Surface Topography of Silicon Microcircuits

Thesis submitted by

Martin Fallon

for the degree of

Doctor of Philosophy.

Edinburgh Microfabrication Facility

Edinburgh University

1992

Table of Contents

Abstract ix

Declaration of Originality x

Acknowledgements xi

Glossary xii

1 INTRODUCTION 1

1.1 HISTORICAL BACKGROUND 1

1.2 THE TRANSISTOR 2

1.3 MINIATURISATION 3

1.3.1 Motivation 3

1.3.2 Scaling 6

1.3.3 Non-Scaleable Constants 8

1.4 THESIS OUTLINE 10

1.4.1 Thesis Motivation 10

1.4.2 Thesis Content 11

REFERENCES 12

2 ISOLATION PROCESSING 14

2.1 INTRODUCTION 14

2.2 OXIDATION 15

2.3 LOCOS 17

2.3.1 White Ribbon 18

2.3.2 2-D Profiles 20

2.3.3 3-D Profiles 23

2.4 LOCOS ALTERNATIVES 24

2.4.1 General 24

2.4.2 Silicon Etch Processes 25

2.4.3 Poly Buffered Processes 27

2.4.4 Other Options 28

2.6 SUMMARY 29

REFERENCES 30

3 DEVICE PHYSICS 34

3.1 INTRODUCTION 34

3.2 THE CRYSTAL LATTICE 34

3.2.1 The Fermi Energy 35

3.2.2 Energy Bands 35

3.2.3 The Fermi Energy 36

3.2.3 Intrinsic Carrier Concentration 37

3.2.4 Law Of Mass Action 38

3.2.5 Impurities 40

3.2.6 Conduction 43

3.2.6.1 Mobility and Drift Velocity 43

3.2.6.2 Diffusion Current 45

3.2.6.2 Minority Carrier Lifetime 46

3.3 THE PN JUNCTION DIODE 47

3.3.1 Reverse And Forward Bias 52

3.4 THE FET TRANSISTOR 54

3.4.1 The Junction FET 54

3.4.2 The IGFET 56

3.4.3 The 2 Terminal MOSFET 57

3.4.4 General Analysis Of The 2 Terminal MOSFET 63

3.4.5 The Three Terminal Device 72

3.4.5 The Four Terminal Device 74

3.5 ION IMPLANTED CHANNELS 77

3.6 SUMMARY 78

REFERENCES 79

4 DEVICE WIDTH 80

4.1 INTRODUCTION 80

4.2 THRESHOLD VOLTAGE 81

4.2.1 Width Dependence Of Threshold Voltage 82

4.2.2 The Parasitic Field Transistor 85

4.3 DEVICE WIDTH 87

4.3.1 The Series Resistance 88

4.3.2 Mobility Modulation 90

4.3.3 Width Definitions 92

4.3.3 Electrical Extraction Techniques 94

4.3.4 Width Formulation 104

4.4 SUMMARY 111

REFERENCES 112

5 SEM AND MODELLED RESULTS 118

5.1 INTRODUCTION 118

5.2 FABRICATION 118

5.2.1 The Process Variations 121

5.3 RESULTS 1: TOPOGRAPHY 121

5.3.1 SEM Profiles 122

5.4 RESULTS 2: SIMULATION 126

5.4.1 Process Modelling 127

5.4.2 Device Models 131

5.5 NOVEL EXTRACTION STRATEGY 137

5.6 SUMMARY 140

REFERENCES 141

6 ELECTRICAL EXTRACTION 143

6.1 INTRODUCTION 143

6.2 SERIES RESISTANCE EXTRACTION 143

6.3 WIDTH EXTRACTION I: $I_{edge} = 0$ 144

6.3.1 Channel Normalisation 145

6.3.2 Channel Conductance Extrapolation 148

6.4 WIDTH EXTRACTION I: $I_{edge} \neq 0$ 151

6.5 ERRORS AND APPROXIMATIONS 154

6.5.1 Channel Normalisation: Error In Series Resistance 155

6.5.2 Channel Normalisation: Error In Threshold Voltage 158

6.5.3 G_t' Extrapolation: Errors In Series Resistance 163

6.5.4 G_t' Extrapolation: Errors In Threshold Voltage 163

6.6 MEASUREMENT ANALYSIS 165

6.6.1 Correlation Coefficient 170

6.7 DATA SET SELECTION 171

6.7.1 Extrapolation Using Two-Stage Processes 175

6.7.2 Data Set Optimisation 179

6.8 RESULTS 185

6.9 SUMMARY 187

REFERENCES 187

7 DISCUSSION AND CONCLUSIONS 189

7.1 INTRODUCTION 189

7.2 RESULTS COMPARISION 189

7.3 REVIEW 192

7.3.1 SEM 192

7.3.2 Simulation 193

7.3.3 Parameter Extraction 192

7.4 RECOMMENDATIONS AND FUTURE WORK 194

REFERENCES 195

APPENDIX 196

PUBLISHED PAPERS 217

Abstract

The shrinking dimensions of silicon microcircuits have reached the point where the vertical and lateral features are comparable in size. The consequence can be seen in each aspect of the manufacture of devices. The 2D layout of the physical routing becomes a convoluted maze when put into fabrication. The diminishing dimensions have focused greater attention on the edge effects since these play a proportionately greater role in the device performance. The consequences of the edge interactions can be categorised into two sections: those on the silicon surface and those on the subsequent layers .

The MOS transistor is directly impacted by the silicon surface profile. A fundamental parameter is the transistor width, which till recently has received little attention. This thesis correlates the different definitions commonly used, and investigates the impact of the individual processing parameters on the surface topography and consequently on the transistor width. Different measurement techniques are used and a novel extraction process is proposed.

The weakness of the current generic electrical extraction technique is exposed and recommendations made to overcome this. Further work on SEM sample preparation and processing is presented.

Declaration

I declare that all the work in this thesis is entirely my own unless otherwise indicated.

Martin Fallon

Acknowledgements

I would like to thank Dr A. J. Walton for his unstinting enthusiasm, A. M. Gundlach for his unfailing *joie-de-vivre*, R. J. Holwill for his good ear, and the staff of the EMF for their constant encouragement throughout.

I also wish to thank my wife, Susan, who has not known married life without the PhD.

Lastly I wish to thank big Sean for all his inspiration: *And that's a fact*

To my wife.

Susan

'Writing a Ph.D. is easy - you just stare at a blank sheet of paper till drops of blood form on your forehead'

List Of Symbols

β	$q/k_B T$	V^{-1}
C'_{ox}	oxide capacitance per unit area	Fm^{-2}
C'_s	silicon capacitance per unit area	Fm^{-2}
ΔW	channel width reduction	m
D	diffusion coefficient	$m^2 s^{-1}$
Ec	conduction band energy	eV
E_F	Fermi energy	eV
E_g	band gap energy	eV
E_v	valence band energy	eV
ε_{ox}	permittivity of silicon dioxide	Fm^{-1}
E	electric field	Vm^{-1}
G	conductance	AV^{-1}
G_P	parallel conductance	AV^{-1}
γ	body effect coefficient	$V^{1/2}$
I_{DS}	source-drain current	A
I_{edge}	edge current	A
J	current density	Am^{-2}
κ	scaling constant	
k_B	Boltzmann's constant	JK^{-1}
k_s	surface rate reaction constant	ms^{-1}
L	drawn channel length	m
L_{bb}	bird's beak length	m
L_{eff}	effective channel length	m

μ	*carrier mobility*	$m^2V^{-1}s^{-1}$
μ_s	*surface carrier mobility*	$m^2V^{-1}s^{-1}$
μ_o	*low field carrier mobility*	$m^2V^{-1}s^{-1}$
μ_c	*chemical potential*	V
n	*n-carrier concentration*	m^{-3}
n_i	*intrinsic carrier concentration*	m^{-3}
N_A	*no. of acceptors per unit volume*	m^{-3}
N_D	*no. of donors per unit volume*	m^{-3}
p	*p-carrier concentration*	m^{-3}
ψ_s	*surface potential*	V
ϕ_{ms}	*metal-semiconductor work function*	V
ϕ_F	*Fermi potential*	V
q	*electron charge*	C
Q_f	*fixed oxide charge/ unit area*	Cm^{-2}
Q_{ot}	*oxide trapped charge/ unit area*	Cm^{-2}
Q_m	*mobile oxide charge/ unit area*	Cm^{-2}
Q_{it}	*interface trapped oxide charge/ unit area*	Cm^{-2}
R_{P0}	*width × resistance product*	Ωm
R_{chan}	*channel resistance*	Ω
R_x	*series resistance*	Ω
ρ	*charge density per unit volume*	Cm^{-3}
T_{pad}	*pad oxide thickness*	m
T_n	*silicon nitride thickness*	m
T	*absolute temperature*	K
τ	*fitting factor*	s
τ_t	*transit time*	s^{-1}

τ_n	*electron minority carrier lifetime*	s
τ_p	*hole minority carrier lifetime*	s
θ	*mobility modulation coefficient*	V^{-1}
t_{ox}	*oxide thickness*	m
U	*recombination rate*	s^{-1}
v_D	*drift velocity*	ms^{-1}
V_{DS}	*drain to source voltage*	V
V_{GS}	*gate to source voltage*	V
V_t	*threshold voltage*	V
V_{G_t}	*voltage above threshold*	V
W_D	*drawn width*	m
W_{eff}	*effective width*	m
x_n	*n region junction depth*	m
x_p	*p region junction depth*	m

Chapter 1

Introduction

1.1. Historical Background

The Computer

As a calculating machine, the computer has existed for thousands of years. The fundamental principles upon which computers are based, are those of mathematics, whereby every complex computation can be reduced to addition and subtraction, and for that matter subtraction is nothing more than the opposite of addition. Thus it was the manual dexterity of the operator that limited the performance of the first computer, the abacus. It was not until 1822 with the realization of a proposal by the mathematician Charles Babbage, which produced a mechanical counting machine, that a new step had been taken in computer technology. His suggestion to mechanize the long, tedious calculations necessary to produce the nautical and astronomical charts of the day was embodied in "Babbage's Engine" [1] which successfully counted and carried by means of gearing. His failure to produce a universal machine rested on the intricate gearing necessary but unobtainable at that time.

In 1930, Vannevar Bush at MIT built a mechanical computer which was driven by electric motors. A World War II problem led to the first electronic computer. A means was sought to calculate quickly ballistic tables for a wide variety of shells. J. W. Mauchly and J. P. Eckbert at the University of Pennsylvania tackled the project. The result was the Electronic Numerical Integrator and Calculator, ENIAC [2] . ENIAC weighed over 30 tons, occupied $150m^2$ of floor space, used 18,850 vacuum

tubes and required 150 kW to keep it operating. The world's first electronic computer was completed in 1946.

On 23rd December 1947 at Bell Labs, John Bardeen and Walter H. Brattain demonstrated the transistor which sounded the death knell for the technology of ENIAC. Along with William B. Shockley, they were awarded the Nobel prize, for what is arguably the most significant technological development of this, or any other, century.

1.2. The Transistor

The Field Effect Principle upon which this was based had been proposed as a means of producing amplifying devices as early as the 1926 by Lillenfield and 1938 by Heil[3] . In 1945 Shockley suggested making a semiconductor amplifier based on this principle. Early experiments failed, however, and it was Bardeen's suggestion of electron trapping in surface states and the possibility of an inversion layer which led to the publication by Bardeen and Brattain in 1947 of the point contact transistor [4] .

The development of the transistor, *per se*, would not have been sufficient to produce the unheralded growth in the industry. The properties of semiconductors and pattern transfer were the key to the integration of discrete components. In 1958, J. S. Kilby of Texas Instruments produced the first I.C. using black wax to mask etching of mesas to produce the active elements [5] . The advantages of this technology over valves were enormous. Valves were costly, bulky and unreliable; of limited life and dissipated large amounts of power; they also had to be manually integrated by solder connections to other components. Interest in semiconductors proliferated. In 1960, Kahng and Atalla fabricated the first modern surface field effect transistor [6] and in 1961 P. K. Weimer presented the first evaporated thin film transistor [7] . Reference 7 is a collection of the early publications in this field.

From a point start three decades ago, worldwide sales of I.C.s have risen to $46.65billion in 1991 [8] .

Figure 1.1 charts this growth.

