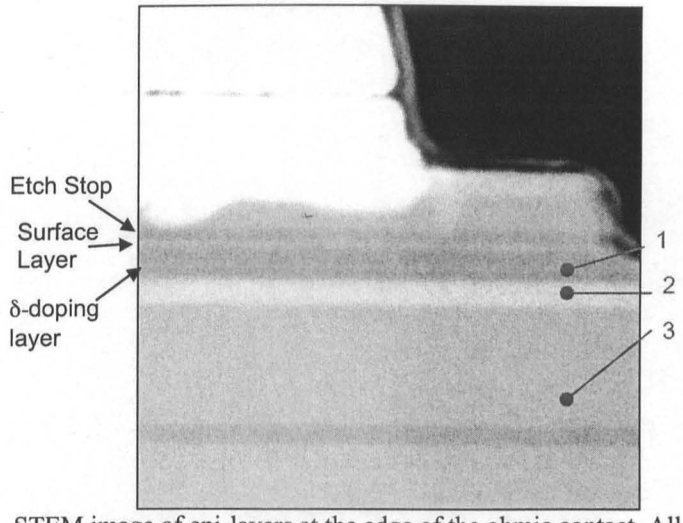


Figure 7.7 STEM image of epi-layers at the edge of the ohmic contact. All the individual layers of the epi-structure are visible.

It is also possible to see the diffusion of the non-annealed ohmic contacts, the metal (white area) has sunk from the surface down through the cap layer to the top of the etch stop layer. By operating the microscope in EDX mode, which uses the ultra fine beam of STEM mode and placing the beam in a particular area of the sample and the emitted x-rays can be collected. It is possible to determine which elements are present in a very small area of interest. As the layer structure used is known accurately from MBE growth sheets this provided an excellent opportunity to test this tool for applications in HEMT samples. The probe was placed in three positions, labelled 1,2,and 3 in Figure 7.7

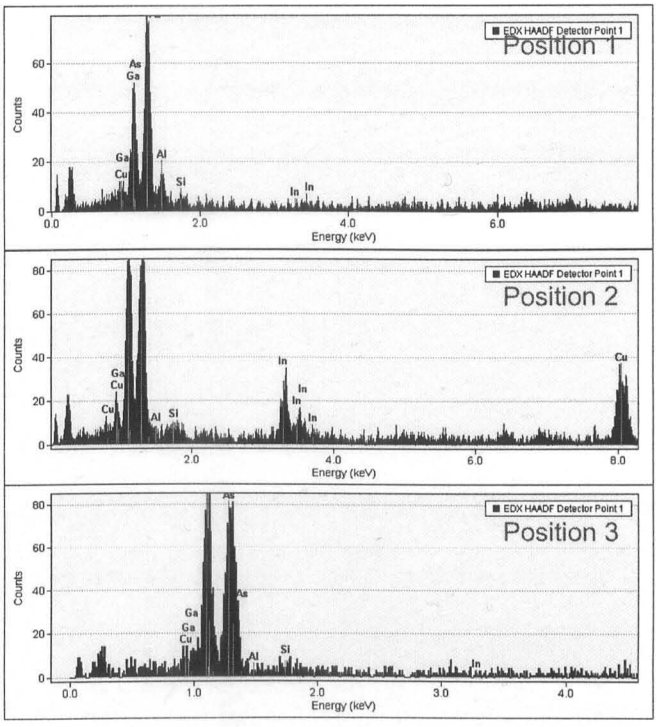


Figure 4.8 EDX measured energy spectra of three position within epi-layer structure.

Figure 4.8 shows the energy spectra of the collected x-rays for each position. In position 1, the spectrum is almost fully made up of gallium, arsenide and some aluminium. There is a slight trace of silicon that can be seen over the background readings. This is as expected, this position is in the δ-doping layer, which consists of a very thin a layer of silicon surrounded by three mono-layers of GaAs, within the AlGaAs barrier layer. The spectrum of position 2 shows gallium and arsenide in addition to a noticeable peak at 3.6eV corresponding to indium. This is what is expected as this position corresponds to the $In_{0.2}Ga_{0.8}As$ channel layer. It is noticed that there is also a large trace of copper arising from the sample being positioned on a copper TEM grid. Position three is within a thick GaAs buffer layer, as expected the spectra shows large peaks for the energies of gallium and arsenide.

It must be noted however that the results presented above are not ideal, due to the orientation of the sample as it was placed in the TEM, the angle that provides high quality images (along the zone axis) corresponds to an angle that provides a very low number of x-rays into the detector, this accounts for the very low number of readings taken. Despite this detail, this work has demonstrated the possibilities of using this technique for finding the elemental makeup of very

small areas of a HEMT device. This could be of use when studying the diffusion of ohmic contact metals to find out where each element of the metallisation is located after diffusion. This could also be of use for reliability studies examining the problem of gate sinking.

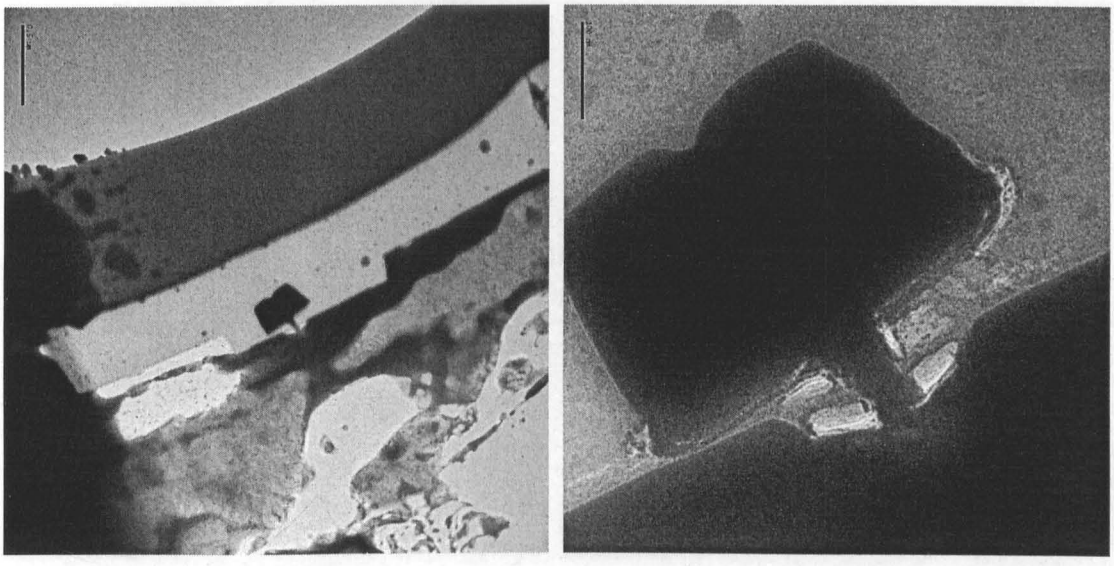
7.3.3 Preparation of samples using HSQ

Figure 7.6 showed the damage to the semiconductor layer located under the gate head caused by the milling process. It was thought that this was caused by the deflection of the gallium ions as they come out of the dense metal gate into the air-gap between the gate head and the substrate. It was believed that if this gap could be eliminated that the surface of the semiconductor could be protected. Two possible methods could be employed to achieve this. It had been noted that Silicon Nitride deposited by Inductively Coupled Plasma (ICP) could be deposited under the head of the gate [1.2]. The second method was to use a spin on oxide, in particular FOx© 14 produced by Dow Corning, hydrogen silsesquioxane (HSQ). This comes in a resin form and consists mainly of Si-O networks and partially formed cage like structures that contain Si, O, and H. The normal process to turn this into an oxide is to bake the sample above 200°C which opens the cage structures and forms a three dimensional network, which gives the film its strength [1.3]. However it has also been found to be a good negative electron beam resist as well, instead of the heat changing the composition, the exposure to electrons turns it into it's dielectric state. [1.4] In this work the transition was caused by the secondary electrons emitted from the sample when the FIB images the sample.

The sample that was studied using this technique was a 40nm HEMT on an InP substrate. The structure of this device is slightly different to the structures that were studied before; as it incorporates a SiO_2 gate support layer. The details of this process are given in Chapter 8. The device sample was coated in undiluted FOx© 14 to provide a layer with a thickness greater than 300nm. The sample was spun at 5000rpm for 60 seconds and then baked for two hours at 180°C

to drive off the solvent. Before milling, the sample was imaged a number of times thus exposing the HSQ, it was noticed that the sample was charging badly so the charge neutraliser was used for the initial milling steps. The sample was milled in the standard way with a size of 20μm allowing for two gates to be included in the sample. The samples were thinned to a thickness of around 100nm.

This was there first time that InP samples like this had been lifted out, there were concerns over the fragility of the InP and whether the sample would be strong enough to withstand the process. It was found that there was no problem in the transfer to the TEM grid. The sample was loaded into the TEM, however when the beam was switched on and focused on the sample a violent reaction happened, and much of the sample was damaged. After the reaction the sample appeared to stable, no other damage was witnessed during imaging. It is unclear exactly what caused this reaction although it is thought that the affected area was very thin, much thinner than at the area of interest, this is often observed in samples that have been milled in this way.



(a) (b)

Figure 7.9 TEM images of InP HEMT sample prepared using a HSQ protection layer (a) Low magnification image showing the extent of damage to InP substrate. (b) Higher magnification of gate region showing the flow of HSQ under the gate head.

Despite this damage it was possible to see the effect of the HSQ protection layer. Figure 7.9 (a)&(b) shows two of the TEM images of the sample, Figure 7.9(a) is taken at a relatively low

magnification (x5k), this shows the extent of the damage (lower right corner). The HSQ protection layer is intact, this is the light grey region layer surrounding the gate, on top of this is the darker platinum layer. Figure 7.9(b) shows a higher magnification image (x38k) of the gate. This shows that the HSQ has flowed underneath the gate region preventing the air-gap below the gate and preventing the damage to the substrate. Due to the poor quality of the sample it was not possible to obtain any more information on the devices.

7.4 Conclusions

This chapter has described the development of a process capable of preparing cross sectional TEM samples from HEMT device samples using FIB milling. This technique of sample preparation allows TEM to be used in the analysis of HEMT devices. TEM is a very powerful analytical tool, capable of much more than the basic things that have been asked of it so far. In this chapter the high resolution of which TEM is capable has only been touched upon. With experience of sample preparation, it will be possible to look at the epi-layer structure on an atomic level to study the roughness of the channel-spacer interface as well as a study of the diffusion process that is present in the non-annealed ohmic contact. These are just two examples of further applications of this technology there will be many others that TEM analysis will be able to shed light on in the future.

7.5 References.

[7.1] FIB User's web page.

http://users.ox.ac.uk/~rml/index.htm

[7.2] Communications. with Xin Cao.

[7.3] Siew et al., "*Thermal Curing of Hydrogen Silsesquioxane*" Journal of Electrochemical Society, **147**, 2000, p335.

[7.4] H. Namatsu et al. "*Nano-patterning of Hydrogen Silsesquioxane Resist with Reduced Linewidth Fluctuations*", Microelectronic Engineering, **41/42,** 1998, p331

8.0 Development of a Sub-50nm T-gate HEMT process

The advantages of gate length scaling were introduced in section 6.1. It was seen that the reduction of the gate length of the HEMT reduces the electron transit time through the device as well as reducing the intrinsic capacitances associated with the gate. Both of these effects enhance the transistor performance. This was demonstrated with the realisation of a HEMT with a 50nm T-gate fabricated on a metamorphic GaAs substrate, which showed excellent RF performance and reduced gate capacitance.