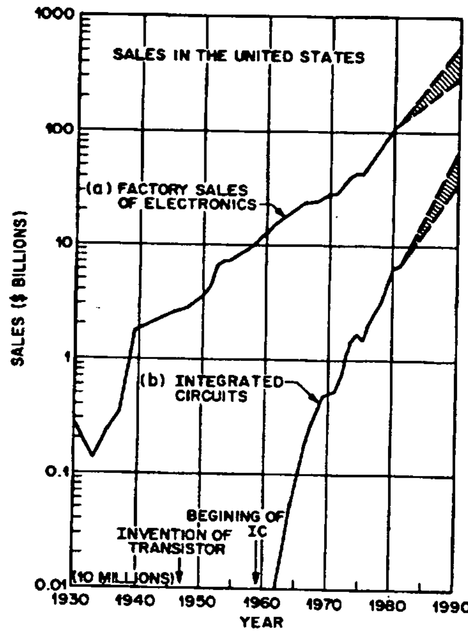


Figure 1.1. Factory sales of I.C.s. [9]

The wide interest generated further technological advances which facilitated continued miniaturisation and integration. Table 1.1 highlights the process innovations, between conception and introduction into manufacturing.

1.3. Miniaturisation

1.3.1. Motivation.

In any industry, technology change is motivated by cost. The continual change in the semiconductor industry is to reduce the size of the devices. In the early days of the industry the benefits of size reduction were evident. Communication electronics in the heavy bomber of the 1940s comprised 25% of the payload. Miniaturisation

History Of Technology Discontinuities		
Tool or Technology	Year Developed	Year in Production
Silicon Epitaxy	1960-61	1964
Atmospheric SiN	1965	1967-68
Ion Implant	1969	1973
TiW	1969	1973
Schottky TTL	1970	1974-7
CCDs	1970	1981
RIE	1975-76	1980
Advanced Schottky	1976	1980
Polysilicon emitter	1976	1984-85
Refractory Gate	1976	1983
SOI - Ion Implant	1978	1989
Trench	1979	1989
Silicide	1978	1985
LDD	1980	1986
TiN Local Strap	1986	1988

Table 1.1. Technology Developments.[8]

was then an economic must. In the electronics of to-day's microchip, the economics are still the driving force but the reasons are more subtle.

The planar process which is used for I.C. fabrication produces uniform device features across a whole wafer. It is the most successful and least recognized example of automation ever produced. Because the devices are 'batch produced' on a single

chip, the chips 'batch produced' on a single wafer, and the wafers batch produced in lots of 25 or 50, the cost of processing one device diminishes as the number of devices/wafer increases.

Another reason for this can be seen in Fig 1.2.

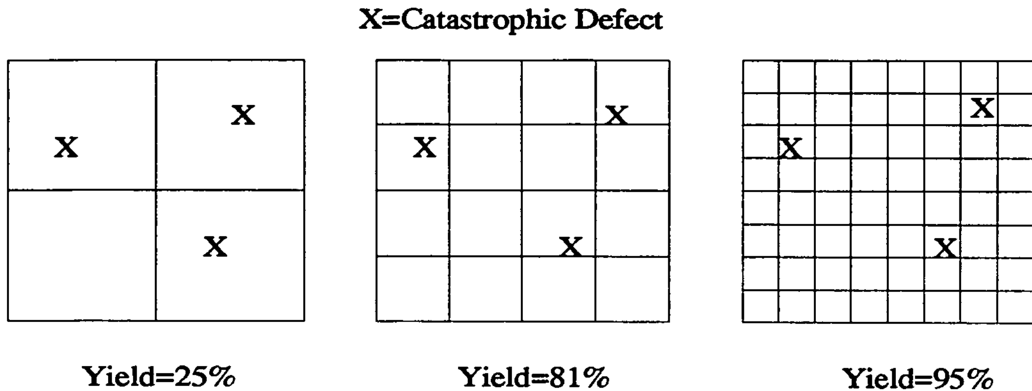


Figure 1.2. Chip size and yield.

A continual war is being fought in the fabrication line against particulate. When chip sizes are large, a larger proportion of chips are lost due to a fixed number of contaminant particles. Thus the yield can be increased dramatically as the device dimensions reduce. So it was that the industry saw the introduction of large scale integration (LSI), ($> 10^3$ components/chip), in the 1970s and VLSI ($> 10^4$ components/chip) in the late 1970's.

The benefits of 'downscaling' are seen as higher speed, increased functionality and lower cost. The DRAM production is the driving force in developing new technology to meet the requirements of the shrinking device. Every 3 years a new DRAM with a ×4 increase in bit density appears on the market as the previous generation begins to wane. At Texas Instruments in late 1990, the 256Mbit DRAM arrived at the research stage, the 64Mbit was in development, and the 16Mbit in product characterisation. The 4Mbit was then ramping up to full manufacture. These 4 phases comprise a 10 year cycle, when with the introduction of the 1024Mbit to the

research stage, the sequence will shift down one step [10] .

1.3.2. Scaling.

The inexorable drive to smaller devices has continued towards the fundamental physical limits which must define an end point to the present basic structures. For example, as feature sizes become comparable to the wavelength of light, the seeming elasticity in the limits of optical lithography will mean that its domination over X-ray and e-beam systems must yield eventually. The study of size reduction has produced the principle of scaling whereby all lateral and vertical dimensions are reduced by a factor of κ. Beyond this, two options are open. These and their consequences are shown in Table 1.2.

In the first instance of 'constant voltage', all the physical dimensions within the circuit are scaled down by a factor of κ, but the supply and threshold voltages are kept constant. Substrate doping must be increased by a factor of κ^2 also, to scale the depletion widths around the source and drain junctions. The constant supply voltage across a reduced gate length increases the power per gate by a factor of κ. Since the gate density increases by a factor of κ^2, the power density increases by a factor of κ^3. Applying these rules and projecting gate lengths down to 0.5 μm, it would appear from geometric considerations, to be possible to put 800,000 gates on a $6 \times 6mm^2$ chip. If this were done however, the power dissipation would be of the order of 1 kWatt.

The other option of constant field scaling overcomes this problem by maintaining a constant power density, but results in system-incompatible supply voltages and reduced noise margin in threshold voltages. In practice, a recipe for scaling is produced by some massaging of these two scaling principles. This however begs the question: how large can κ become? This is essentially asking what the minimum effective channel length can become. A rough approximation can be

Paramater	Constant Voltage	Constant Field
Feature Size	$1/\kappa$	$1/\kappa$
Gate/Field Oxide Thickness	$1/\kappa$	$1/\kappa$
Gate Capacitance	$1/\kappa$	$1/\kappa$
Junction Depth	$1/\kappa$	$1/\kappa$
Substrate Doping Conc.	κ^2	κ
Poly Sheet Res.	κ	κ
Linewidth	$1/\kappa$	$1/\kappa$
Transistor Gain	κ	κ
Supply Voltage	1	$1/\kappa$
Enhancement/Deplt V_T	1	$1/\kappa$
Current/Gate	κ	$1/\kappa$
Relative RC	1	1
Power/Gate	$1/\kappa$	$1/\kappa^2$
Power Density	κ^3	1
Gate Density	κ^2	κ^2
Gates for 1W	$1/\kappa$	1

Table 1.2 Scaling Factors [11]

obtained by imposing 2 limitations on the MOSFET [10] :

1. $L_{\min} > 2 \times$ drain depletion width

2. N_A must support the drain junction for the supply voltage.

For typical values

$$N_A = 1.5 \times 10^{17},\ V_{DD} = 5V,\ L_{\min} = 0.18 \mu m$$

Other proposals have been suggested based on maintaining long channel characteristics. Ratnakumar and Meindl, [12] set the criterion that the change in threshold voltage be less than 10%, while Brews *et al* [13] set a 10% relative limit on increased drain current at $V_G = V_T$, as V_D is increased.

These approaches assume that the fundamental structures remain as they are. The expected limit beyond which radical change is necessary is the 1Gbit device whereupon physical limits and quantum mechanical effects will be significant.

1.3.3. Non-Scaleable Constants

Whilst scaling principles may be applied to devices, and judicious selections may provide an operating compromise, some features remain stubborn to miniaturization.

The increased complexity in topography has focused attention on interconnect. The maze of wiring can now surpass the chip area several times over. As shown in table 1.2, the *RC* time constant does not scale. Therefore, as chip sizes and layer numbers increase, the *RC* product becomes unacceptably large. This problem has been attacked through the use of alternative materials, most notably in the area of refractory metal silicides [14] . Accompanying this problem is the increase in the number of contacts and contact resistance, the latter increasing by κ^2. This again can be addressed by materials and process development in the first instance.

A common problem for interconnect and device scaling is that of edge-effects. As dimensions reduce and complexity increases, edge-effects play a significantly larger role in performance specification. There are two reasons for this.

(i) Firstly, an increase in the number of layers produces awkward topography, through which the interconnect tracks must route. This increases the unwanted parasitic capacitances. Also, since it may not be possible to reduce

the thickness of the film due to step coverage constraints [15] and the length of interconnect may increase rather than decrease, the edge electric fields form a larger percentage of the total electric field. In this case thoughtful layout can improve device performance.

(ii) Secondly, as line/space pitch is reduced the transition or overlap between the two remains constant. The effect of scaling by κ, then, is to increase the region of transition by κ also.

This can be seen in figure 1.3 where the pitch, p, is scaled to p' by the scale factor κ. The transition region s does not scale and therefore, since,

$$p' = \kappa p$$

then

$$s/p' = \kappa s/p$$

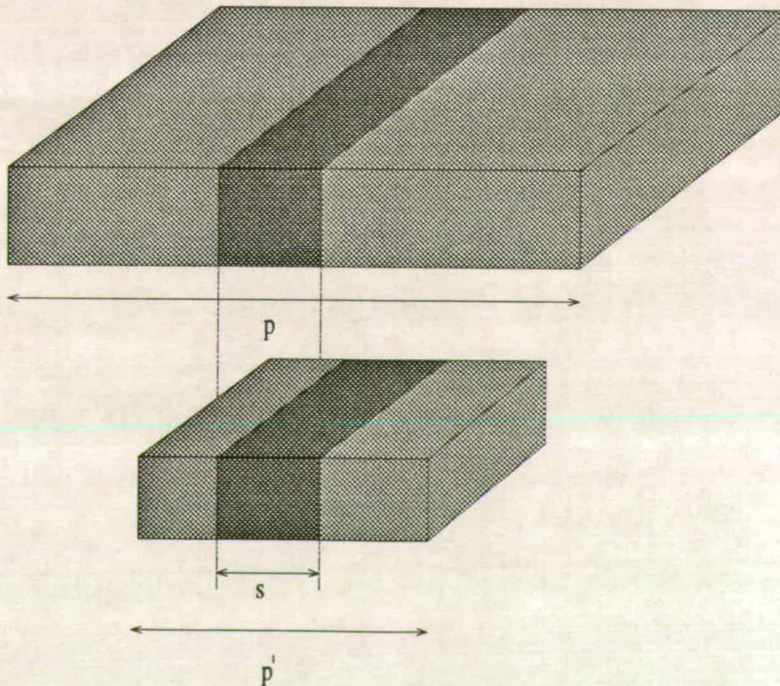


Figure 1.3. Effect of Scaling.

At some point, this increase will become intolerable and a new technology will be

forced. In the MOS transistor structure, this effect is seen in two places. Firstly, the gate-drain overlap remains fixed for a geometric scaling. The theoretical reduction limit here is constrained by twice the depletion depth as discussed in section 1.2.2. Secondly, the active area-isolation transition also remains fixed. This region of the device defines the width of the MOSFET and has till recently been regarded as a second order effect compared to the gate length parameters. With the advent of VLSI this can no longer be the case and forms the subject matter for this thesis.

1.4. Thesis Outline

1.4.1. Thesis Motivation

As the device dimensions have reduced, the feature sizes have become comparable with that layer thickness. This has had a three-fold impact on fabrication.

(i) The complex topography now faces difficulties in processing, requiring planarization and edge tapering.

(ii) The measurement of the feature size is difficult since the edge of the feature is no longer clearly defined.

(iii) The steps and corners produced during fabrication can result in high electric fields as these points and edges mirror the lightning rod phenomena, acting as a focus for electric field lines.

The last of these three is perhaps the most subtle, since the effect cannot be seen till the process is complete. Since many features are now measured electrically, these edge effects must now also be examined.

The surface topography examined herein is that produced at the initial stage of processing, impacting single transistor performance and defining the topography of the first layer. The process studied is the LOCOS process. This was invented in 1970 and despite a plethora of alternatives is still in use even at small geometries.

1.4.2. Thesis Content

Chapter 2 reviews isolation technologies currently available and discusses the LOCOS processing parameters and their impact on the bird's beak profile.