It was recognised from the outset that there would be a limit to the gate length that could be achieved using a copolymer/PMMA resist system. Therefore a further technology was developed in conjunction with the sub-100nm copolymer process. This technology would be designed to produce devices with a gate length below 50nm. This chapter describes the work carried out in developing this process.

8.1 Proposed T-gate process

The aim of this gate process was to produce the T-gate with as small a gate length as possible. Unlike the copolymer process described in Chapter 6 the yield and reproducibility of the process was not the major concern. The initial task was to decide what materials would be used to fabricate these devices and how these materials would be used. The first task was to choose which resist system would be used.

8.1.1 UVIII/PMMA Bilayer

It was decided that the process would be based on the PMMA/UVIII 70nm T-gate process that had been developed recently in Glasgow [8.1]. This process uses a PMMA base layer to define the gate foot and a UVIII layer to define the larger gate head. This process was chosen over other resist systems such as copolymer/PMMA and UVIII/ZEP for a number of reasons. Compared to the PMMA/UVIII system, the PMMA/copolymer stack suffers from a poor sensitivity ratio. This results in a much more complicated dose (see section 6.2.1). The choice of using PMMA rather than ZEP as the base layer was not so straightforward. The sensitivity ratios of ZEP/UVIII and PMMA/UVIII are 5-10[8.2] and 12.5 respectively [8.3] effectively allowing the head and the foot of the gate to be considered separately. Both PMMA and ZEP are capable of high resolution lithography, ZEP 520 was used as the base layer of the 30nm T-gate realised on a planar substrate [8.4]. Furthermore, ZEP520 in addition has good contrast and excellent dry-etch capabilities. However removal of ZEP is extremely difficult, requiring an overnight soak in dedicated remover, making the lift-off of the delicate gate difficult. PMMA does not suffer from this problem it is quickly removed in acetone. As a result, the PMMA/UVIII gate stack was chosen as the resist strategy for this work. In order to prevent cross linking of the PMMA and UVIII layer a separation layer of Lift-Off Resist (LOR) from Microchem was included.

8.1.2 SiO_2 Gate Support Dielectric.

As the length of a T-gate is reduced the physical stability becomes a problem. The important transistor figure of merit, f_{max}, is strongly dependent on the resistance of the gate, to maintain the low gate resistance as the gate length is scaled the head of the gate must increase, or, at the very least, stay the same size. This leads to a large gate head (300-500nm wide) supported by a gate foot that is very small. This can lead to poor gate yield during lift-off. A number of papers describing the fabrication of ultra small gate length HEMTs have used a dielectric layer such as SiO_2[8.5,8.6] SiN_x [8.4] to support the gate foot.

There are other benefits to using the dielectric definition layer. It has been reported that a SiN_x layer prevents distortion of the resist mould during metallisation and therefore provides a better metallised profile.[8.4] It is also possible by good control of the dry etching to reduces the gate foot below the lithographic size.

It was decided to use SiO_2 as the gate support layer, this was because it was originally planned to use low-damage sputtered oxide which gives good control of the thickness. Further, SiO_2 has a lower dielectric constant than SiN_x, leading to reduced parasitic gate capacitance.

8.1.3 Lithography

The most important consideration in performing lithography below 50nm is the forward scattering of electrons as they pass through the resist. The first step taken to reduce this scattering was to write the gates at 100keV, it is seen from Figure 4.5 that the forward scattering experienced by the electrons is greatly reduced by using 100keV instead of 50keV. Figure 4.5 also shows that for a beam energy of 100keV, the thickness of resist greatly influences the scattering of the electrons although the effect shown is exaggerated (as the electrons are passing through up to 1μm of PMMA) the behaviour is the same. As the electrons pass through the UVIII and LOR layers they will scatter away from the beam axis increasing the effective beam

size and so increasing the size of the developed gate foot. In order to prevent this effect a two-stage lithography process was proposed.

This two-step process, labelled "Bi-lithography" used the fact that the standard UVIII/LOR/PMMA process used two separate developers. The UVIII and LOR layers are developed in an alkali based developer CD26. The PMMA layer is then developed in o-xlyene. The use of two separate developers makes this process an ideal base for the development of a two-step lithography process. The large gate head can be written first at 50keV, developed and the sample placed back in the beamwriter for the foot to be written at 100keV. This eliminates the scattering through the thick UVIII and LOR layers and reduces the lithography challenge to one of writing small lines in 70nm of PMMA. There is of course the issue of the alignment of the gate head to the gate foot, for the standard device structure (i.e. not self-aligned) however, by adopting suitable writing strategies, the excellent overlay accuracy of the e-beam tool can be exploited resulting in alignment of better than150nm without major problems.

8.1.4 Fabrication Plan

Using the materials and techniques introduced above, a process was devised that should be capable of realising sub-50nm T-gates. This relatively complex, six stage process flow is illustrated in Figure 8.1.

Figure 8.1 (a) shows the layers that are used to define the gate, the upper layer is UVIII with a thickness of around 350nm below this is a thin layer of LOR (<100nm). LOR is a lift-off resist designed to create large undercuts to aid metal lift off, in this application it is being used merely to separate the UVIII and PMMA layers.

The initial exposure uses a very small dose, typically 80$\mu C/cm^2$, to define the wide gate head and the feed layer in the UVIII layer. The sample is then developed in CD26 developer, the standard developer for UVIII which also etches the "separator" LOR layer. To ensure that all the LOR is removed the sample is slightly over-developed leading to an undercut between the UVIII and PMMA, the result is shown schematically in Figure 8.1(b).

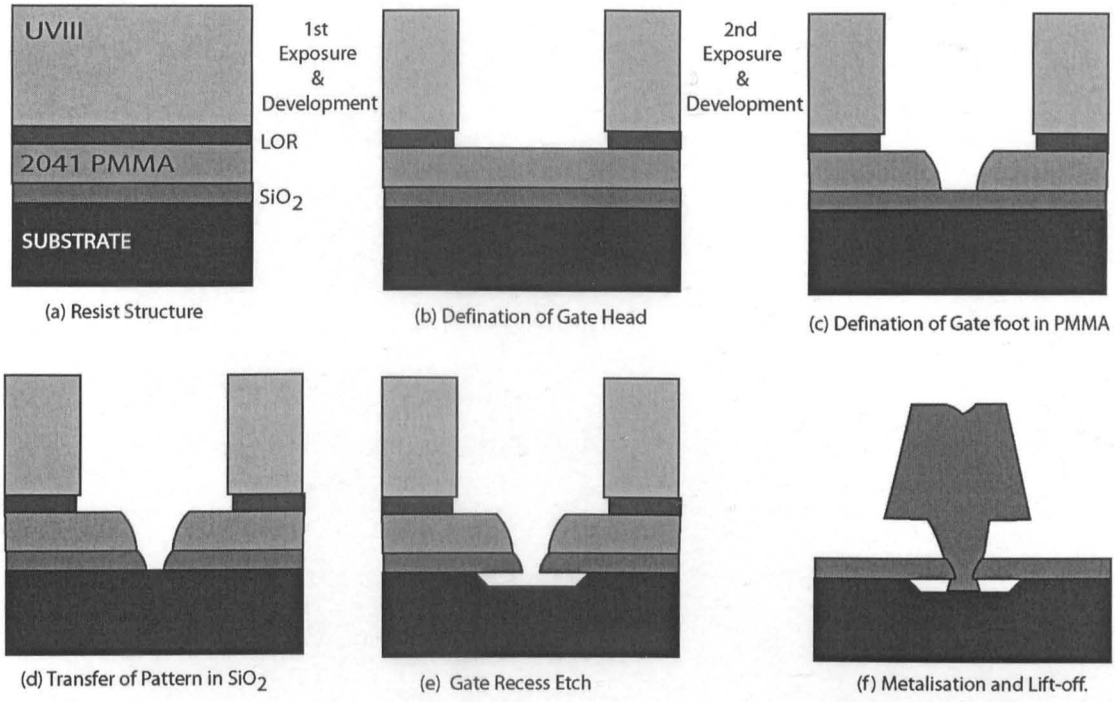


Figure 8.1 Illustration of proposed process for the fabrication of sub-50nm T-gate devices.

The sample is then placed back in the Beamwriter for the definition of the gate foot, achieved by writing a 10nm central line with a very large dose and defining the gate feeds in the PMMA layer using the same pattern as for the UVIII layer but using a higher dose. The PMMA layer is then developed in o-xlyene for 60 seconds at a temperature of 23°C. The result is shown schematically in Figure 8.1 (c). The gate foot pattern is then transferred into the SiO_2 layer by etching, (Figure 8.1 (d)). The wet chemical gate recess is then performed before gate metallisation. (Figure 8.1(e)) The final structure is shown schematically in Figure 8.1(f), illustrating the purpose of the support layer, although the gate foot is very small the gate structure is robust.

8.2 Development of processes for the fabrication of T-gates

The proposed sub-50nm T-gate process flow contained a number of new processes that required development and their compatibility with the fabrication of the HEMTs assessed. This section will deal with the development of these key processes, namely dielectric deposition, pattern transfer into the dielectric and the lithography issues.

8.2.3 Dielectric Deposition

The chosen dielectric was SiO_2 deposited by PECVD, with a thickness of 25nm. The thickness was chosen as a comprise; it is important that the support layer is strong enough to support the gate and not deform during metallisation, but as this layer must be etched and metal pass through the mask onto the substrate the oxide cannot be too thick or the gap will close during metallisation results in gate breakage, an example of this is shown in Figure 8.2. The choice of 25nm for the thickness gives a 1:1 aspect ratio with the desired gate length.

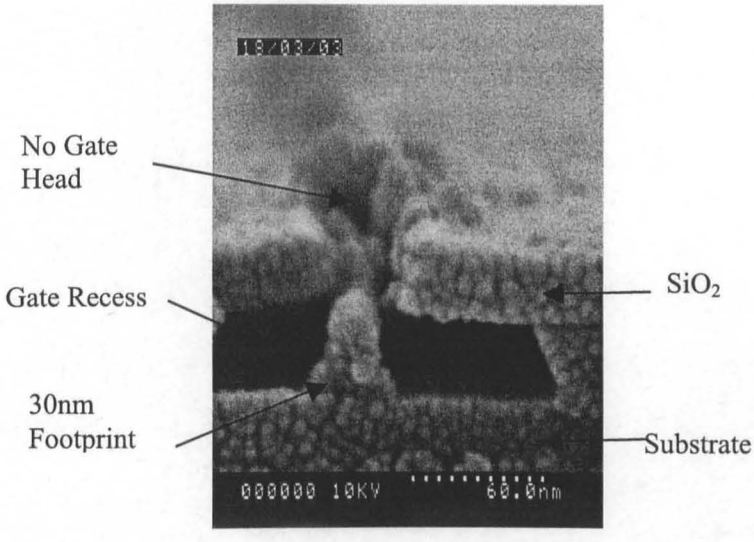


Figure 8.2 SEM image illustrating the problem of filling a etched gap in oxide with metal.

Deposition of such a thin layer of oxide was not a standard process in Glasgow, the average thickness normally deposited is around 300nm. Because of this, the deposition rate is much faster than would be used for depositing a thinner layer. Over a number of months it was found

that using the standard deposition process, targeting a 20nm layer thickness produced an oxide thickness film of 25±2nm. This systematic difference between the requested and measured depth of the oxide was thought to be due to an over estimation of the induction time before deposition starts.