Chapter 3 discusses the basic device physics, focusing on the parameters which are chosen to characterise a MOST.

Chapter 4 reviews state-of-the-art work in the area of width parameter extraction. The impact of the constituent process steps which define the device in cross section are also examined.

Chapter 5 describes the experiment and presents the topographic profiles obtained by SEM cross-section from actual devices. Both process simulation and device simulation results are presented. A novel electrical parameter extraction technique is proposed to permit direct comparison with the simulation results.

Chapter 6 contains the results of electrical measurement. The effective width of the MOSFET is examined with respect to processing parameters and its sensitivity to gate voltage. Detailed analysis of the electrical techniques is performed.

Chapter 7 is presents the conclusions and a final discussion of the results.

References

1. H.I. Sharlin, in *The Making Of The Electrical Age*, Aberlard-Schuman Press.

2. J. J. Staller and A. H. Wolfson, “Miniaturisation In Computers”, in *Miniaturisation*, ed. H. D. Gilbert, Rienhold Publishing, New York.

3. J. Nishizawa, “The Transistor - A Look Back And A Look Ahead”, *Solid State Technology*, Dec 1987.

4. J. Bardeen, *Solid State Technology*, Dec 1987.

5. J. S. Kilby, “The Invention of the Integrated Circuit”, *IEEE Trans. On Electron Devices*, vol. Vol ED-23, no. 7, July 1976.

6. D. Kahng, “A Historical Perspective On The Development Of MOS Transistors And Related Devices”, *IEEE Trans. On Electron Devices*, vol. ED-23, no. 7, July 1976.

7. P. K. Wiemer, *The TFT - A New Thin-Film Transistor,* Holt, Rheinhart and Winston Inc, New York. Reprinted in "Micro and Thin Film Electronics"

8. J. Handelman, *Semiconductor International*, Jan 1991.

9. S. M. Sze, in *VLSI Technology*, McGraw-Hill.

10. J. Meindl, “Opportunities For Gigascale Integration.”, *Solid State Technology*, Dec 1987.

11. “DRAM Manufacturing In The 90’s”, *Semiconductor International*, May 1991.

12. K. N. Ratnakumar and J. D. Meindl, "Short Channel MOST Threshold Model", *IEEE Journal of Solid State Circuits* , vol. SC-17 (5), Oct 1982.

13. W. Fichtner, E. H. Nicollian, S. M. Sze, and J. Brews, "Generalized Guide for MOSFET Miniaturization", *IEEE Electron Device Letters*, vol. EDL -1,2, 1980.

14. S. Murarka, in *Silicides for VLSI Fabrication.*, Academic Press, New York.

15. R. T. Bate, "Nanoelectronics", *Solid State Technology*, Nov 1989.

Chapter 2

Isolation Processing

2.1. Introduction

The isolation regions in any chip are essentially wasted space and it is therefore highly desirable to reduce these areas. The function of the isolation region is to prevent cross-talk between transistors. In practice however, the demands made on the technology are high. Ideally all of the following must be satisfied:

(i) Negligible leakage current between active devices

(ii) Minimal 'wasted' space

(iii) Controllable process

(iv) No impact on active device characteristics

(v) Minimal mask offset.

(vi) In some cases, temperature and radiation hard.

Traditionally, two approaches have been used. One of these uses junction isolation and the other oxide isolation. The former was favoured in bipolar technology and the latter in MOS. In MOS, the leakage paths are primarily localized at the semiconductor surface, so the oxide isolation is preferable. Combinations of these two techniques are increasingly used, in both bipolar and CMOS technology and with the advent of BiCMOS processes this trend has continued.

There are basically two current approaches to VLSI isolation. One provides an oxide thick enough to prevent surface inversion and the other is trench isolation which is analogous to the junction isolation process, whereby deep trenches in the silicon are filled with a dielectric. This more recently developed process has, in solving some problems, introduced others of its own, particularly in reducing these leakage paths round the interface.

2.2. Oxidation

Thermal oxidation of silicon was first proposed in 1959 by Atalla [1] for use as a gate dielectric, though it was 5 years later before the problematic ionic contaminants were identified and eliminated to establish the process as an industrial concern [2] . The mechanism of the oxidation process was examined by Deal & Grove [3] . They produced a model based on equating the three fluxes as shown in figure 2.1: the flux from the gas to the surface, F_1, the flux across the oxide F_2, and the flux corresponding to the reaction at the interface, F_3.

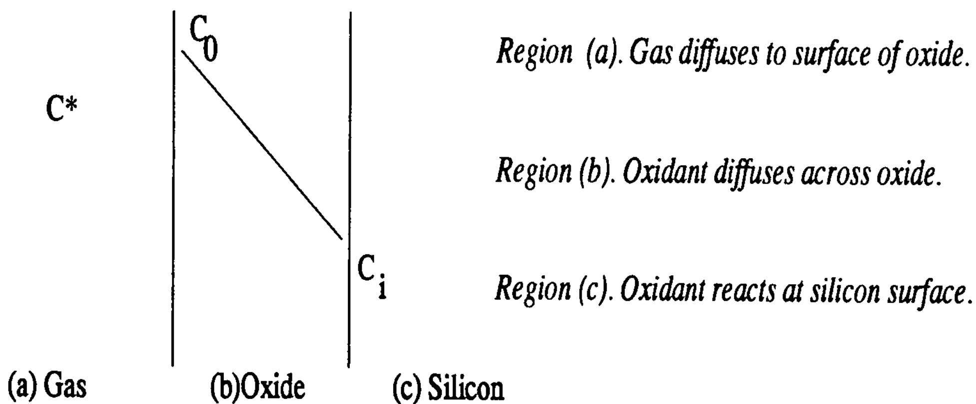


Figure 2.1. Deal-Grove Model.

Three initial equations were used.

$$F_1 = h(C^* - C_0) \tag{2.1}$$

$$F_2 = -D_{eff}(\frac{dC}{dx}) \tag{2.2}$$

$$F_3 = k_s C_i \tag{2.3}$$

where

h is the gas-phase mass transfer coefficient

C^* is the concentration of the oxidant in the gas

C_0 is the concentration of the oxidant in the outer surface of the oxide

D_{eff} is the effective diffusion coefficient of the oxidant in the oxide.

$\frac{dC}{dx}$ is the concentration gradient of the oxidizing species in the oxide.

C_i is the concentration of oxidant in the interface

k_s is the surface rate reaction constant

Equating these fluxes produced the linear-parabolic oxidation law:

$$x_0 + Ax^2 = B(t + \tau) \tag{2.4}$$

where

x_0 is the oxide thickness grown

t is the time

τ is a fitting factor for thin oxides

B is the parabolic rate constant

B/A is the linear rate constant

B and B/A can be expressed in terms of the parameters in (2.1), (2.2) and (2.3)

This model works well for oxides down to about 25 nm and exponential terms have been added to extend this model for thinner oxides [4] . It is however a one dimensional model and as will be seen the 2D (and 3D) effects play a different but important role in the isolation process.

2.3. LOCOS

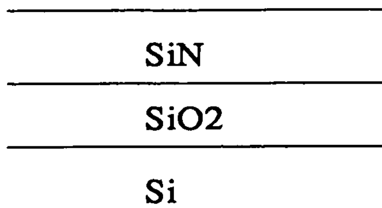
Early isolation techniques used a thick field oxide through which a window was cut for the active area, producing a steep transition between active area and field. This proved problematic later in the process, giving rise to weak spots in the metallization. In 1970, Philips Research Laboratories published a new oxide isolation process which took advantage of the difference in oxidation rates of silicon and silicon nitride. With a rate of oxidation approximately 2 orders of magnitude lower, silicon nitride proved to be an excellent oxidation mask. The process is outlined in Figure 2.2. This process for selective or local oxidation of the silicon regions was given the acronym LOCOS. - for *LOC*al *O*xidation of *S*ilicon. The advantages here are three-fold.

(i) A better window definition is achieved.

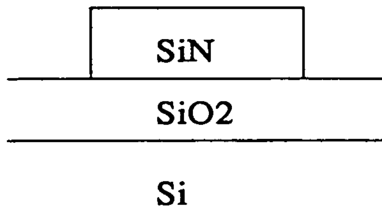
(ii) The field inversion implant is self-aligned to the isolation mask

(iii) Because the oxide is semi-recessed, the thickness can be increased without step cover problems.

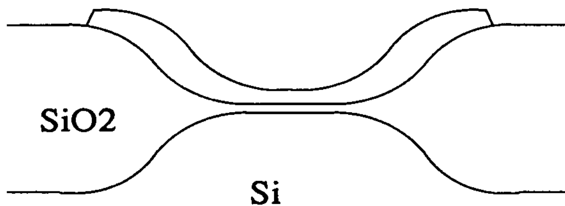
The process proved to be the workhorse for the industry, though with the advent of LSI, inherent problems became more apparent. The transition region between active area and field is not abrupt and is characterised by the 'bird's beak' profile. This arises from the diffusion of oxidant under the edge of the nitride mask and thence to the silicon substrate where oxidation takes place. The bird's beak feature is approximately the same dimension as the thickness of field oxide grown. The rigidity



(a). Silicon thermally oxidised. SiN deposited by CVD.



(b). SiN patterned and etched.



(c). Field Oxide grown.

Figure 2.2. The *LOCOS* process

of the nitride layer can lead to stacking faults in the silicon and enhanced diffusion can occur during the oxidation processes.

2.3.1. White Ribbon

Another feature of the *LOCOS* process was the **white ribbon**. The term was used because through an optical microscope it looked like a white ribbon surrounding the isolation region, extending about 1 μm into the active area [5] . In fact, it turned out to be a thinning of the oxide in the active area. The reduction of thickness of the gate oxide introduced a reliability problem, with this region susceptible to low breakdown.

The mechanism for this phenomena is shown in figure 2.3.

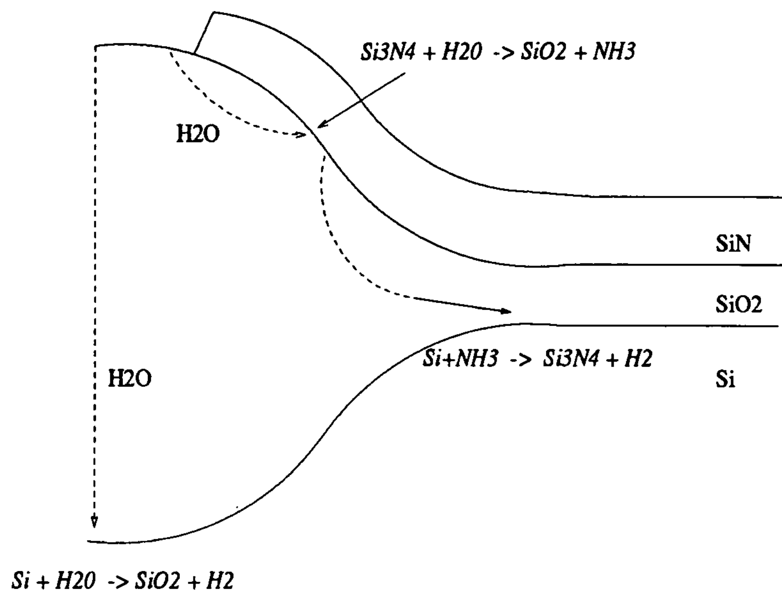


Figure 2.3. White Ribbon Formation.

Diffusion of OH or H_2O from the ambient to the underside of the SiN layer reacts to produce SiO_2 and NH_3. Subsequent diffusion of the NH_3 to the substrate enables the reaction with the silicon to occur and the formation of Si_3N_4 and H_2O. The newly formed SiN at the substrate inhibits oxidation from occuring. Nakajima *et al* [6] showed a similar effect in the body of the active area, away from the bird's beak. Here the NH_3 produced at the surface of the SiN layer was seen to diffuse through the nitride layer itself to produce defects in the substrate. This problem can be overcome in both instances by oxidising in a dry ambient. However, as the field oxide is the thickest thermal oxide grown during the process, the time factor is critical. SiN etching fails to remove this defect consistently and similar problems are seen when the SiN is crystalline as opposed to amorphous [7] .

The solution to the problem is to use a second, wet, sacrificial oxidation performed prior to the gate oxidation [8] .

2.3.2. 2-D Profiles

The importance of the bird's beak transition from active area to field increased as feature sizes became comparable to layer thicknesses. Using a similar approach to the Deal-Grove 1-D model, Wu *et al* [9] developed a 2-D equation relating processing parameters to the topographic profile of the bird's beak. Figure 2.4 shows the LOCOS profile with the parameter definitions.