There is of course no point in developing a process for the fabrication of ultra-small gate lengths if the performance benefits of the scaled devices are negated by any damage caused by the fabrication process. To monitor this, tests were performed to determine the transport characteristics of the device quality material before and after oxide deposition. Van der Pauw samples were fabricated on wafer A1774, a lattice matched InP device wafer with a barrier thickness of 10nm a spacer thickness of 4nm and a cap layer of 20nm, suitable for the fabrication of 70nm gate length devices. Although this is much larger than the proposed devices it will provide information on how the oxide deposition affects shallow epi-layer structures. Subsequently 25nm SiO_2 using PECVD was deposited, the oxide was removed from the pads using HF but left on the active region. The mobility, μ, and carrier concentration, n_{sh} were then measured at room temperature. The remaining oxide was then removed by rinsing the whole sample again in HF. The results of these experiment are summarised in Table 8.1.

Before Deposition		With Oxide		After Removal	
μ (cm^2/Vs^{-1})	n_{sh} (cm^{-2})	μ (cm^2/Vs^{-1})	n_{sh} (cm^{-2})	μ (cm^2/Vs^{-1})	n_{sh} (cm^{-2})
4832	$4.11x10^{12}$	4703	$4.41x10^{12}$	4894	$4.12x10^{12}$

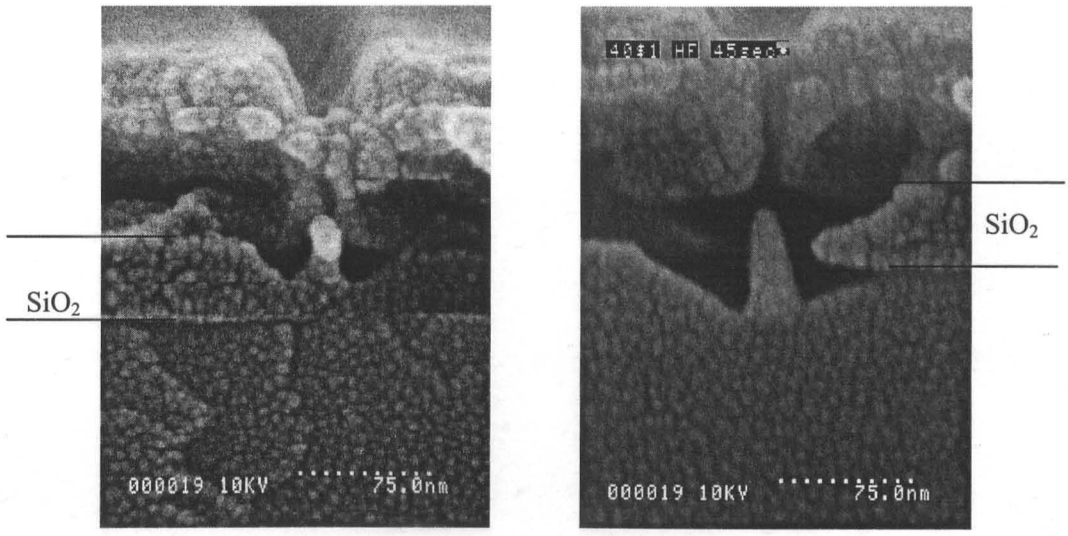
Table 8.1 Summary of effect of Oxide Deposition on electron transport parameters.

Table 8.1 shows that the oxide deposition does not damage the transport characteristics of the sample, it is noted that the mobility falls and the sheet concentration increases when the oxide is deposited but this is likely to be caused by a change in the surface condition due to the oxide/semiconductor interface which modifies the depletion region depth in the cap layer. This

is supported by the fact that both mobility and sheet carrier concentration return to their original values upon removal of the oxide. It was concluded that this method of deposition was compatible with HEMT fabrication.

8.2.4 Pattern Transfer in SiO_2 layer

The next process to be developed was that of transferring the pattern from the PMMA layer into to the oxide layer. Initially it was decided to use a wet etch process to perform this procedure, as a wet etch would have the benefit of being quicker to develop and easier to control, whereas the dry etch process is complex, very sensitive to many parameters such as etch gas composition, flow rate chamber pressure, RF power, DC bias [8.7] and particularly sensitive to the developed surface conditions. The wet etch investigated was using a dilute HF solution. As HF is known to etch SiO_2 very quickly, for example the removal of 25nm of oxide in the previous section was carried out with a 20s etch using 10:1 H_2O:HF, it was clear that the etch had to be made more controllable by dilution. This was achieved by mixing 10ml of 10:1 HF with 30ml of H_20. An array of gates was written on a sample covered with 25nm SiO_2 in 70nm of PMMA with a linewidth of around 30nm. The pattern from the PMMA layer was transferred into the oxide by etching the sample for 30,45, and 60 seconds before a non-selective etch of the GaAs substrate and deposition of 15:15:50 Ti:Pd:Au. The SEM cross sections of the 30 and 60 seconds samples are shown in Figure 8.3



(a) (b)

Figure 8.3 SEM image of tests of HF pattern transfer etch showing lateral creep of etch (a) 30 seconds etch, the etch has not passed though the oxide layer (b) 60 seconds etch, the etched has passed through the oxide layer however the lateral etch is to large.

Figure 8.3(a) shows the sample that had been etched for 30 seconds, the dilute HF has etched ~10nm into the oxide layer oxide layer with an etch width was around 60nm. By increasing the etch time to 60 seconds the oxide layer is fully etched (Figure 8.3(b)) however the lateral etch is even larger. This has the effect that the gate metal passes through the PMMA mask and straight through the oxide without any effect; the oxide layer serves no purpose, neither as a support nor as a mask.

From this it was seen that an isotropic etch was needed, meaning a dry etch process would have to be developed. This development work was carried out with the aid of Hai-ping Zhou the dry etch process engineer in Glasgow. The first step was to develop an etch that was capable of etching SiO_2. This etch was based on a process previously used to etch thin layers of SiN_x for HEMT fabrication. The initial etch conditions are summarised in Table 8.2.

Gas	CHF_3
Flow (sccm)	30
RF Power (W)	30
DC Bias (V)	65
Pressure (mT)	16

Table 8.2 Summary of Etch Conditions

Using these conditions, it was found that the oxide layer was not etched. This was thought to be due to the low RF power chosen to minimise any substrate damage. Tests were therefore carried out to find the optimum power that would etch the oxide layer whilst not damaging the PMMA mask layer or the underlying substrate. To assess the impact of the etch process on the PMMA resist mask, a test pattern of lines in 2.5% 2041PMMA with a width of around 30nm was defined. These samples were then etched using different RF power levels, keeping all of the other etch parameters constant. It was found that 80W provided a good etch profile while not excessively damaging the PMMA layer. Figure 8.4 shows the resultant etch. The profile of the oxide etch is good with a nearly vertical profile and a depth of 30nm. Due to a thicker than expected oxide thickness the etch has not reached the surface of the GaAs substrate, however with the correct oxide thickness this would be achieved. The second feature to note is the geometry the PMMA mask. The etch has thinned the resist by approximately 20nm however, the PMMA is still intact. The etch has produced a triangular profile in the PMMA, which is ideal as the wider profile at the top of the resist layer will aid the metallisation of the gate and add strength to the structure.

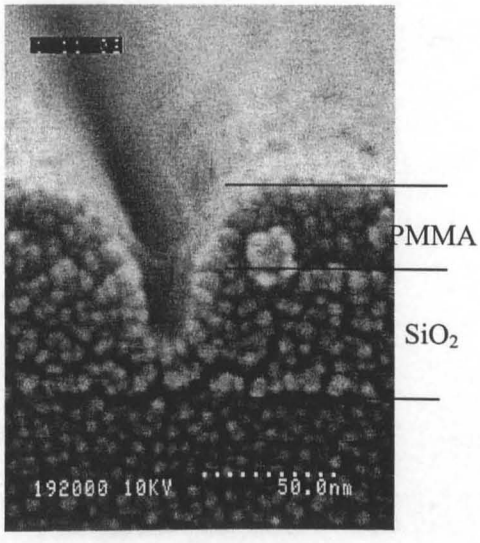


Figure 8.4 SEM cross section of dry etch test, depth of etch is 30nm the oxide thickness is too large, around 50nm.

The next stage of the development of the dry etch process was to assess the damage that the etch caused to the transport properties of the material. Van der Pauw structures were fabricated on wafer A1774 using optical lithography. A 25nm thick layer of SiO_2 was deposited using

PECVD as described in section 8.2.3, a further stage of optical lithography defined a resist mask which protected the sample surface with the exception of the centre of the Van der Pauw pattern, the sample was then dry etched using the above process parameters with a RF power of 80W. Thereafter, the mobility and the carrier concentration was measured. It was seen that the etch did affect the transport properties of the material, the mobility of the material increased from 4997 to 6050cm^2/Vs^{-1} and the carrier concentration decreased from $3.99x10^{12}cm^{-2}$ to $2.1x10^{12}cm^{-2}$. This suggests that there has been some damage to the material. It was suspected that that was due to damage to the cap layer reducing the active doping concentration. The reduction in carrier concentration and increase in the mobility resulting from a reduced contribution of the cap layer to parallel conduction in the material. The increase in mobility suggests the transport properties of the device channel have not been adversely effected however. In addition, the geometry of the Van der Pauw sample is very different from that of the devices. The active area of the Van der Pauw structure was approximately 1x1mm, compared with the actual device, which has the dimensions of 25nmx50μm. The much larger area of the Van der Pauw test structure means that the sample would etch more quickly and thereby leaving the surface unprotected for a longer period. It was decided that the damage caused to the actual device would be much smaller than suggested from this Van de Pauw sample. As a consequence of the above, it was concluded that the CHF_3 oxide etch process was compatible with a full HEMT process flow.

A process capable of depositing a very thin layer of SiO_2 through which the high resolution gate pattern could be transferred by means of dry-etching had been developed. The next stage of the process was to develop the lithography capable of producing a resist profile that could be transferred in to oxide layer and metallised to produce a T-gate with a gate length of less than 50nm. This process is discussed in the next section.

8.2.5 Lithography

The basis for the lithography development of sub-50nm T-gates was the 70nm T-gate process using a UVIII/LOR/PMMA tri-layer. As discussed section 8.1.1 this was used because of the large difference in sensitivities of the UVII and PMMA. The processing is quite different to that used for the copolymer/PMMA process described in section 6.2.1. This process shall be quickly discussed below.

70nm UVIII T-gate Process

The foot of the T-gate is defined in a 90-100nm thick layer of 2041 PMMA, achieved by spinning 2.5% concentration of 2041PMMA for 60s seconds at 2000rpm. The sample is then baked for 1 hour. The LOR layer is applied by spinning a 1:4 concentration of LOR and cyclopentanone at 5000rpm, then baked for 15 minutes to drive off the solvent. The UVIII layer is then spun, using a 80% solution of UVIII diluted using ethyl-lactate spun at 5000rpm giving a thickness of 300nm. The sample is then baked on a hotplate at a temperature of 120°C for 60 seconds. The sample is then ready for e-beam lithography. At this stage the UVIII layer is light sensitive and must be kept covered.

After exposure a Post Exposure Bake is performed, on a 120°C hotplate for 90 seconds. As was discussed in section 4.3.2 UVIII is a chemically amplified resist. This post bake is requires to start the reaction that allows the UVIII to develop. In the two step development process the sample is first developed for 60 seconds at room temperature in CD26 to develop the UVIII and to etch the LOR layer. The PMMA layer is then developed in o-xlyene at 23°C for 60 seconds, after each development step the sample is thoroughly rinsed in water.