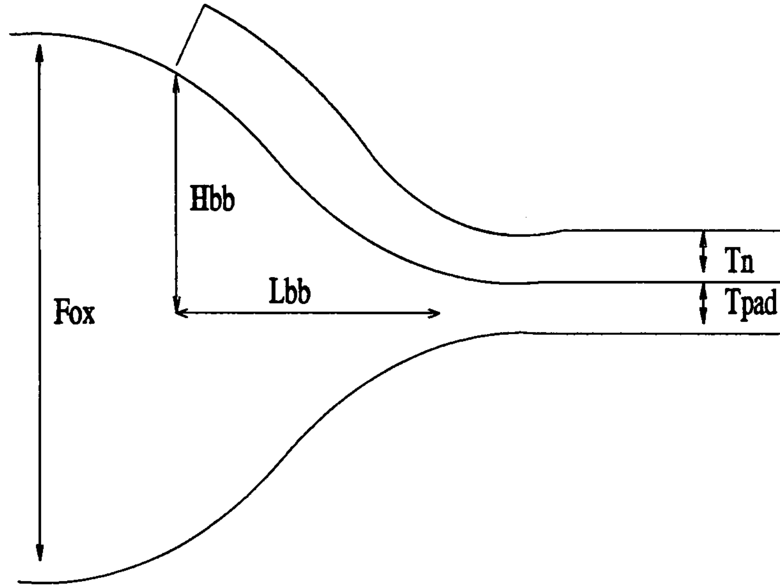


Figure 2.4. *LOCOS* dimensional parameters

Equation (2.5) shows an exponential decay of bird's beak height against distance into the active area. Five processing parameters can be examined as to their individual or combined impact on the profile.

$$\frac{R(y,t)}{R_o(t)} = \frac{C_s(t)}{C_i(t)}\left[\frac{1}{\left(1+T_{pad}\dfrac{k_s}{2D}\right)}\right]\exp(-\Gamma y) \tag{2.5}$$

where

$$\Gamma = \left[\frac{k_s}{T_{pad}D}\,\frac{1}{1+\dfrac{T_{pad}k_s}{2D}}\right]^{1/2} \tag{2.6}$$

As can be seen, similar terms are present in this model as in the 1-D Deal-Grove model, with rate limiting terms such as surface rate reaction, diffusion coefficient and concentrations appearing. The left hand side relates the ratio of oxidation rate at a distance y along the surface of the silicon, under the nitride mask, to the oxidation of the unmasked region. The equation predicts an exponential decay of the growth rate under the nitride mask. This model proved successful but demonstrated the complexity of the processing conditions. An alternative approach was used by Guillemot *et al* [10] who began with experimentally determined profiles and fitted equations to these using empirical parameters. In so doing, another processing parameter was included: the thickness of the silicon nitride. The nitride film is usually a dense, rigid film. The expanding oxide under the delineated edges lifts the nitride resulting in the bird's beak. A thick nitride film will suppress this but too thick a film will crack. As may be expected with such a large number of process variables, trade-offs must be made. The nitride film must be thick enough to act as an oxidation barrier. Thick nitride also acts as a diffusion barrier to the NH_3 responsible for the defects in the gate oxide [6] and further reduces the bird's beak.

Using this approach the experimentally observed "step" in the silicon surface as a result of the field oxidation could be included. These sharp corners could not be predicted from the initial analysis by Wu [9] since no nitride stress effects were included in the model. The topographic step, or kink, is seen according to the T_{pad}/T_n conditions and may impact the electrical performance of the device.†

Another subtle effect which may only be evident by its consequences on electrical characteristics is the stress in the nitride film which produces dislocations and defects in the silicon crystal. This is produced, not by the difference in thermal expansion of the silicon and silicon nitride, but by coulombic forces arising from the

† Discussed in Chapter 4.

mismatch of the two layers at their interface [11] . This will impact dopant diffusion and consequently produce uncertainty in the final device performance. The presence of the intermediate pad oxide in the silicon - silicon nitride sandwich acts as a stress relief layer to alleviate this problem. Further, direct deposition of nitride onto silicon can result in doping of the surface with n-type material during the nitride removal in H_3PO_4 [12] .

Guillemot's motivation in this approach was to produce an analytical model for computer simulation. Thus a defined range of interest of input parameters was necessary to produce a look-up table. The sole advantage over the numerical models available is in the CPU time. However the disadvantages is in the limited range of usefulness. Current numerical simulation models incorporate stress effects of both nitride and oxide. The evaluation of stress in the nitride layer can be measured across the wafer according to the amount of wafer bending, but this value can increase dramatically when the wafer is patterned, resulting in different, localised stress values.

Oxide has till now been modelled as a non-viscous, incompressible material. However it has been recognised that the oxide may flow at temperatures around 950°C [13] . Viscous flow will relieve the stress in the oxide layer and reduce the number of point defects introduced into the substrate and the subsequent impurity diffusion. Point defect diffusion is orders of magnitude greater than substitutional diffusion and consequently must be minimised, to allow control at small geometries.

The number of input processing parameters which influence the bird's beak profile means that a lot of experimental work is necessary to produce a definitive calibration table. Table 2.1 shows the qualitative interactions between the major parameters

Parameter	$T_{pad}\uparrow$	$T_N\uparrow$	$k_s\uparrow$	Temp ↑	$F_{ox}\uparrow$	Press ↑
$\frac{L_{bb}}{F_{ox}}$	↓	↑	↓	↓	↓	↑

Table 2.1. Effect of process parameters on bird's beak.

2.3.3. 3-D Profiles

The compressive pressure of the nitride mask, which causes the retardation of the oxide growth below the mask edge, can be seen to vary according to the pattern etched in the film. Van der Plas *et al* [14] demonstrated possible geometrical cases whereby the bird's beak profile will vary across the circuit. These are identified as

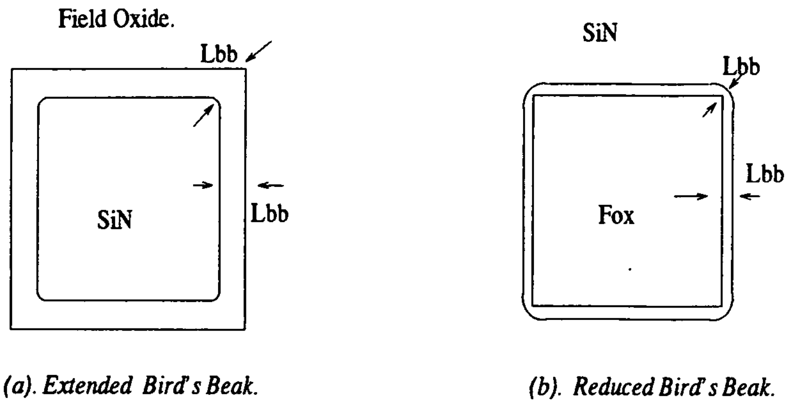
(i) Convex corners where the oxidant can diffuse from a greater solid angle thus increasing the oxidation rate at this corner.

(ii) Concave corners where the solid angle open to the oxidant is reduced causing a reduction in the bird's beak.

(iii) Narrow mask features which are more resistant to bending thus reducing the beak length

(iv) Narrow mask openings, which were also studied by Mizuno *et al* [15] and results in a field oxide thinning but maintains the length of the beak.

Hand-in-hand with the effect of reducing the solid angle open to the oxidant is an increased resistance of the mask to bending. These results are seen in the hole and island structures studied by [16] and are shown in figure 2.5. One consequence of this is that packing density will be severely impacted, since the extension of the bird's beak at the corner may be 2-3 times that of a long mask edge.



(a). Extended Bird's Beak. *(b). Reduced Bird's Beak.*

Figure 2.5. Geometry Dependent Bird's Beak

A final parameter impacting this feature is the etch back of the oxide after field oxidation. Since this is usually done with a wet isotropic etch, etching beyond the surface of the silicon can be rapid, and can have serious consequences on the electrical characteristics of the transistor. Thus a compromise must be reached, whereby the etch-back leaves a small bird's-beak in the wide openings to avoid over-etching the narrow openings.

2.4. LOCOS Alternatives.

2.4.1. General

The drive for alternatives was motivated by the length of the bird's beak. Since this feature was approximately $1\mu m$, the drawn active area size would of necessity, have to be larger to achieve the dimension on the wafer. i.e. incorporating $1\mu m$ loss of feature size on either side of the active area, set a $+2\mu m$ offset on the drawn dimension. Further, as shown by Mizuno [15] oxidation windows below $1\mu m$ showed a sharp attenuation of field oxide thickness. Therefore, on purely topographical grounds the minimum pitch must be $3\mu m$ greater than the desired final active area width. Since this $3\mu m$ serves no other function than isolating active devices, it is seen as an undesirable overhead.

Because of these limitations other approaches have been suggested. The option of producing a planar process was an attractive one, since the subsequent layer depositions would avoid step coverage problems, and eliminate potential weak spots. The alternatives can be classified into 2 main groups. The first of these consists of a silicon etch prior to field oxidation, whereas the second uses a polysilicon buffered layer which can be selectively oxidised to produce the field oxide.

2.4.2. Silicon Etch Processes

The option of etching the silicon prior to field oxidation to provide a more planar structure was presented by Appels *et al* in the original LOCOS paper [17] . This opened the possibility of creating mesas, either by silicon etching or by successive oxidation and oxide etching. Bassous *et al* [12] studied both recessed and semi-recessed oxides with a view to planarising the surface and minimising the bird's beak. Optimum conditions were seen to be a 5 nm pad oxide and a 200 nm nitride film when using fully recessed oxide. The silicon etch process used *KOH* to anistropically etch the substrate to leave the sidewalls exposed to the field oxidation. Thus, although the bird's beak was minimised the characteristic profile remained. Chui *et al* [18] employed a similar approach, using plasma etching to recess the silicon substrate. The *S*ide*W*all *M*asked *I*solation technique[19] eliminated the white ribbon problem and produced a more planar surface. In this case the exposed silicon sidewalls are oxidised and coated with a second silicon nitride and planar etched to leave the pad oxide and nitride on the active areas and sidewalls before undergoing field oxidation. However this structure is susceptible to the dislocations symptomatic of the thick nitride structures previously discussed. Since the second nitride film coats the sidewall, the thickness of the film perpendicular to the substrate is equal to the size of the step ($\approx 330nm$). Volume expansion due to oxidation under these sidewalls encourages stress defects in the silicon substrate.

The problem was attacked by sloping the sidewalls, in a controlled etch process, thus reducing the effective thickness of the second nitride perpendicular to the oxide growth. However, by requiring a slope in the silicon prior to field oxidation, the minimum drawn isolation width will be constrained by the etched depth and the angle of the slope.

The *S*ealed *I*nterface *L*ocal *O*xidation, (*SILO*), tackles the bird's beak problem by depositing nitride directly onto the silicon surface before the pad oxide and masking nitride. Silicon etching recesses the substrate prior to field oxidation. Excellent planarity is achieved and the bird's beak is eliminated [20] . However, the stability of the active area has still to be proved.

In each of the above cases, the presence of *SiN* is required to eliminate the bird's beak. However, the nature of the film is conducive to stress dislocations in the substrate, which may increase greatly when patterned and etched [18, 21] . These defects give rise to

(i) Excessive junction leakage

(ii) Device degradation

(iii) Reliability problems by providing deep recombination centres

Trench, or BOX, isolation uses the silicon etch but does not incorporate silicon nitride in the process. Here the isolation region is back filled once the exposed silicon surface has been sealed with a dielectric and the entire structure is etched back to from a planar surface [22] . This technique poses 2 major problems. One is in the stray leakage paths along the trench sidewalls. The other is in satisfactory planarising the range of topographic profiles which are present over the circuit.

Figure 2.6 shows the 2-d profiles of these structures before field oxidation.