A number of modifications were made to this process to make it suitable for fabricating smaller gate lengths. The first step was to reduce the thickness of the PMMA layer from around 100nm to 70nm, this was achieved by increasing the spin speed from 2000 to 5000rpm. This improves

the achievable gate length as well as reducing the aspect ratio of the gate foot aiding metallisation. The thickness of the UVIII layer was also reduced, by reducing the concentration of from 80% to 58% while maintaining the same spin speed.

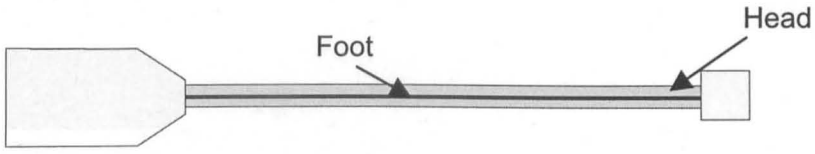


Figure 8.5 Design of the sub-50nm gate. (Not to scale) The central line has a thickness of 10nm, the rectangle defining the gate head has a width to 300nm.

The design of the gate shown schematically in Figure 8.5 was modified from that used for the 70nm process. The gate consists of two rectangles, each written with a different dose. The wide rectangle has a width of 300nm and defines the head of the gate, the foot of the gate is defined by the thin central line. In the 70nm process a central line of width 20nm is used. To reduce the gate length this was modified to 10nm. The gate feeds are written on a third layer, this includes a 1μm square block at the end of the gate, which serves as an anchor, making good contact to the substrate and preventing the gate peeling off during lift-off. The anchor block is offset by 250nm to reduce the probability of the shorts between it and the mesa caused by misalignment of the isolation level.

It was originally intended that the sub-50nm T-gate process would be based on the two step "bi-lithography" process. However, the requirement to align the gate foot with the previously exposed and developed head, necessitated that the test samples required alignment markers to enable registration. This, coupled with the two-stage lithography process meant that for each test piece three levels of lithography were required making the process development cylce excessively long. Work was therefore carried out using the same resist structure but using only one lithography step.

The development of the sub-50nm was carried out in a similar way to the sub-100nm process discussed in chapter six. Dose tests were carried out, however because of the large difference in

the sensitivities of the resists used the two dose regions could be considered to be independent. The relative doses were defined in a ccfa in the same manner as the 50nm copolymer process. However since the doses were essentially independent the file was written in such way that a large dose range of the central line could be tested while having a relatively constant dose to the head region. An example of this of the file is given below.

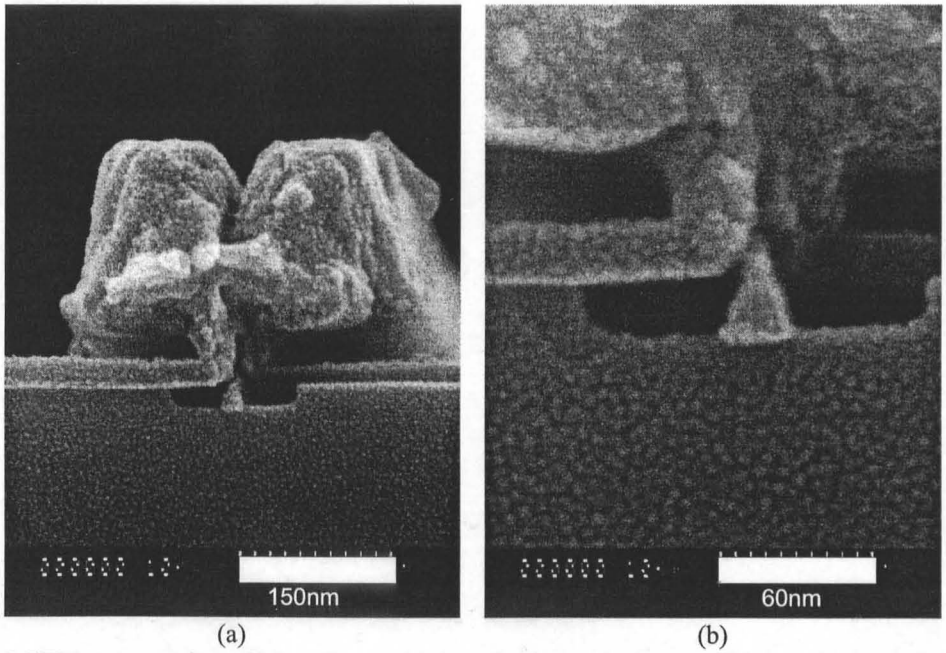
```
sort by ccfa
1.1     !Dose 0, dose of 110 for layer 6 (Head)
29      !Dose 1, dose of the gate foot.
layer 6
0
layer 31
1
```

The dose for the head is the base dose multiplied by 1.1, the central dose is multiplied by 29, this means that for a typical base dose range of 80-95$\mu C/cm^{-2}$ range of the central dose ranges from 2320 to 2755$\mu C/cm^{-2}$. The feed layer is fractured separately and written using a much a much larger spot, 160nm, compared to 12nm for the gate region to minimise writing time.

The doses tests were originally carried out by cleaving and sputtering the developed resist profiles before examination in the S900 SEM however this was later changed to a more accurate but time consuming preparation. The test structures were written on GaAs pHEMT material with a 30nm GaAs cap which had 25nm of SiO_2 deposited. After development the pattern was transferred into the oxide by dry etch. A gate recess was then formed before the gates were metallised and lifted off. This sample preparation was needed for a number of reasons; using the SEM to view such small lines in resist can alter the resist profile that is being observed. It is therefore difficult to determine the resist profile caused by the exposure and development as opposed to the observation. The second reason is that it had been noticed that the patterns which had looked to be developed by SEM inspection were not being etched into the SiO_2. It had been noted that the dry-etch process was very sensitive to resist scum on the sample surface, to eliminate this a 40 seconds oxygen ash was included after development. The final reason for this elaborate sample preparation is that the aim is to develop a metallised gate with as small a gate length as possible, the etching of the oxide and the metallisation will affect the final gate

length achieved. There is little point in a 25nm wide line defined in the resist if it is not possible to transfer this pattern in to oxide and then evaporate metal to define this foot width in metal.

Dose tests were carried out using the method outlined above. A dose of 2500μC/cm^{-2} repeatably produced metallised gates with a gate length of 35nm. Figure 8.6 shows the cross section of the resultant T-gate. Figure 8.6(a) shows that the gate head is well defined, the stalk of the gate is around 40nm in height, less than that of the 50nm and 120nm process but large enough to separate the gate from the highly doped cap layer, reducing the parasitic capacitance. The gate foot region is shown with a higher magnification (x500k) in Figure 8.6(b), this allows the gate length to be determined with more confidence, the foot width of this device is 33nm. It can also be seen that the thickness of the oxide layer is around 25-30nm as required. The recess offset achieved with a 25 second selective succinic acid recess etch is too large, at 53nm. This etch time was reduced in the final device sample to 20 seconds.



(a) (b)

Figure 8.6 SEM cross section of 33nm T-gate fabricated using a single stage lithography process and a SiO_2 gate definition layer

Figure 8.6 shows a single stage lithography process is capable of producing 35nm T-gates on a planar substrate. However a gate process serves no purpose if it cannot be incorporated into a HEMT device, it is obviously a requirement to realise such a structure between source and drain

ohmic contacts and to ensure that the gate is continuous along the whole width of the gate. The next section will discuss the incorporation of this process into the fabrication of a HEMT.

8.3 Process Flow of sub-50nm T-gate HEMTs

The previous section showed that there are a number of processes required to the fabrication of sub-50nm T-gates that are not in the standard fabrication process of HEMTs. The most obvious is the incorporation of a thin SiO_2 layer. The immediate question is at what stage should this be layer deposited? It could be deposited after the ohmic contacts had been metallised or after mesa isolation. It was decided that the oxide should be deposited at the very start of the process flow, this was for a number of reasons; if the oxide layer is deposited at the very start of the fabrication process i.e. immediately after the wafer has been grown it will act as a encapsulation layer and will prevent the cap layer from oxidising. It is also important that the oxide layer is uniform, if it is deposited after the ohmic contacts have been formed the thickness may vary within the source-drain gap causing problems in the etching and lithography steps in addition, the SiO_2 would have to be removed from the ohmic contacts to provide a good electrical contact with the bondpad level.

Depositing the oxide layer at the start of the process of course means that it has to be removed from the surface before the isolation and ohmic contact stages can be performed.

8.3.1 Selective Removal of SiO_2 layer

Section 8.2.2 showed that hydrofluoric acid etches SiO_2 very quickly; it was therefore an obvious choice for oxide removal prior to device isolation and contact defination. However it was also seen that the etching of oxide in HF is highly anisotropic. In a thin oxide layer this leads to the HF undercutting the edges of the resist mask. The optimum etch time must remove the oxide but not massively undercut the oxide. This time was found by experiment, using a

TLM test structure. This is an ideal testbed as the narrow gaps of the TLM structure closely resemble the source-drain geometry of a HEMT. In addition, the contact resistance can be measured to ensure that there are no remnants of oxide that could hinder ohmic contact formation. The samples were etched in 10:1 HF for 20 seconds before the evaporation of a Ni:Ge:Au ohmic contact and then were studied in the S4700 SEM which provides high resolution with the benefit of taking large wafer pieces.

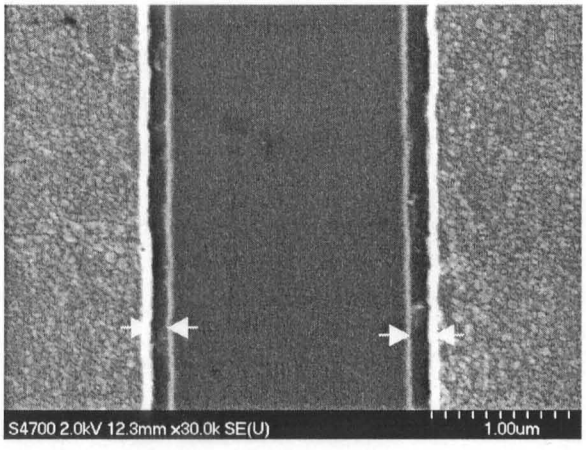


Figure 8.7 SEM image of the undercutting of the PMMA by the HF SiO_2 etch. The width of the undercut is 150nm.

The undercutting of the oxide can be seen in Figure 8.7 (highlighted by white arrows), the length of the undercut was 150nm. This was considered too large so the etch time was reduced to 15 seconds. The contact resistance measured was identical to that measured for the standard process. This etch was included in addition to the de-oxidation step before the isolation etch and ohmic contact metallisation. The full process for the fabrication of HEMTs of sub-50nm HEMTs using a SiO_2 layer is given in Appendix F.

8.4 Device Fabrication

Devices were fabricated on wafer A1602, this is a lattice matched InP HEMT. This wafer has a 20nm $In_{0.53}Ga_{0.47}As$ $1x10^{19}$ cm^{-3} doped cap layer.The thin cap is ideal for the sub-50nm T-gate process, the distance between the SiO_2 mask and the etched surface is reduced from that used for

tests which were performed with 30nm cap layers. This reduction reduces the "flaring" of the gate metal and should result in smaller footprint sizes. The barrier layer thickness is 7.5nm with a 4nm spacer layer, this gives a gate to channel separation of 11.5nm making the material suitable for short gate length devices. However, the very thin barrier layer thickness will mean that the gate offset must be kept small to avoid depletion of the channel and the devices operating in enhancement mode. This behaviour was seen in 50nm devices fabricated on A1881, which had a very similar layer structure.

The wafers were scribed into quarters cleaned and de-oxidised in $HCl:H_2O$ before deposition of 25nm of $Si0_2$. At this stage the thickness of the oxide layer was tested by simply scratching the film and measuring the thickness using AFM, which was found to be 27nm. The marker layer was then written and metallised with the use of the HF etch prior to de-oxidisation to ensure good adhesion. The ohmic contacts were then defined, the oxide removed as above and metallised with Ni:Ge:Au. The samples were annealed at a temperature of 240/280°C for 30/60 seconds. The samples were then isolated electrically using the non-selective orthophosphoric acid/hydrogen peroxide/water etch until the current dropped to less than 100nA at 2V in the dark in the isolation test structure described in section 4.4. At this stage the contact resistance and transport characteristics were measured. The contact resistance was determined to be 0.06Ω.mm, the mobility was found to be $5767cm^2/Vs^{-1}$ and the carrier concentration was $1.13x10^{13}$ cm^{-2}. All of these parameters, particularly the combination of the high mobility and sheet concentration suggest that the material is well suited to for the fabrication of high-speed transistors.

Gates were written using the one stage lithography process. The dose of the central line was $2800\mu C/cm^{-2}$, larger than used for the planar substrate; which would give a gate length of 40nm and ensure better uniformity. The dose of the head and the feed were 80 and $600\mu C/cm^{-2}$ respectively. It was found that the gate feeds of the devices were overexposed, leading to the

two feeds joining together in the UVIII, (Figure 8.8) this would cause problem during lift-off and lead to increased capacitance.

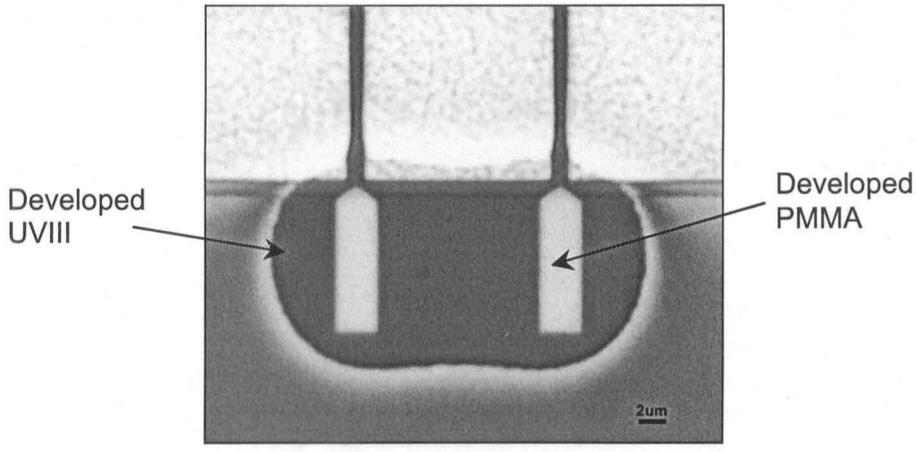
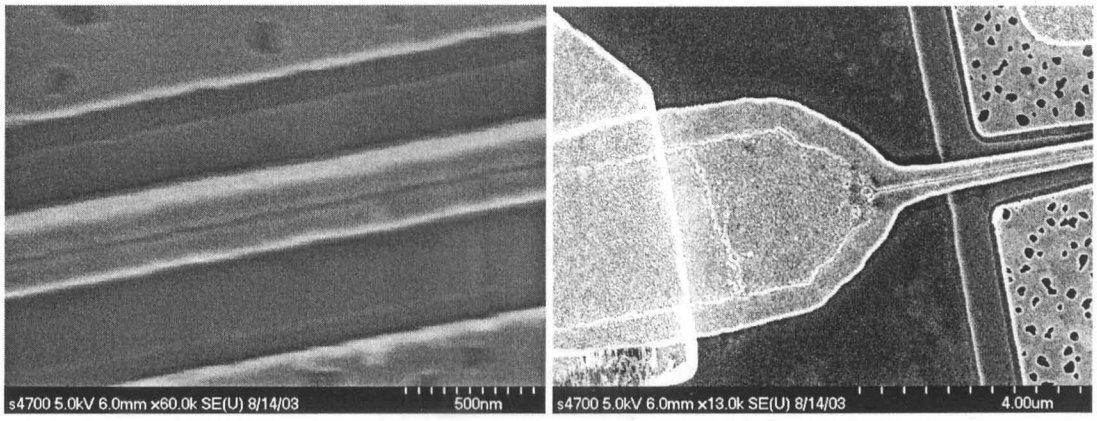


Figure 8.8 Optical Micrograph of overexposed gate feeds during the fabrication of sub-50nm devices on A1604.

A dose test resolved this problem. On reworking with modified gate feed exposure, the mechanical yield of the devices was very high, attributed to the gate support layer. However on completion of the devices it was found that the devices suffered from poor gate control indicating that the gates were not in contact with the substrate across the full width of the device. The samples were studied using the S4700 SEM, however due to the high gate yield, it was not possible to see the gate foot region as it was shielded by both the gate head and the oxide layer. Figure 8.9 shows SEM images of the gates of devices in this batch, it is not possible to see the gate foot since the head is still intact, however the gate is present and well defined with the etching of the oxide layer around the ohmic contacts visible. Figure 8.9(b) also shows the overhang of the gate feed metal caused by the large differences in the resist sensitivities when exposed simultaneously.



(a) (b)

Figure 8.8 SEM image of fabricated sub-50nm HEMTs using SiO_2. (a) Image of centre of gate, the gates looks continuous including the indentation of the stalk. The SiO_2 etching is well controlled with a undercut of around 75nm. (b) Low Magnification showing the feed region, the overexposure of the feed caused by the sensitivity difference is clearly seen.

Further attempts were made to thin the UVIII and LOR layers by increasing spin speed, this however did not work, the gates were still discontinuous. During the AFM study on resist thickness detailed in section 6.6 it was found that thickness of the PMMA layer remained approximately the same within and outside a source drain gap. (The two processes use the same thickness of 2041PMMA) The extra resist thickness in the source-drain gap is therefore almost wholly caused by the LOR and UVIII layers, removing these layers would reduce the problem to writing high resolution lines in PMMA. As described in section 8.1.4 a process had already been developed to do this, "bi-lithography". The use of two stage lithography had additional benefits, without the scattering from the UVIII layer the lithographic resolution would be improved, the problem with overexposure of the gate feeds would be eliminated, it would be possible to write the UVIII layer at the clearing dose, develop and then pattern the PMMA layer at a much higher dose.

A dose test was carried out on planar substrate. A epi-layer test piece was spun with 2.5% 2041 at 3000rpm, to create a layer thickness of 70nm. A array of gate patterns was written and developed in o-xylene for 60 second before being ashed, etched and metallised with 15nm Ti:15nm Pt :15nm Au and lifted off. The lowest dose found that gave the continuous gate along the 50μm gates was 2170μC/cm^{-2}, it should also be noted that this should be reduced slightly for

the writing of the gate as the resist will already had the dose from the previous lithography stage. It is important for this two-stage process that the alignment of the two levels is better than 200nm and ideally less than this. The central alignment mark was used to define the foot to provide the best focus point and therefore the smallest spot size for the writing of the small features. (See section 6.4) By using the focus marker for both the layers any difference in the height caused by mounting or tilt would be negated and so the field size of the patterns written at separate times would be the same. This would improve the alignment of the two layers.

Devices were fabricated using this method, the physical yield of theses devices was high with over 75% devices intact. The use of a "bi-lithography" greatly reduced the over development of the gate feeds as shown the optical micrograph in Figure 8.10.

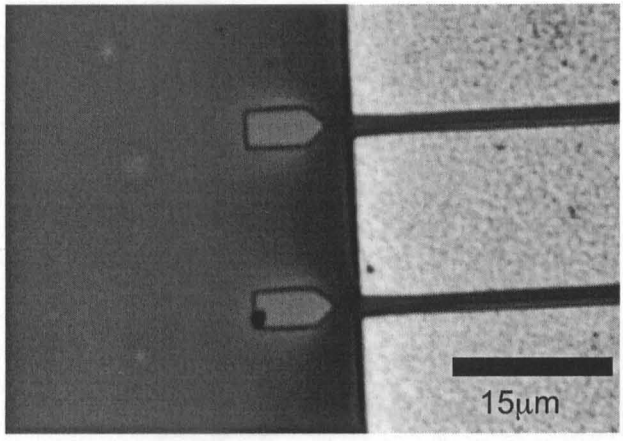


Figure 8.10 Optical Micrograph of gate feeds written using the two-stage lithography process. The overexposure of the UVIII has been eliminated.

The devices were completed with the addition of CPW feed lines and then measured at DC. The devices did not demonstrate a saturating behaviour that is expected from a transistor, suggesting that, for some reason the gate recess etch had not been successful. The samples were cleaved to produce cross sectional samples for viewing in the S900. Inspection of these devices showed that although the gate foot was developed the oxide had not been fully etched and so no etching of the cap was visible. In areas where the oxide had etched, the recess etch had been successful and the gate foot successfully metallised. Figure 8.11(a) shows the profile of the whole gate at a

relatively low magnification (x180k). This shows a number of things; the stalk of the gate is still relatively tall, with a height of around 40nm, much less than 110nm for the 120nm gate process but it is enough to separate the large gate head from the cap layer. It can be seen that excellent alignment that is achieved using the two-stage lithography using central focus marker. (Estimated to be offset by around 20nm to the right). This leads to a symmetric gate and a strong structure. At the base of the gate stalk it is possible to see the silica layer, the recess etch, and the gate foot. This is shown at higher magnification in Figure 8.11(b), the magnification used was x600k and the scale bar in the bottom right corner reads 50nm. It is difficult to determine the exact size of the gate foot as the gate recess etch is small, however the gate foot is seen as the grey area which has a length of 25nm.

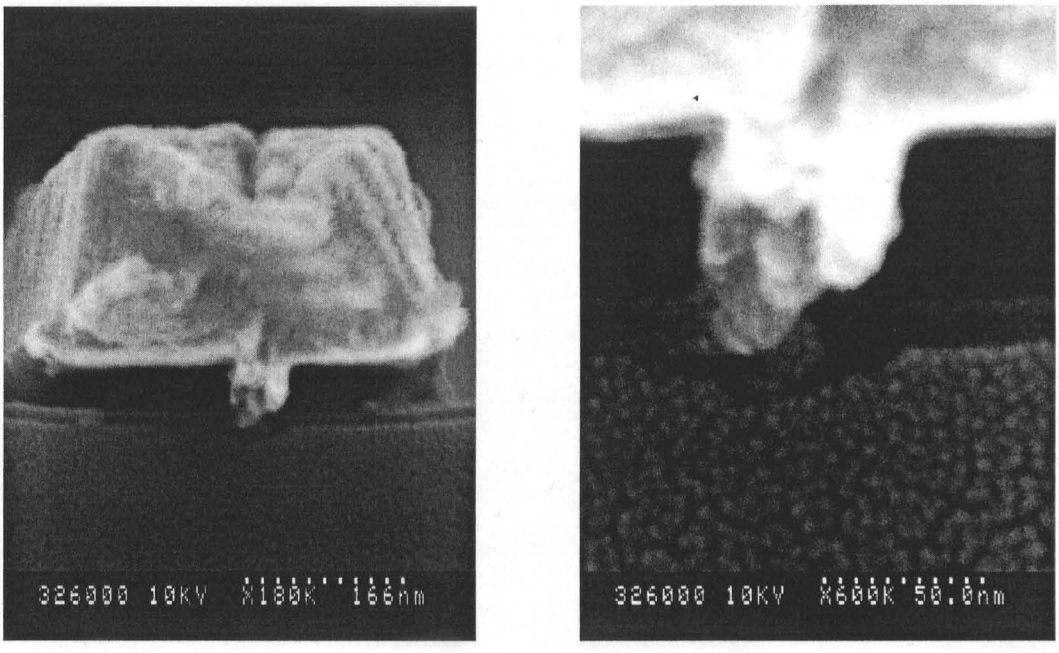


Figure 8.11 SEM Micrographs of 25nm T-gate fabricated using the Bi-lithography process.