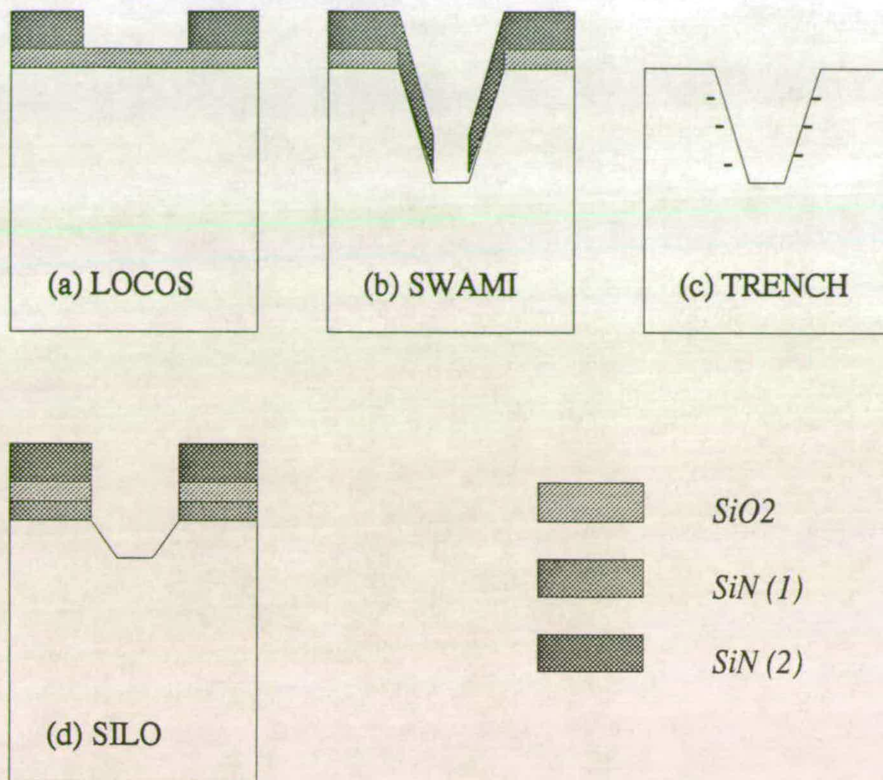


Figure 2.6. Isolation Profiles.

2.4.3. Poly Buffered Processes.

Using a nitride film to reduce the presence of the bird's beak produces stress along the perimeter of the active area possibly causing defects which may lead to junction leakage and/or low breakdown. The use of a polysilicon 'sandwich' layer between the silicon oxide and silicon nitride has been proposed as a way of minimising these effects [23] . Figure 2.7 shows the generic process for which there are two alternatives. In the first instance the polysilicon film is thick enough to provide the entire field oxide. This leaves a non-recessed oxide and corresponding large steps in the transition region. The stress effects of the nitride film are then absorbed in the oxidised polysilicon instead of the silicon substrate. Consequently a thicker nitride film can be used, to minimise the bird's beak. Alternatively, a thinner polysilicon film may be used and consequently a thinner nitride film, which recesses the field oxide but is more likely to suffer from the stress dislocations associated with

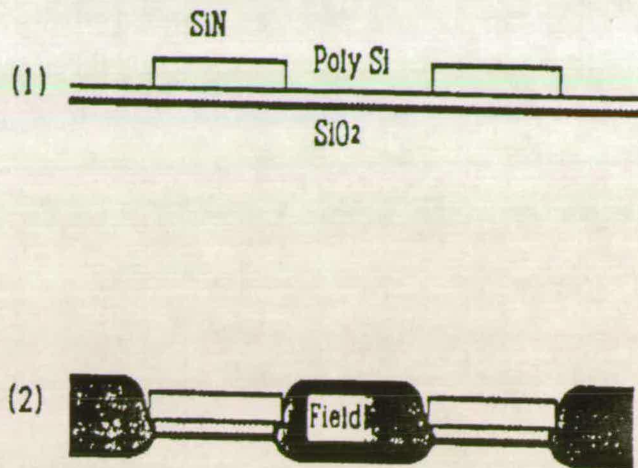


Figure 2.7. Poly Buffered Process.

conventional LOCOS. In neither case is any evidence of the white ribbon effect seen.

Optimised values of the three stack layers [24] suggest values for the pad oxide, polysilicon and nitride as 20 nm, 55 nm, 150 nm respectively, while compromising the restrictions due to stress dislocations arising from the thicker nitride films.

As with conventional LOCOS, the final topographic width is dependent on the etchback prior to gate oxidation. However the thinning of the field oxide must be compensated by a higher channel stop implant. This added etchback can extend the width of the active area by as much as 0.4 μm over the conventional poly buffered process and has been successfully applied to the manufacture of 0.5μm CMOS technolology [25] .

2.4.4. Other Options

Whilst the isolation technologies described above are the most commonly used, they are by no means complete. Oxynitrided layers act as oxidation inhibitors [21] but are prone to problems similar to the SILO structures. Epitaxial silicon has also

been used [26] to form device islands in an approach similar to SOI [27] . Another area of isolation research is in implantation, whereby oxygen is implanted deep in the silicon by high energy implantation to provide an SOI structure [28] .

The common disadvantage that these alternatives suffer is the increased complexity of processing, an option which does not sit well with high volume manufacture. This is reflected in the reluctance for volume manufacture to switch from the LOCOS process. Whilst other options have been presented, the viability of the original LOCOS process is demonstrated by current processes such as that of AT&T which operates a 1.5 μm active area/isolation pitch.

2.5. Summary

Chapter 2 has reviewed the more common isolation processes and highlighted some of the inherent difficulties. The range of options reflects the importance of the isolation process in the I.C. performance. This is of ever increasing relevance as geometries continue to shrink. One of the difficulties in assessing the strategies chosen is the large number of input parameters for any one option. A second problem is need for electrical testing. The effects of isolation technique on the electrical parameters may be difficult to interpret and of course can only be done at the end of process. Thus the process matrix to encompass even a few parameters involves a great deal of further work before devices are ready for electrical characterization. The most popular of the isolation processes, however, is the *LOCOS* process which will be used as the vehicle of study for this thesis.

References

1. M. M. Atalla, *US Patent 3,206,670*, 1960.

2. E. H. Snow, *J. Appl. Physics*, vol. 36, p. 1664, 1965.

3. B. E. Deal and A. S. Grove, "General Relationship for the Thermal Oxidation of Silicon", *Journal of Applied Physics*, vol. 36, no. 12, p. 3770, 3778, 1965.

4. H. Z. Massoud, J. D. Plummer, and E. H. Irene, "Thermal Oxidation Of Silicon in Dry Oxygen - Growth Rate Enhancement in the Thin Regime", *Journal of the Electrochemical Society*, vol. 124, p. 2685, 1985.

5. E. Kooi, J. G. Van Lierop, and J. A. Appels, "Formation Of Silicon Nitride at a Si-SiO2 Interface During Local Oxidation Of Silicon And During Heat-Treatment Of Oxidised Silicon In NH3 Gas", *Journal Of The Electrochemical Society: Solid State Science & Technology*, vol. 23, no. 7, July 1976.

6. O. Nakajima, N. Shiono, S. Muramoto, and C. Hashimoto, "Defects In A Gate Oxide Grown After The LOCOS Process", *Japanese Journal Of Applied Physics*, vol. 18, no. 5, pp. 943-951, May 1979.

7. H. Tompkins, M. Dydyk, and P.W. Deal, "Unintentional Formation Of Etch-Resistant Silicon", *Journal Of The Electrochemical Society*, vol. 137, no. 6, 1990.

8. T. Shankoff, T. T. Sheng, R. B. Marcus, S. E. Haszko, and T. E. Smith, "Bird's Beak Configuration And ELimination Of Gate Oxide Thinning Produced During Selective Oxidation", *Electrochemical Society: Solid State Science And Technology*, pp. 216-222, 1980.

9. T. C. Wu, W. T. Stacy, and K. N. Ritz, "The Influence Of The LOCOS processing Parameters On The Shape Of The Birds Beak Structure", *Journal Of The Electrochemical Society: Solid State Science & Technology*, vol. 130, no. 7, pp. 1563-1556, July 1983.

10. N. Guillemot, G. Pananakakis, and P. Chenevier, "A New Analytical Model Of The Birds Beak", *IEEE Transactions On Electron Devices*, vol. ED-34, no. 5, pp. 1033-1038, May 1987.

11. J. G. Dill, J. W. Barsten, R. D. J. Verhaar, and A. E. T. Kuiper, "LOCOS With Thick And Thin Nitride Masking", *Philips Journal Of Research*, vol. 40, no. 2, 1985.

12. E. Bassous, H. N. Yu, and V. Maniscalco, "Topology Of Silicon Structures With Recessed SiO2", *Journal Of The Electrochemical Society: Solid State Science & Technology*, vol. 123, no. 11, Nov 1976.

13. Eernisse, *Appl. Phys. Lett.*, vol. 30, p. 290, 1977.

14. P. A. van der Plas, N. A. H. Wils, and R. de Werdt, "Geometry Dependent Birds Beak Formation For Submicron LOCOS Isolation", in *ESSDERC*, pp. 131-134, 1989.

15. T. Mizuno, S. Sawada, S. Maeda, and S. Shinozaki, *Oxidation Rate Reduction In The Submicrometer LOCOS Process.*

16. H. Umimoto, S. Odanaka, I. Nakao, and H. Esaki, "Numerical Modelling Of Nonplanar Oxidation Coupled With Stress Effects", *IEEE Trans. Comp.-Aided Des*, vol. 8, no. 6, pp. 599-607, June 1989.

17. J. A. Appels, E. Kooi, M. M. Paffen, J. J. H. Schatorje, and W. H. C. J. Verkuylen, "Local Oxidation Of Silicon And Its Application In Semiconductor Technology", *Philips Research Reports*, no. 25, pp. 118-132, 1970.

18. R. C. Fang, K. Y. Chiu, and J. L. Moll, "Defect Characteristics And Generation Mechanism In A Birds Beak Free Structure By Sidewall Masked Technique", *Journal Of The Electrochemical Society: Solid State Science & Technology*, vol. 130, no. 1, Jan 1983.

19. K. Y. Chui, J. L. Moll, and J. Manoliu, "A Birds Beak Free Local Oxidation Technology Feasible For VLSI Circuits Fabrication", *IEEE Transactions On Electron Devices*, vol. ED-29, no. 4, April 1982.

20. V. K. Dwivedi, "A Birds Beak Free Sealed Interface Local Oxidation Technology For Submicron Ultra Large Scale Integrated Circuits", *Journal Of The Electrochemical Society*, vol. 137, no. 8, Aug 1990.

21. H. H. Tsai, C. L. Yu, and C-Y Wu, "A Birds Beak Reduction Technique For LOCOS in VLSI Fabrication", *IEEE Transactions Electron Device Letters*, vol. EDL-7, pp. 122-123, Feb 1986.

22. G. Zwicker, P. Lang, P. Staudt-Fishbach, and W. Windbracke, "Field Isolation Using Shallow Trenches For Submicron CMOS Technology", in *ESSDERC*, pp. 147-149, 1989.

23. N. Matsukawa, H. Nozawa, J. Matsunaga, and S. Kohyama, "Selective Polysilicon Technology for VLSI Isolation", *IEEE Transactions On Electron Devices*, vol. ED-29, no. 4, April 1982.

24. M. Ghezzo, E. Kaminsky, Y. Nissan-Cohen, P. Frank, and R. Saia, "LOPOS: Advanced Device Isolation For A 0.08μm", *Journal Of The Electrochemical Society*, vol. 136, no. 7, July 1989.

25. J. Hayden, F. K. Baker, S. A. Ernst, R. E. Jones, J. Klein, M. Lein, T. F. McNelly, T. C. Mele, H. Mendez, B-Y. Nguyen, L. C. Parrilo, W. Paulson, J. R. Pfeister, F. Pintchovski, Y-C. See, R. Sivan, B. M. Somero, and E. O. Travis, "A High Performance Half Micrometer Generation CMOS Technology For Fast SRAMS", *IEEE Transactions Electron Device Letters*, vol. 38, no. 4, pp. 876-886, April 1991.

26. Y. C-S. Yu and A. Witkowski, "New Birds Beak Free Device Isolation Technology", *Journal Of The Electrochemical Society: Solid State Science & Technology*, vol. 135, no. 10, Oct 1988.

27. S. M. Sze, *Physics Of Semiconductor Devices,* Wiley & Sons, New York, 1981.

28. K Yallup, "Oxygen Implantation For SOI Device Technology", in *ESSDERC*, p. 533, 1987.

Chapter 3

Device Physics

3.1. Introduction

The workings of semiconductor devices lie in the theories of solid state physics. The fundamental structures of solids determine their electrical properties. The transition from the good conductors (e.g. metals) to good insulators, is the realm of semiconductors. Their ability to conduct electricity to a varying but controllable degree, means that their potential as electronic components is vast. The techniques used in wafer fabrication allow very precise alteration of the semiconductor properties. Understanding the physics behind this necessitates an awareness of the quantised nature of the energy levels, or bands, in the semiconductor. For the duration of this discussion the semiconductor will be silicon, though the qualitative arguments are applicable to any semiconductor.