This demonstration of the fabrication of a HEMT with a 25nm T-gate is encouraging. From the examination of the HEMT cross sections of a number of devices across the wafer, it was concluded that this sample failed for two reasons. The first was the non-uniformity of the dry etch of the oxide layer which is believed to be linked to the second reason. It is thought that the dose for the PMMA exposure was fractionally too low, meaning that although the sample had

developed out correctly over large areas there was still a residue of resist that could not be removed by oxygen ashing. The dry etch process is very sensitive to resist residue and hence non-uniform. Ideally this would be repeated using a slightly higher dose, however the filament of the Beamwriter failed, changing the beam profile, which means repeating the doses tests. Unfortunately this occurred at the end of the project and there was insufficient time to rework this process.

8.5 Conclusions

This chapter has investigated the fabrication of T-gates with gate length of less than 50nm. This has included the demonstration of a full HEMT process flow with a T-gate with a footprint of 25nm written within a source-drain gap. This gate was fabricated using a SiO_2 gate definition layer and a two step lithography process. It is expected that using this technique it is possible to fabricate functional HEMT devices with a gate length as low as 25nm in the near future.

It is felt that this will be the limit of gate lengths that can be produced using this resist structure and e-beam tool. A number of "tricks" have been played with the resists to reduce the effect of the forward scattering therefore reducing the gate length. This included using a two-stage lithography technique that eliminates the scattering through the resist layers that define the gate head. The thickness of the PMMA layer was also reduced to minimise the effect of forward scattering. This thinning of resist has reached its limit, as the stalk of the T-gate must be tall enough to prevent large parasitic capacitance between the gate and the cap.

The tool that used throughout of the course of this work, the Leica EBPG-5HR has a minimum spot size of 12nm. This is relatively large when compared to modern beam writer tools such as the Leica VB6 which has a spot size of ~4-5nm and the converted Jeol TEM with a spot size of around 3nm.

The reason for the limit of gate length of around 25nm with a spot size is 12nm is the way that the resist is exposed. As discussed in section 4.3.2 PMMA is not exposed by the high-energy primary electrons but the lower energy secondary electrons, which are produced as the primary electrons pass through the resist, the range of these low energy electrons is around 5nm and tend to travel perpendicular to the primary beam. [8.8] This production of secondary electrons coupled with the width of the electron beam leads to a limit to the size of the patterns that can be produced with the tool and resist combination. Accounting for the effect of forward scattering, this gives a limit of around 25nm.

Therefore there are two methods available for reducing the gate length of devices even further. The first is to use a more exotic resist as the gate definition layer such as the composite resist used by Suemitsu et al. [8.6] which consists of a mixture of ZEP520A and C_{60} fullerene molecules. The addition of the fullerene molecules is said to improve the sensitivity and the contrast of the resist. [8.9] The second method is to write the gate using a e-beam tool with a much smaller beam size, normally this would be infeasible due to the expensive of new Beamwriter, however, as this work was nearing completion the converted Jeol TEM in the Nanoelectronics Research Centre was becoming operational. As discussed in section 4.2.3 this beam writer is capable of writing patterns with extremely high resolution, regularly capable of writing patterns with features as small as 12nm. This is therefore a perfect tool for the continued development of ultra-small gate length devices. Work was started on transferring the sub-50nm T-gate process to the Jeol including consideration of factors that are unique to this type of tool, in particular achieving alignment accuracy, however due to time constraints is was not possible to complete this work.

8.6 References

[8.1] Y.Chen et al. "*Fabrication of T-shaped gates using UVIII chemically amplified DUV resist and PMMA*" Electronics Letters , **35** , 1999, p338

[8.2] Y. Chen, D. Macintyre, and S. Thoms, "*T-gate fabrication using a ZEP520A/UVIII bilayer*" Microelectronic Engineering, **57**, 2001, p939

[8.3] Y.Chen et al. "*Effect of Sensitivity Ratio on T-gate fabrication*", J. Vac. Sci. Technol. B, **19**, 2001, p2494

[8.4] Y.Chen et al., "*Fabrication of 30nm T-gates using SiN_x as a supporting and definition layer* ", J. Vac. Sci. Technol. B, **18**, 2000, p3521

[8.5] K. Shinohara et al. "*Ultrahigh-speed Pseudomorphic InGaAs/InAlAs HEMTs with 400-GHz Cutoff frequency*" IEEE Electron Device Letters, **22**, 2001, p507

[8.6] T. Suemitsu et al. "*30nm Gate InP based Lattice Matched High Electron Mobility Transistors*", JJAP, **38**, 1999, p L154.

[8.7] Communications with Hai-Ping Zhou

[8.8] B. Wu and A.R. Neureuther, "*Energy Deposition and transfer in electron beam lithography*", J. Vac. Sci. Tech. B, **19**, 2001 p2508

[8.9] T. Ishii et al., "*Nanocomposite resist system*" App. Phys. Lett., **70**, 1997, p1110

9.0 Conclusions

The aim of this work was to develop new fabrication technologies for the realisation of advanced millimetre-wave III-V High Electron Mobility Transistors (HEMTs). Within this rather wide-ranging aim two goals were targeted. First, the investigation of the performance benefits achieved by the reduction of parasitic resistances using a self-aligned device geometry. The second goal was to develop fabrication processes that would enable the realisation of ultra-short gate length HEMTs. This work included the development of two gate lithography processes capable of the realisation of 50nm and 30nm T-gates using a co-polymer and SiO_2 gate support layer respectively. To allow in depth analysis of these short gate length HEMTs a method for the preparation of cross sectional HEMT samples for TEM was also developed. The key outcomes of this research are summarised below.

The investigation of the self-aligned HEMT demonstrated that is was possible to realise a self-aligned GaAs pHEMT based on a 120nm T-gate process. However the expected performance benefits were not seen, the standard device structure fabricated within the same cell demonstrated better RF behaviour with a higher f_t and f_{max}. This low RF performance was the result of lower transconductance, which was most probably caused by the effect of the filling of traps close to the gate region of the device. This work did have a number of interesting spin-offs however, the first of these is the succinic acid based selective etch initially developed for the fabrication of self-aligned HEMTs, which has since found use in the fabrication of GaAs pHEMTs, metamorphic GaAs HEMTs and lattice matched HEMTs on InP substrates. The non-annealed ohmic contact developed for the self-aligned HEMT provides an extra degree of freedom in the process flow and allow the gates to be deposited much earlier in the process flow. [9.1] This technology also allowed the realisation of the first HEMT fabricated using Nano-Imprint Lithography [9.2,9.3].

The work carried out using the PMMA/Co-Polymer resist system was an exploration of the ultimate device gate length that could be achieved using the same resists as the existing 120nm T-gate process. This demonstrated that is was possible to use this resist system to fabricate devices with a gate length as small as 50nm. The electrical yield was exceptionally high with uniformity of the threshold voltage, of –0.445 with a standard deviation of 0.005 demonstrating the quality of the gate process. These devices showed excellent DC and RF performance with a DC transconductance of 1200mS/mm, a maximum cut-off frequency of 330GHz and a maximum frequency of oscillation, f_{max}, of 260GHz comparable with the highest performing metamorphic HEMTs. An additional benefit of this gate process is its ease of use, with only two layers of resist and an overnight bake the sample is easily prepared. Since the process uses standard resists as opposed to chemically amplified resists there is no post exposure bake. Further, the process is insensitive to the delay before exposure and development as well as UV light. This increases the manufacturability of the process compared to others capable of producing HEMTs with a gate length of 50nm such as the UVIII/LOR/PMMA also developed at Glasgow.

The gate length scaling of HEMTs was explored further in the development of a process for the realisation of HEMTs with a gate length of less than 50nm. This process used two stages of lithography to reduce forward scattering through the resist stack and a SiO_2 gate support and mask layer to help define the small gate foot and support the T-gate structure. This work included the controlled deposition of a thin SiO_2 layer, pattern transfer into the SiO_2 layer by a low damage dry etch process and lithography capable of producing a T-gate structure with a gate length of less than 30nm.

The gate lithography process was incorporated into a flow compatible with the fabrication of InP HEMTs. This process was used to fabricate HEMTs with a gate length as small as 25nm, the smallest HEMT device fabricated at the University of Glasgow and of the order of the smallest HEMT devices that have been reported.

The ultimate size of the gate produced was found to be limited to by a combination of the range of secondary electrons produced by the incident electrons, which travel perpendicular to the beam axis and have a range of 5-10nm in addition to the minimum spot size capable of being produced by the existing tool.

9.1 References

[9.1] E. Boyd et al. "*120 nm gate length e-beam and nanoimprint T-gate GaAs pHEMTs utilizing non-annealed ohmic contacts*", 2002 International Symposium on Compound Semiconductors, Lausanne, 2002, p291

[9.2] Y. Chen et al. "*High electron mobility transistors fabricated by nanoimprint lithography*", Microelectronic Engineering, 67-68, 2003, p189

[9.3] Y. Chen et al, "*Fabrication of high electron mobility transistors with T-gates by nanoimprint lithography*", J. Vac. Sci. Technol. B, 20(6), 2002, p2887

10.0 Further Work

Over the course of this work a number of processes and technologies have been developed, however as one technology is developed the need for a new process becomes clear. Due to time constraints it has not been possible to complete all of the work one would wish to. This chapter serves as a summary for the continuation of this work.

It was seen in chapter 6 that the 50nm T-gate process using a bi-layer of PMMA and copolymer is reliable and very uniform. However, it has also been shown that the gate recess etching of these shallow materials for ultra-short gate length HEMTs is problematic. The recess etch used for the 50nm T-gate HEMT results presented in chapter 6 was very short etch time to ensure the devices operated in depletion mode. As a result, the very short recess offset meant that the gate was often in contact with the highly doped InGaAs cap, leading to significant gate leakage. The short recess offset also contributed to the low breakdown of these devices. In future, if the devices are to be incorporated into a M^3IC process, this problem will have to be addressed. One possible solution may be to use a slightly thicker barrier layer, (e.g. 10nm instead of 8nm) and a double recess strategy. This would allow a larger recess offset in the cap layer leading to better diode and breakdown characteristics. The second etch would be narrower, shallower and would etch slightly into the barrier layer allowing the optimum gate-channel separation to be maintained. This could be achieved using either a combination of wet etch and dry etch or two different wet etch chemistries. This development would require effort but it is envisaged that the device performance would benefit.

The work on the preparation of the HEMT samples for TEM analysis using a thick layer of HSQ shows promise. The HSQ was seen to flow beneath the gate, protecting the recessed

region. Future work would concentrate on perfecting the milling process, possibly using a smaller beam current to thin the samples further to achieve higher resolution images.

The fabrication of T-gates with a gate length of 25nm was demonstrated in chapter 8, however these devices didn't operate due to the discontinuous gate caused by irregular etching of the SiO_2 gate definition layer. This was caused by a small amount of resist residue along the foot of the T-gate. This problem could be solved with further dose tests with the beam writer filament in a similar condition. Recent experiments have also suggested that it may be possible, with careful processing to lift-off 30nm gate length structures with high yield without the use of a SiO_2 support layer. This would remove the dry-etching stage, which has been seen to be problematic. It is envisaged that in the near future a gate length of 30nm could be fabricated in this manner.