3.2. The Crystal Lattice

At 0K the silicon substrate is an array of well-ordered silicon atoms, held together by valence electrons forming a pure covalent bond. The electrons are held firmly in place, and charge neutrality is observed. At temperatures greater than 0K, the lattice vibrates due to thermal energy and electrons 'shake' free. If the motions of the electrons can be coordinated, then a flow of current can be achieved.

3.2.1. The Fermi Energy

Quantum physics tells us that the energy levels in an atom are discrete. Only specific energies are permitted within one atomic structure. At 0K the maximum energy of a ground state electron is given by E_F, called the Fermi energy. At temperatures greater than 0K, the energy distribution of electrons is given by the Fermi-Dirac distribution,

$$f(E) = \frac{1}{e^{(E-\mu_c)/k_BT} + 1} \tag{3.1}$$

This equation gives the probability that an orbital at energy E will be occupied. At absolute zero, $\mu_c = E_F$.

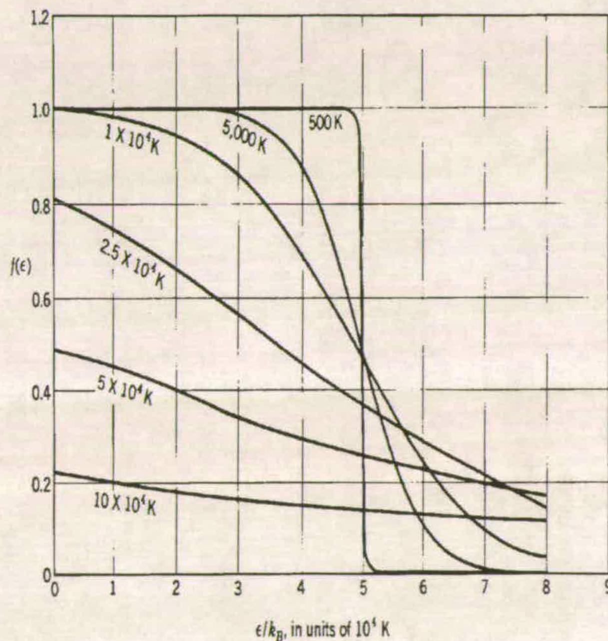


Figure 3.1. The Fermi-Dirac Function [1]

Figure 3.1 shows the Fermi-Dirac distribution function as a function of energy at different temperatures and to a first order in T, we may replace the chemical potential μ_c, by the Fermi energy, E_F. The Fermi energy can be related to the electron

concentration by solving the Schroedinger equation for a fixed volume, V. This results in

$$E_F = \frac{\hbar^2}{2m}\left(\frac{3\pi^2 N}{V}\right)^{2/3} \tag{3.2}$$

which can be re-arranged to give the total number of orbitals of energy $\leq E_F$,

$$N = \frac{V}{3\pi^2}\left(\frac{2mE_F}{\hbar^2}\right)^{3/2} \tag{3.3}$$

The number of orbitals per unit energy range is then given by dN/dE,

$$D(E) = \frac{V}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2} E^{1/2} \tag{3.4}$$

3.2.2. Energy Bands

The concept of energy bands is an important tool in understanding the mechanisms of current transport in semiconductors.† In extending the above to a single crystal lattice, we see that standing waves are set up within the crystal wherein no solutions of the Schroedinger equation are found [1] . Thus forbidden regions or band gaps within the energy distribution are produced, leaving bands of permitted energies on either side. The nature of these energy bands dictates the properties of the material, categorising it into conductor, semiconductor or insulator. The crystal behaves as a metal if one or more of the bands is partly filled (10% - 90%), an insulator if the bands are entirely filled or entirely empty, and a semiconductor if all the bands are entirely filled except for one or two bands slightly filled or slightly empty. In silicon this band gap is 1.12eV.

---

† An alternative approach using *contact potentials* is given in reference 2.

3.2.3. Intrinsic Carrier Concentration

Consider figure 3.2, representing a silicon lattice.

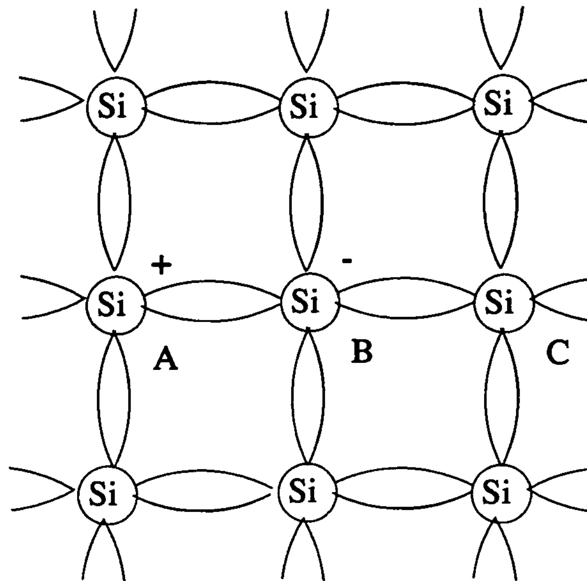


Figure 3.2. The Silicon Lattice.

Each silicon atom is bonded to 4 other atoms, using valence electrons in a covalent bond. At 0K, there are no free electrons and the semiconductor is a perfect insulator. At temperatures above 0K, thermal vibration dislodges electrons which are free to roam throughout the lattice. Atom A is one such atom which has lost an electron. Therefore there is a positive charge associated with A. The vacancy created by the freed electron may now be filled by an electron from a neighbouring atom. This in turn leaves atom B with a net positive charge which may be neutralized by the same process whereby a neighbouring atom, C, donates an electron. Thus we can see that there are two mechanisms for transporting charge around the lattice.

(i) The motion of free electrons, each carrying a single negative charge

(ii) The motion of valence electrons in the opposite direction, from bond to bond, at each stage leaving behind a single positive charge.

It is convenient to think of this second mechanism as the motion of positively charged 'holes' in the opposite direction. (The commonly used analogy is that of bubbles in a liquid). If a hole and electron meet in a lattice then they undergo what is termed recombination, whereby they annihilate each other. In a pure, or *intrinsic,* lattice at room temperature, the concentration of these free carriers is $\sim 1.45 \times 10^{10} cm^{-3}$. The density of silicon is 5×10^{22} atoms cm^{-3}. Therefore only about 3 atoms in every 10^{13} contribute an electron-hole pair. In the intrinsic semiconductor every electron which breaks free is matched by a hole. Let the concentration of electrons be n and the concentration of holes be p. Then since these occur in pairs,

$$n = p = n_i \tag{3.5}$$

where n_i is known as the intrinsic carrier concentration.

3.2.4. Law Of Mass Action

It is convenient to calculate this parameter n_i, in terms of the energy levels, E_c, E_V and μ_c, the chemical potential. In semiconductor physics μ_c is called the Fermi level, E_F. First we calculate the number of electrons per unit volume excited to the the conduction band. This is given by n where,

$$n = \int_{E_c}^{\infty} D(E) f(E) dE \tag{3.6}$$

$D(E)$ is given from (3.4) and $f(E)$ is the Fermi-Dirac distribution function (3.1). The integral therefore becomes

$$n = \int_{E_C}^{\infty} 4\pi \left[\frac{2m_e}{h^2} \right]^{3/2} E^{1/2} \frac{1}{1 + e^{(E-E_F)/k_B T}} dE \tag{3.7}$$

where m_e is the effective mass of the electron.

At temperatures of interest, where $E - E_F \gg k_B T$, the Fermi-Dirac distribution

reduces to

$$f(E) \approx e^{-(E-E_F)/k_BT} \tag{3.8}$$

The integral is easily evaluated to

$$n = N_c e^{-(E_C-E_F)/(k_BT)} \tag{3.9}$$

where

N_C is the effective density of states in the conduction band

and

$$N_C = 2\left[\frac{2\pi m_e k_B T}{h^2}\right]^{3/2} \tag{3.10}$$

Because a hole is the absence of an electron, the distribution function for holes can be written as

$$f_h = 1 - \frac{1}{e^{(E_F-E_V)/k_BT} + 1} \tag{3.11}$$

Using a similar approach then, yields the hole concentration in the valence band

$$p = N_V e^{-(E_F-E_V)/(k_BT)} \tag{3.12}$$

where

m_h is the effective mass of the hole

N_V is the effective density of states in the conduction band

and

$$N_V = 2\left[\frac{2\pi m_h k_B T}{h^2}\right]^{3/2} \tag{3.13}$$

The product of (3.9) and (3.12) is known as the Law of Mass Action

$$np = 2\left[\frac{2\pi k_B T}{h^2}\right]^3 (m_e m_h)^{3/2} e^{-(E_C-E_V)/k_BT} \tag{3.14}$$

This is a useful expression since the result is independent of the presence of impurities. The only assumption is that the material is non-degenerate. viz. the Fermi level is $> 3k_BT$ away from the edge of the conduction or valence band, within the band gap. From (3.5)

$$n = p = n_i$$

Therefore,

$$n_i = 2\left[\frac{2\pi k_B T}{h^2}\right]^{3/2} (m_e m_h)^{3/4} e^{-(E_C - E_V)/k_B T} \tag{3.15}$$

Thus the intrinsic carrier concentration depends exponentially on the temperature. Also since $n = p$, equating (3.9) and (3.12) for an intrinsic semiconductor gives

$$E_F = \frac{E_V + E_C}{2} + \frac{3}{4} k_B T \ln\left(\frac{m_e}{m_h}\right) \tag{3.16}$$

which places the Fermi level very close to the middle of the band gap for the intrinsic semiconductor.

3.2.5. Impurities

Because the product of electrons and holes is a constant, introducing added quantities of one must reduce the other. This simple observation has profound impact in practice. It possible then to enormously reduce a carrier concentration by addition of the opposite type. This phenomenon is known as **compensation**. Figure 3.3 demonstrates the mechanism of doping of a crystal lattice.

Figure 3.3(a) shows the pure silicon case. Figure 3.3(b) shows the presence of an atom (boron) from group III in the periodic table and 3.3(c) the presence of an atom (arsenic) from group V in the periodic table. Since silicon lies in group IV, in neither case will there be a complete matching of bonds. In the case of boron with 3 valence electrons, an extra electron is 'accepted' to form 4 covalent bonds around the

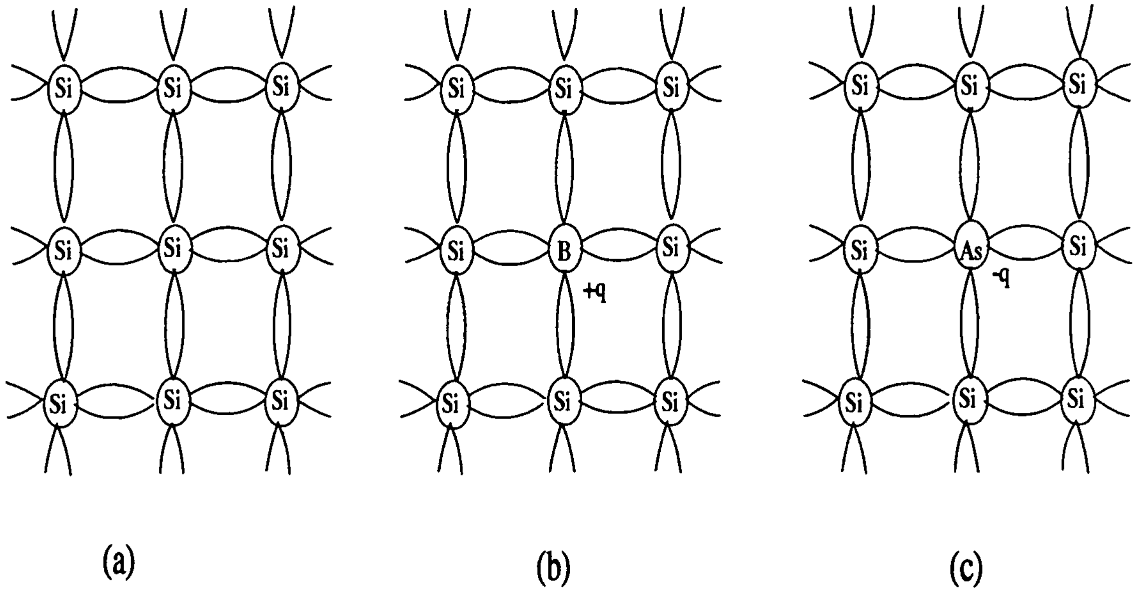


Figure 3.3. The Doped Lattice.

boron atom and a positively charged hole is created in the valence band. The silicon is then p-type because of the addition of a positively charged carrier and the boron is known as an acceptor. In the case of arsenic an electron is lost or 'donated' from the valence band to the conduction band. The silicon is then n-type because of the addition of a negatively charged carrier and the arsenic is known as a donor. Since at room temperature most donors and acceptors are ionised, charge neutrality can be expressed as

$$n + N_A = p + N_D \tag{3.17}$$

In an n-type semiconductor

$$n \approx N_D \tag{3.18}$$

and in a p-type semiconductor

$$p \approx N_A \tag{3.19}$$

Equations (3.9) and (3.12) can be re-arranged to provide the relations

$$E_C - E_F = k_B T \ln\left(\frac{N_C}{n}\right) \tag{3.20}$$

and

$$E_F - E_V = k_B T \ln\left(\frac{N_V}{n}\right) \tag{3.21}$$

which using (3.18) and (3.19) become

$$E_C - E_F = k_B T \ln\left(\frac{N_C}{N_D}\right) \tag{3.22}$$

and

$$E_F - E_V = k_B T \ln\left(\frac{N_V}{N_A}\right) \tag{3.23}$$

The conclusion of this is to shift the Fermi level *relative* to the conduction or valence bands according the to quantity of dopant introduced. Figure 3.4 shows the related energy band diagrams for the intrinsic, p-type and n-type cases.