One of the conclusions of this thesis was that with the use of an electron beam writer tool with a smaller beam size it would be possible the produce T-gates with a gate lengths below 25nm. This is possible with the Jeol converted TEM which has a spot size of around 3nm. However, since the Jeol is essentially still a TEM rather than a dedicated lithography tool there are some limitations. The beam is only well defined over a very small area, around 15x15μm compared to a minimum block size of 160x160μm using the Leica EPBG5HR. Outside this range, aberrations in the beam occur which will affect the writing of high resolution patterns. This dictates the width of gates that can be written to be around 15μm compared to 20μm used in the majority of devices fabricated using the ultra-short gate processes described in this work.

The alignment accuracy of the Jeol is around 200nm, which means that it is not possible to use the "bi-lithography" process presented in chapter 8 - a single step process must be employed. The use of a one step process means that the thickness of the upper layers of resist will have to

be carefully controlled to minimise forward scattering. This could include replacing the LOR separation layer with 5-10nm aluminium layer, this was originally part of the UVIII/PMMA 70nm T-gate process but was removed to make the process compatible with the post etch HF dip. This aluminium will be deposited by evaporation which allows very thin layers to be put down with much greater control than can be achieved by spinning LOR. By using a non-annealed ohmic contact technology such as the one described in chapter 5 the problems of gate lithography within a source-drain contact could be eliminated. Using the above techniques it should be possible in future to fabricated HEMT devices with a T-gate of 20nm or less.

Appendix A

120nm GaAs PHEMT Process

Step No.	
	CLEAN
1	5mins Ultrasonic Acetone
2	5mins Ultrasonic IPA
3	N2 Blow Dry
4	Inspect1 Optical Microscope
	MARKERS
5	Spin 12% 2010ELV 5K 60secs
6	Bake 180C 1 hour
7	Spin 4% 2041 ELV 5K 60secs
8	Bake 180C 2 hour
9	Inspect 2 Optical Microscope
10	Expose Markers level
	Dose 310, Res 150, spot 300
11	1:1 IPA : MIBK 30 sec 23C
12	30 sec IPA
13	N2 Blow Dry
14	Inspect 3 Optical Microscope
15	Barrel O2 Etch 60sec
16	1:4 HCL:H20 30sec
17	10sec H20
18	N2 Blow Dry
19	evap markers in plassys (20nm Ti : 130nm Au)
20	1 hour warm acetone
21	Pipette Lift Off
22	30sec IPA
23	N2 Blow Dry
24	Inspect 4 Optical Microscope
	Ohmic
25	Spin 2010 12% ELV 5K 60secs
26	Bake 180C 1 hour
27	Spin 4% 2041 ELV 5K 60secs
28	Bake 180C 2 hour

29	Inspect 5 Optical Microscope
30	Expose Ohmic level
	Dose 260, Res 0.1, spot 160
31	1:1 IPA : MIBK 30 sec 23C
32	30 sec IPA
33	N2 Blow Dry
34	Inspect 6
35	Barrel O2 Etch 60sec
36	1:4 HCL:H20 30sec
37	10sec H20
38	N2 Blow Dry
39	evap ohmic in plassys
40	1 hour warm acetone
41	Pipette Lift Off
42	IPA rinse
43	Inspect 7 Optical Microscope
44	RTA
45	Inspect 8 Optical Microscope
	and DC TLM test

	ISOLATION
46	Spin 12% 2010 ELV 5K 60secs
47	Bake 180C 1 hour
48	Spin 4% 2041 ELV 5K 60secs
49	Bake 180C 2 hour
50	Inspect 9 Optical Microscope
51	Expose Isolation level
	Dose 225, Res 0.15, spot 300
52	1:1 IPA : MIBK 30 sec 23C
53	30 sec IPA
54	N2 Blow Dry
55	Inspect 10 Optical Microscope
56	Barrel O2 Etch 60sec
57	120 C 30 min Post Bake
58	1:4 HCL:H20 30sec
59	10sec H20
60	N2 Blow Dry
61	1:1:200 NH4:H202:H20
62	30sec H20
63	N2 Blow Dry
64	1 hour warm acetone
65	Pipette Lift Off
66	30sec IPA
67	N2 Blow Dry
68	INSPECT 11 Optical Microscope

	and Dektak and DC TLM

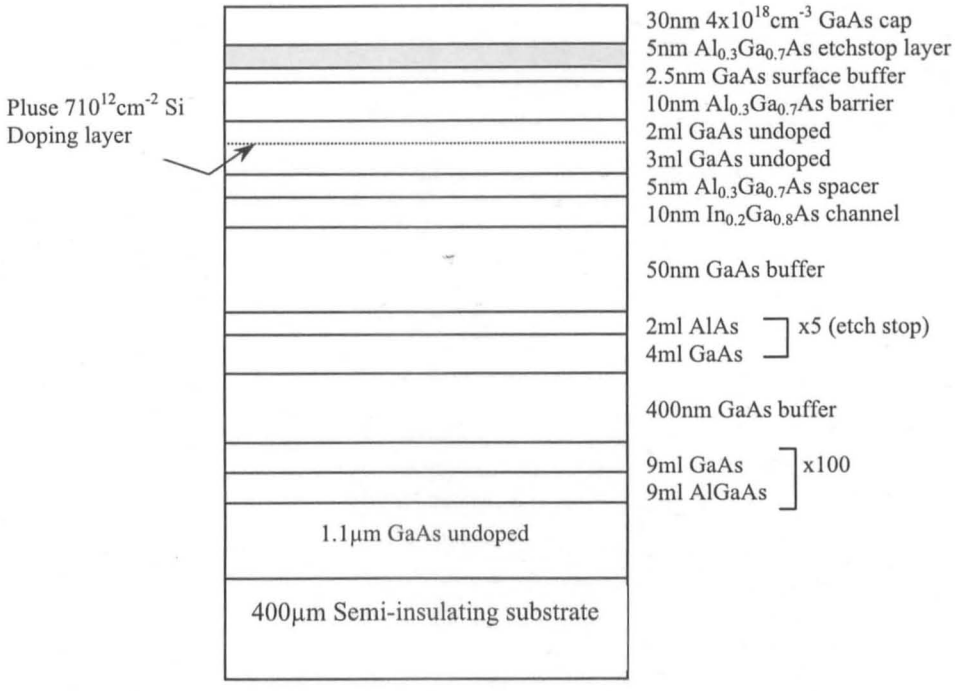
	GATES
69	Spin 4% 2041 ELV 5K 60secs
70	Bake 1 hour 180C
71	Spin 9% Co_Poly 5K 60secs
72	Bake 1hour 180C
73	Spin 2.5% 2010 ELV 5k 60secs
74	Bake overnight 180C
75	Inspect 12 Optical Microscope
76	expose gates 118uC/cm 20nm res 40nm spot
77	develop 2.5:1 IPA:MIBK 30sec 23C
78	30sec IPA
79	N2 Blow Dry
80	Scribe Test piece
81	Inspect 13 S900(test piece #1)
82	Dry Etch SiCl4/SiF4/O2 (test piece #2)
83	HF 10 sec 10:1
84	H20 30 sec
85	Inspect 14 S900(test piece #2)
86	Dry Etch SiCl4/SiF4/O2 (sample)
87	HF 10 sec 10:1
88	H20 30 sec
89	Inspect 14 S900(test piece #3)
90	Evaporate gates (15nmTi : 15nm Pd : 180:Au)
91	2 hour warm acetone
92	Pipette Lift Off
93	30sec IPA
94	Inspect 15 S900 (test piece #4)

	CPW Bond-pads
95	Spin 15% 2010 ELV 5K 60secs
96	Bake 120C 1 hour
97	Spin 4% 2041 ELV 5K 60secs
98	Bake 120C 2 hour
99	Inspect 17 Optical Microscope
100	Expose Bondpad level
101	Dose 330, Res 150, spot 300
102	1:1 IPA : MIBK 30 sec 23C
103	30 sec IPA
104	N2 Blow Dry
105	Inspect 18 Optical Microscope
106	Barrel O2 Etch 60sec

107	1:4 HCL:H20 30sec
108	10sec H20
109	N2 Blow Dry
110	Evaporate BondNCR (50nmNiCr :400nm Au)
111	Acetone
112	30sec IPA
113	Inspect 19 Final Test
114	Store Sample

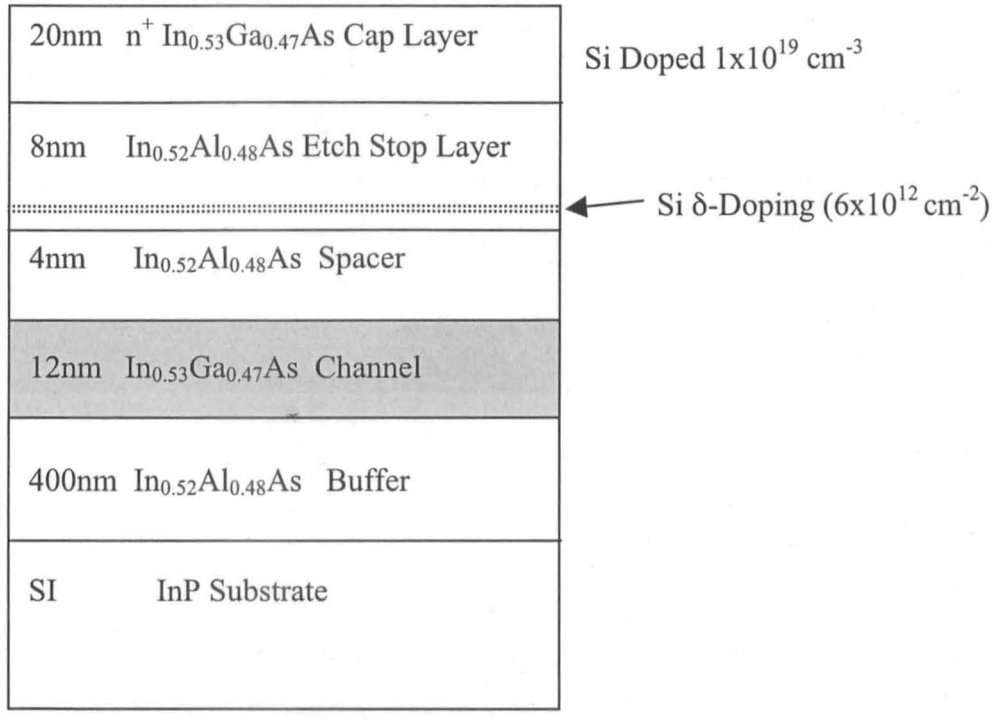
Appendix B

Layer Structure of A1258 GaAs pHEMT



Appendix C

Layer Structure of A1602 Lattice Matched HEMT on InP



Appendix D

120nm Lattice-Matched and Metamorphic HEMT Process

MARKERS - M1

SAMPLE CLEAN	**SPIN**	**EXPOSE**	**DEVELOP**	**OXYGEN ASH**	**DE-OXIDISE**	**EVAPORATE**
In water bath	12% ELV 2010 5k 60s	Beamwriter	1:1 IPA:MIBK	Barrel etcher	Wet IPA	(plassys)
5 min Acetone	Bake 1hr 180°C			For 60s	1:4 HCL:H2O	
Rinse		310μC/cm2	For 30s @ 23°C		For 30s	20nm Ti
5 min IPA	4% ELV 2041 5k 60s	150nm Res	IPA for 30s		Rinse H2O	130nm Au
Rinse	Bake 2hr 180°C	300nm Spot	Blow dry		Blow dry	
Blow Dry						Base Pressure
				Power Check		< 1x10-6
Spin Resist Immediately				**< 110 Watts**		
			INSPECTION			**LIFTOFF**