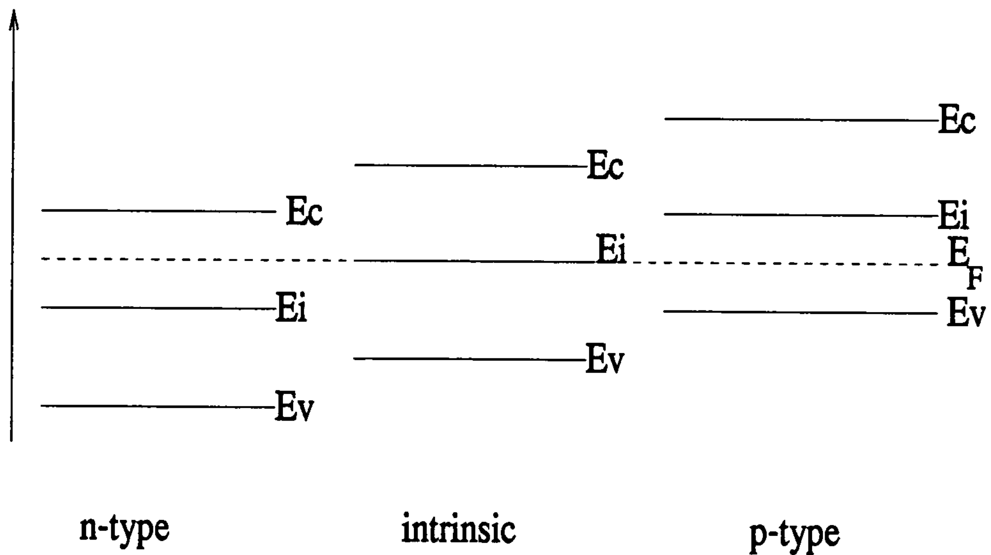


Figure 3.4. Energy Band Diagram for $p-$type, $n-$type and intrinsic silicon.

An alternative approach yields an equivalent expression whereby the concentrations are expressed in terms of the intrinsic carrier concentration and the intrinsic energy level

$$n = n_i e^{(E_F - E_i)/k_B T} \tag{3.24}$$

and

$$p = n_i e^{(E_i - E_F)/k_B T} \tag{3.25}$$

In the limiting case where the Fermi levels are shifted because of excessive carrier doping to a few ($\approx 3k_BT$) of the conduction band, then these equations are no longer valid and the material is said to be **degenerate**.

3.2.6. Conduction

If the motion of the free carriers can be coordinated then a directional current can be achieved. The can be done by applying an electric field E. The motion of the carriers under these conditions is complicated because they accelerate, collide and accelerate again in discrete charge packets. Further, recombination will occur. Generally there are two mechanisms for current flow: drift current and diffusion current.

3.2.6.1. Mobility and Drift velocity

The following discussion assumes no recombination. Assume a current I, is flowing in the semiconductor shown in figure 3.5

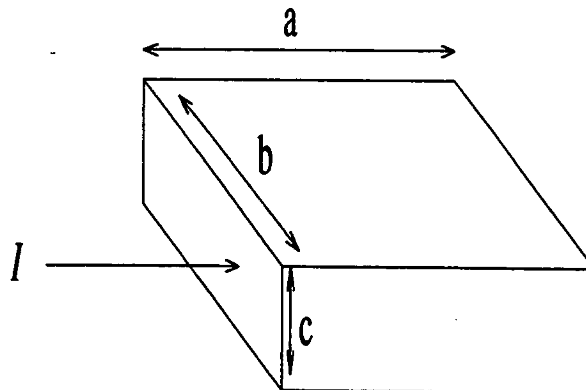


Figure 3.5. Current Flow

The average time taken for a charge to cross the region is defined as τ_t. The motion of the carriers caused by the E field is known as the drift velocity v_d. The value of this velocity depends on

(i) the type of semiconductor

(ii) the type of dopant

(iii) the type of carrier

(iv) the concentration of the carrier

(v) the temperature

When a carrier is accelerated through a high electric field, the subsequent loss of energy to the lattice becomes dominant and the velocity is said to *saturate*. Below this, the velocity is proportional to the electric field and the constant of proportionality is known as the mobility, μ. For silicon a low electric field is $< 0.3V/m$ for electrons, and $< 0.6V/m$ for holes. In general $\mu = aT^{-x}$, where a and x are constants.[2] . Analytic models have been developed to express the reduction of the bulk mobility with the introduction of dopant [3, 4] . The introduction of dopant reduces the mobility according to [2] †

$$\mu = m^{1/2} N^{-1} T^{-3/2} \tag{3.26}$$

The transit time, τ_t, then can be related to the drift velocity.

$$\tau_t = \frac{a}{|v_D|} \tag{3.27}$$

where a is the length of material through which the current flows.

---

†It should be noted that here we have discussed the bulk mobility, μ_B, which is *not* equivalent to the surface mobility μ_s.

Also

$$E = \frac{V}{a} \tag{3.28}$$

Therefore,

$$\tau_t = \frac{a}{\mu E} \tag{3.29}$$

and

$$\tau_t = \frac{a^2}{\mu V} \tag{3.30}$$

Therefore for a fixed voltage, V, and mobility μ, the transit time, τ_t, reduces as the length increases. The dependence is a square law, reflecting the two reasons for this.

(i) As a increases the carriers have further to travel

(ii) As a increases, E, $(= V/a)$, reduces.

The current flowing through the bar is then given by

$$I = \frac{Q}{\tau_t} = nq\frac{abc}{\tau_t} \tag{3.31}$$

Applying (3.30)

$$I = nq\mu\frac{bc}{a}V \tag{3.32}$$

which is equivalent to Ohm's Law where $\left[nq\mu\frac{bc}{a}\right]^{-1}$ is the resistance of the material.

3.2.6.2. Diffusion Current

Where a non-uniform distribution of particles exists, then there occurs a phenomenon known as *diffusion*. i.e. wherever there exists a concentration gradient the motion of the thermally stimulated particles tends towards a uniform distribution over time. This is not due to electric fields. As may be expected, the higher the concentration gradient, the higher the diffusion current. In fact, it can be shown [2] that in the absence of an external electric field,

$$J_n = qD_n \frac{\partial n}{\partial x} \tag{3.33}$$

for electrons and

$$J_p = -qD_p \frac{\partial p}{\partial x} \tag{3.34}$$

where J is the current density, n and p refer to electron and hole concentrations and D is the diffusion coefficient. The total current density made up from the two mechanisms of drift and diffusion can be written as

$$J_n = q\mu_n E + qD \frac{\partial n}{\partial x} \tag{3.35}$$

and

$$J = J_n + J_p \tag{3.36}$$

The diffusion coefficient can be related to the mobility through the Einstein Relation [2] :

$$D_n = \frac{k_B T}{q} \mu_n \tag{3.37}$$

$$D_p = \frac{k_B T}{q} \mu_p \tag{3.38}$$

3.2.6.3. Minority Carrier Lifetime

If the equilibrium is disturbed such that $n_i^2 \neq np$, then the tendency is to return to the equilibrium. This can be seen in the case of an n-type semiconductor under illumination, whereby carriers are generated at a rate dp/dt, equal to the net generation minus the recombination. Recombination processes can be modelled by a parameter known as the minority carrier lifetime (τ_n, τ_p). whereby the nett rate of recombination, U, is proportional to the number of carriers generated under illumination, p_n, minus the number of minority carriers in equilibrium, p_{no} [5] . The constant of proportionality is given by τ_p such that

$$U = \frac{1}{\tau_p}(p_n - p_{no}) \tag{3.39}$$

The processes for recombination are complex, either by direct band-to-band, when an electron in the conduction band recombines with a hole in the valence band, or by trapping between the bands, when an electron or hole capture by a suitable donor or acceptor occurs in the semiconductor. This term must be included however in the current continuity equations

$$\frac{\partial p}{\partial x} = \frac{1}{q}\nabla J_p - U_p \tag{3.40}$$

3.3. The P-N Junction Diode

If a region of silicon is selectively doped with an n-type dopant, and an adjacent region doped with p-type dopant the resultant structure will produce what is known as a **pn junction**. Figure 3.6 shows the schematic.

Here we have assumed that the n-type region is more highly doped than the p-type, and that the transition from one to another is abrupt - the so-called **abrupt** or **step junction**. In this structure electrons will diffuse from the n side across the boundary to the p side. Similarly holes will diffuse from the p to the n region. In so

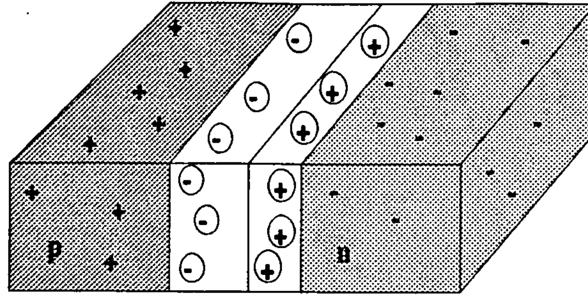


Figure 3.6. *pn* Junction Diode

doing, these carriers leave behind a charged ion of opposite polarity. Therefore the *n* region close to the boundary becomes an area of fixed positive charge. Similarly the *p* region close to the boundary becomes an area of fixed negative charge. As this process progresses, the electric field between the charged ions increases, in such a direction as to oppose the further diffusion of the carriers. Eventually the process will stop when the electric field across the junction balances the diffusion flux. The region within which the electric field builds up is approximated as being completely depleted of carriers, and enclosed within sharply defined edges. This then is called the **depletion approximation**, and the ionised region is known as the space-charge or **depletion region**. The equilibrium condition can be viewed using energy band diagrams. Figure 3.7 shows the energy bands across the *pn* junction.