OHMIC CONTACTS **JOB NAME:**

SPIN	EXPOSE	DEVELOP	OXYGEN ASH	DE-OXIDISE	EVAPORATE	RTA
12% ELV 2010 5k 60s Bake 1hr@180°C 4% ELV 2041 5k 60s Bake 2hr @180°C	Beamwriter 260μC/cm2 100nm res 160nm spot With Proxeco Dose 176μC	1:1 IPA:MIBK For 30s @ 23°C IPA for 30s **INSPECTION**	Barrel etcher For 60s **Power Check** **< 110 Watts**	Wet IPA 1:4 HCL:H2O For 30s Rinse H2O Blow Dry	(plassys) ncohmic **10 nm Ni** **50 nm Ge** **80 nm Au** Base Pressure < 1x10-6 **LIFTOFF**	Recipe 1 10s ramp **20s 240** 10s ramp **30s 280** 10s ramp in N2

MESA ISOLATION **JOB NAME:**

SPIN	EXPOSE	DEVELOP	OXYGEN ASH	DE-OXIDISE	MESA ETCH	RESIST STRIP
12% ELV 2010 5k 60s Bake 1hr@180°C 4% ELV 2041 5k 60s Bake 2hr @180°C	Beamwriter 225μC/cm2 150nm res 300nm spot	1:1 IPA:MIBK For 30s @ 23°C IPA for 30s **INSPECTION**	Barrel etcher For 60s **Power Check** **< 110 Watts** **POSTBAKE** **Bake 120°C 30m**	Wet IPA 1:4 HCL:H2O For 30s Rinse H2O Blow Dry	1:1:100 H3PO4 (70%) H2O2:H2O steps of 20s I <100nA, ±2V Light Off	1hr Water Bath, Acetone 30s IPA Rinse IPA Blow Dry **Talystep Mesa** **Measure TLMS**

GATES **JOB NAME:**

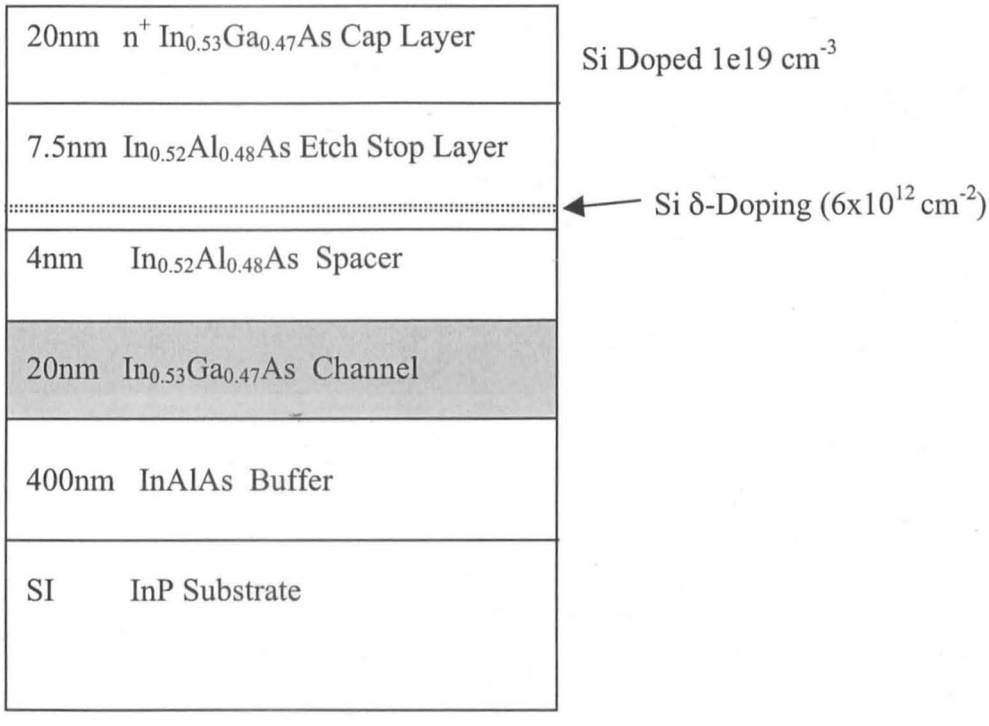
SPIN	EXPOSE	DEVELOP	DE-OXIDISE	WET ETCH	EVAPORATE	
4% ELV 2041 5k 60s	Beamwriter	**2_:1**	Wet IPA		(Plassys)	
Bake 1hr@180°C		IPA:MIBK	1:4 HCL:H2O	1:1:200		
	118μC/cm2	For 30s @ 23°C	For 30s	H2O2:H3PO4:H2O	15nm Ti	
9% Copoly 5k 60s	20nm res	IPA for 30s	Rinse H2O	1% FC-93	15nm Pd	
Bake 1hr 180°C	40nm spot		Blow dry		260nm Au	
				Etch Time ~30s		
2.5% ELV 2010 5k 60s				+ steps of 10s		
Bake night 180°C		**INSPECTION**		Measure Current		
		TEST PIECE			LIFTOFF – 2hrs Warm Acetone	

BONDPAD - M2 **JOB NAME:**

SPIN	EXPOSE	DEVELOP	OXYGEN ASH	DE-OXIDISE	EVAPORATE	
15% ELV 2010 5k 60s	Beamwriter	1:1	Barrel etcher	Wet IPA	(Plassys)	
Bake 1hr @ 120°C		IPA:MIBK	For 60s	1:4 HCL:H2O	Bondncr	
					50nm NiCr,	
	330μC/cm2	For 30s @ 23°C		For 30s	1200nm Au	
4% ELV 2041 5k 60s	150nm res	IPA for 30s				
Bake 2hr @120°C	300nm spot				Check Xtal Health	
		INSPECTION	**Power Check**			
Note: 15%			**< 110 Watts**		Base Pressure	
Note: 120°C		**IMPORTANT:**			< 1x10-6	
		TALYSTEP RESIST & RECORD			**LIFTOFF**	

Appendix E

Layer Structure of A1881 GaAs metamorphic HEMT



Appendix F

30nm Lattice Matched and Metamorphic HEMT Process

SiO2 DEPOSITION

CLEAN	DE-OXIDISE	DEPOSITION	CHECK 1			
In water bath	Wet IPA	System 100 DEP	Scratch Resist			
5 min Acetone	1:4 HCL:H2O	Depth : 20nm	In each corner			
Rinse IPA	For 30s	Cold Load				
5 min IPA	Rinse H2O		AFM Depth			
Rinse IPA	Blow dry					
			Thickness			
			25 +- 5nm			

MARKERS - M1

SAMPLE CLEAN	SPIN	EXPOSE	DEVELOP	OXYGEN ASH	DE-OXIDISE	EVAPORATE
	12% ELV 2010 5k					
In water bath	60s	Beamwriter	1:1 IPA:MIBK	Barrel etcher	Wet IPA	(plassys)
5 min Acetone	Bake 1hr 180°C			For 60s	1:10 HF 15sec	
Rinse		310μC/cm2	For 30s @ 23°C		For 30s	20nm Ti
5 min IPA	4% ELV 2041 5k 60s	150nm Res	IPA for 30s		Rinse H2O	130nm Au
Rinse	Bake 2hr 180°C	300nm Spot	Blow dry		1:4 HCL:H2O 30s	
Blow Dry					Rinse H2O	Base Pressure
				Power Check	Blow Dry	< 1x10-6
Spin Resist Immediately				**< 110 Watts**		**LIFT OFF**

OHMIC CONTACTS **JOB NAME:**

SPIN	EXPOSE	DEVELOP	OXYGEN ASH	DE-OXIDISE	EVAPORATE	RTA
12% ELV 2010 5k 60s Bake 1hr@180°C 4% ELV 2041 5k 60s Bake 2hr @180°C	Beamwriter 225µC/cm2 With Proxecco 100nm res 160nm spot	1:1 IPA:MIBK For 30s @ 23°C IPA for 30s **INSPECTION**	Barrel etcher For 60s **Power Check** **< 110 Watts**	Wet IPA 1:10 HF 15sec Rinse H2O For 30sec 1:4 HCL:H2O 30s Rinse H2O Blow Dry	(plassys) ncohmic **10 nm Ni** **50 nm Ge** **80 nm Au** Base Pressure < 1x10-6 **LIFTOFF**	Recipe 1 10s ramp **20s 240** 10s ramp **30s 280** 10s ramp in N2

MESA ISOLATION **JOB NAME:**

SPIN	EXPOSE	DEVELOP	OXYGEN ASH	DE-OXIDISE	MESA ETCH	RESIST STRIP
12% ELV 2010 5k 60s Bake 1hr@180°C 4% ELV 2041 5k 60s Bake 2hr @180°C	Beamwriter 225µC/cm2 150nm res 300nm spot	1:1 IPA:MIBK For 30s @ 23°C IPA for 30s **INSPECTION**	Barrel etcher For 60s **Power Check** **< 110 Watts** **POSTBAKE** **Bake 120°C 30m**	Wet IPA 1:10 HF 15sec For 30s Rinse H2O 1:4 HCL:H2O 30s Rinse H2O Blow Dry	1:1:100 H3PO4 (70%) H2O2:H2O steps of 20s I <100nA, ±2V Light Off	1hr Water Bath, Acetone 30s IPA Rinse IPA Blow Dry **Dektak Mesa** **Measure TLMS**

GATES **JOB NAME:**

SPIN	EXPOSE 1	EXPOSE 2	DRY ETCH	WET ETCH	POST ETCH	EVAPORATE
2.5% ELV 2041 3k 60s	Beamwriter 50kV	Beamwriter 100kV	system 100			(PlassysII)
Bake 1hr@180°C	90μC/cm2	2150 C/cm2	CHF3	10g Succinic Acid	1:1:100	
	100nm res	10nm res	Flow 30 sccm	8ml H2O2	H3PO4 (70%)	15nm Ti
1:4 LOR 5k 60s	160nm spot	12nm spot	Power 80W	5ml Ammonia	H2O2:H2O	15nm Pt
Bake 15mins 180°C	**DEVELOP 1**	**DEVELOP 2**	Pressure 22mbar	Balance pH 5.5	4 sec	200nm Au
	Hotplate 120°C 90s		Time 6mins	Time 20sec	H20 Rinse	
58% UVIII 5k 60s	CD26 Room Temp	o-xylene 23°C 60s			Blow Dry	**Pressure**
Hotplate 120°C 60s	60 sec	H20 rinse				**< 3x10-7**
				LIFTOFF		
Keep In BAG	H20 Rinse			**– At least 2hrs in Warm Acetone**		**(1hr Pump Down)**

BONDPAD - M2 **JOB NAME:**

SPIN	EXPOSE	DEVELOP	OXYGEN ASH	DE-OXIDISE	EVAPORATE	
15% ELV 2010 5k 60s	Beamwriter	1:1	Barrel etcher	Wet IPA	(Plassys)	
Bake 1hr @ 120°C		IPA:MIBK	For 60s	1:4 HCL:H2O	Bondncr	
	330μC/cm2	For 30s @ 23°C		For 30s	50nm NiCr,	
4% ELV 2041 5k 60s	150nm res	IPA for 30s			400nm Au	
					Check Xtal	
Bake 2hr @120°C	300nm spot				Health	
		INSPECTION	**Power Check**			
Note: 15%			**< 110 Watts**		Base Pressure	
Note: 120°C		**IMPORTANT:**			< 1x10-6	
		TALYSTEP RESIST & RECORD			**LIFTOFF**	