In equilibrium no current flows between the two regions. Differentiating equation (3.24)

$$\frac{\partial n}{\partial x} = \frac{n}{k_B T}\left[\frac{\partial E_F}{\partial x} - \frac{\partial E_i}{\partial x}\right] \tag{3.41}$$

The nett flux of electrons is given by (3.35). Using the Einstein relationship in (3.37) we see that in the absence of an external electric field, ($E = 0$),

$$J_n = 0 = \mu_n n \frac{\partial E_F}{\partial x} \tag{3.42}$$

Similarly,

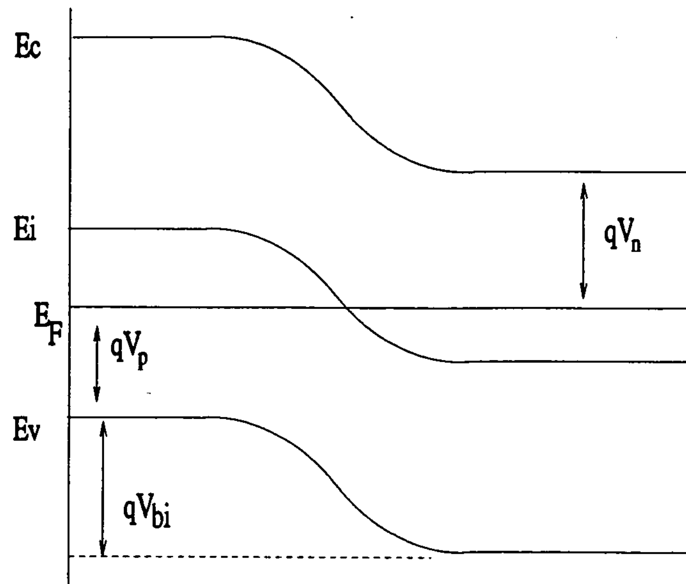


Figure 3.7. Energy Band Diagram For *pn* Diode.

$$J_p = 0 = \mu_p p \frac{\partial E_F}{\partial x} \tag{3.43}$$

Therefore the Fermi level must be constant throughout the regions. The relative positions of the valence and conduction bands can be obtained from eqs (3.22) and (3.23), whereby the *n*-type region shows a shift together of the conduction band and Fermi level as the *n*-type dopant increases, and a similar drawing together of the valence band and Fermi level in the *p*-type region. Crossing the boundary, then, implies a change of potential. This value is given by the difference in the conduction band, or valence band potentials of either side and is known as the **built-in** potential. For the step junction, this can be evaluated from figure 3.8 as

$$qV_{bi} = E_g - \left(qV_n + qV_p\right) \tag{3.44}$$

and can be solved [2] to give

$$V_{bi} \approx k_B T \ln\left(\frac{N_A N_D}{n_i^2}\right) \tag{3.45}$$

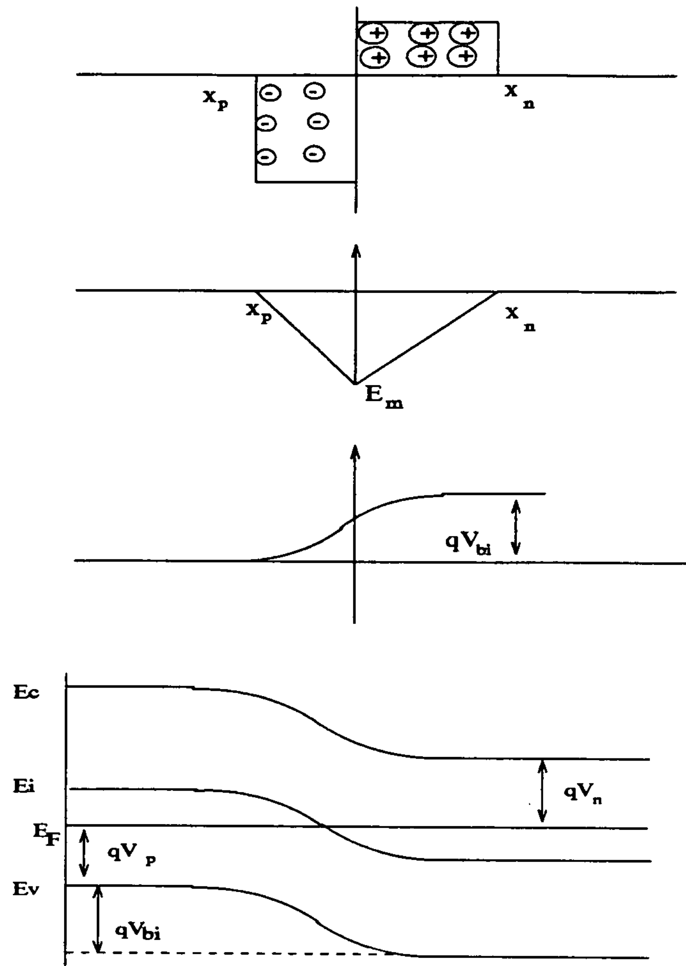


Figure 3.8. Field and Potential Across *pn* Diode.

In thermal equilibrium, the electric field in the neutral regions is zero. Therefore the total charge /unit area on either side of the junction must be equal and opposite. i.e.

$$N_D x_n = N_A x_p \tag{3.46}$$

Applying Poisson's equation gives

$$-\frac{\partial^2 V}{\partial x^2} = \frac{\partial E}{\partial x} = \frac{q}{\varepsilon_s}\left[p(x) - n(x) + N_D(x) - N_A(x)\right] \tag{3.47}$$

which approximates to

$$-\frac{\partial^2 V}{\partial x^2} \approx \frac{q}{\varepsilon_s} N_D \quad \text{for} \quad 0 < x \le x_n \tag{3.48}$$

and

$$-\frac{\partial^2 V}{\partial x^2} \approx \frac{-q}{\varepsilon_s} N_A \quad \text{for} \quad -x_p \le x \le 0 \tag{3.49}$$

Referring to the coordinates of Figure 3.8b, this produces a solution of the form

$$E(x) = \frac{qN_D(x - x_n)}{\varepsilon_s} \qquad \text{for} \quad 0 < x \le x_n \tag{3.50}$$

$$E(x) = -\frac{qN_A(x + x_p)}{\varepsilon_s} \qquad \text{for} \quad -x_p \le x < 0 \tag{3.51}$$

The maximum field E_m is at the point $x = 0$ and is given by

$$|E_m| = \frac{qN_D x_n}{\varepsilon_s} = \frac{qN_A x_p}{\varepsilon_s} \tag{3.52}$$

Integrating the electric field over the depletion region, solves for the potential distribution across the region

$$V(x) = E_m\left(x - \frac{x^2}{2W}\right) \tag{3.53}$$

whence

$$V_{bi} = \frac{1}{2} E_m W = \frac{1}{2} E_m (x_n + x_p) \tag{3.54}$$

where W is the total depletion width, $W = x_n + x_p$.

Eliminating E_m from equations (3.53) and (3.54) gives

$$W = \left[\frac{2\varepsilon_s}{q}\left(\frac{N_A + N_D}{N_A N_D}\right)V_{bi}\right]^{1/2} \tag{3.55}$$

Figure 3.8 shows the charge, electric field, potential distributions and energy band diagrams for this structure. The same analysis (solving Poisson's equation) can be used in the case of the linearly graded junction to arrive at the electric field and potential distributions.

3.3.1. Reverse and Forward Bias

The application of an electric potential across the junction alters the energy band diagram according to the sign and quantity of the voltage. If the *n*-type region is positively biased, then the potential drop across the region is simply the sum of the built in voltage and the applied voltage. As a result of this, the width space charge region on both sides of the junction will increase (in fact the width increases as the square root of the sum of the applied voltage and the built in voltage [2]). Figure 3.9 shows this condition along with the equivalent electric field distribution and energy band diagram.

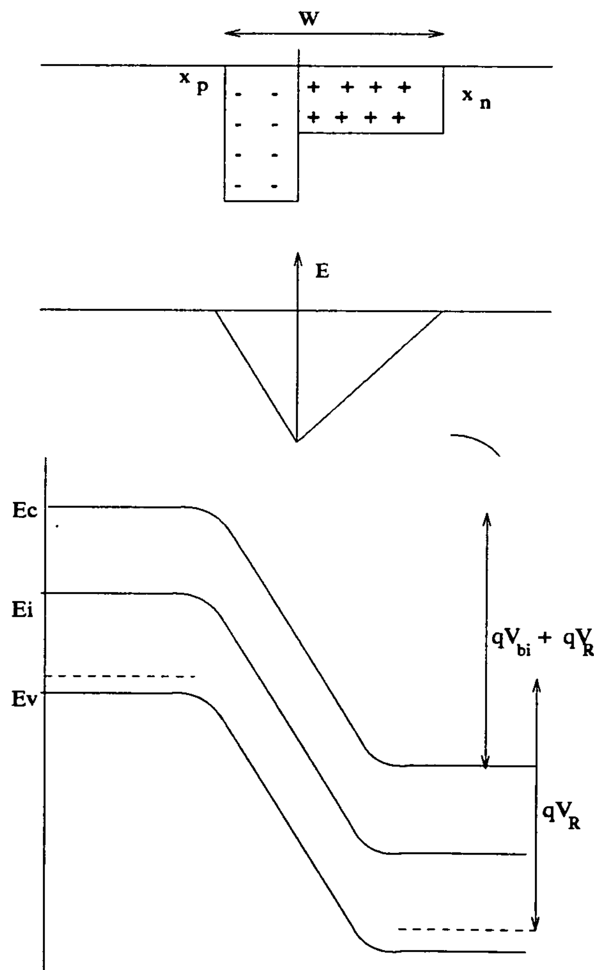


Figure 3.9. *pn* Diode Under Reverse Bias.

Since the junction is no longer in equilibrium, the Fermi level is no longer constant

throughout the silicon. In the limiting case the junction will breakdown. There are three modes of breakdown possible

(i) Thermal Instability

This is particularly relevant in semiconductors with small bandgaps. High reverse bias across the diode, increases the reverse current and consequently the junction temperature. At a critical voltage, this temperature causes an increase in the reverse current producing a negative resistance and the diode will be destroyed.

(ii) Zener (or tunneling)

In this case the electric field is so high as to transfer a valence electron from one side of the junction directly to the conduction band on the other side. In silicon the critical value for this to occur is $10^6 V/cm$ when $N = 5 \times 10^{17}/cm^3$. This form of breakdown may be recoverable.

(iii) Avalanche breakdown (or impact ionization)

This is the more likely breakdown mechanism in regions where the doping density is low. In this case an electron gains sufficient energy from the applied field to cause electron-hole generation on collision with a lattice site. The resulting electron-hole pairs are then swept apart by the electric field whereby the same collisions and results may occur, forming a chain reaction. Thus the junction may not recover when the field is removed.

In silicon *pn* junctions with breakdown voltages less than $4E_g/q$, the mechanism is found to be tunneling and with breakdown voltages greater than $6E_g/q$, avalanche. In between these voltages a mixture of the two breakdown mechanisms is found to occur [2] .

The application of a forward bias acts to reduce the barrier height and the depletion

width. The effect is to build up the carrier concentration to increase the concentration gradient such that the balance with the drift component term in the equilibrium case is no longer observed and a net current flow is created. The *pn* junction is still in existence, though the current flow is opposing the electric field in the space charge region. To determine the value of the current, it is necessary then to determine the concentrations of electrons and holes. At this point it is useful to introduce a new parameter, the **quasi-fermi level** or **imref**. This is defined as the value which, when substituted in place of the Fermi level, gives the concentration of that carrier under non-equilibrium conditions.

3.4. The Field Effect Transistor

The FET is a unipolar multielectrode semiconductor device in which current flows through a narrow conducting channel between two electrodes and is modulated by an electric field applied at a third electrode. FET devices differ from junction transistors in that the current flow is dominated by one carrier instead of two. Within the group of FET transistors, two further classifications are possible. One is the junction FET (also called the JFET), and the other is the insulated gate FET (also called the IGFET).

3.4.1. The Junction FET.

The junction FET passes current along a resistive channel which is flanked on either side by two highly doped regions of opposite type to the channel. Biasing one end of the bar (called the drain end) causes a current flow along this region. The current then can be modulated by application of a bias to the highly doped regions. Figure 3.10 shows a layout of this device.

The current flow can be controlled in either of two ways. In each case a voltage is applied to the drain end to set up a potential difference along the channel. Two

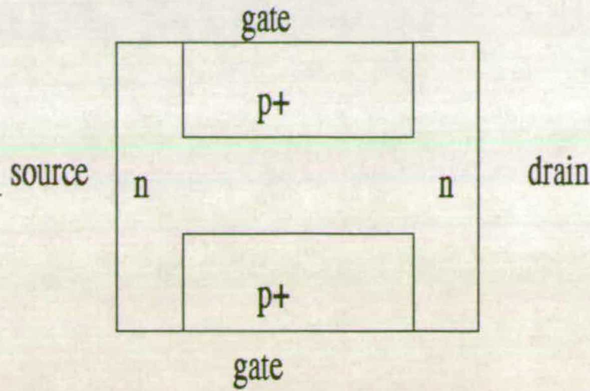


Figure 3.10. The Junction FET.

opposing effects are present. An increased drain voltage will cause an increased current according to Ohm's Law. However, Ohm's Law is only valid with constant resistance. The increase in the drain current will extend the depletion regions around the gate-drain junctions. This has the effect then, of reducing the width of the channel available for conduction at the drain end, which in turn increases the resistance. Thus the resultant current produced by the applied drain voltage will deviate from the linear Ohm's law relationship through a reduction in the expected drain current. At some critical voltage these depletion regions will extend from either side of the gate regions until they meet at the drain end of the channel as shown in figure 3.11.

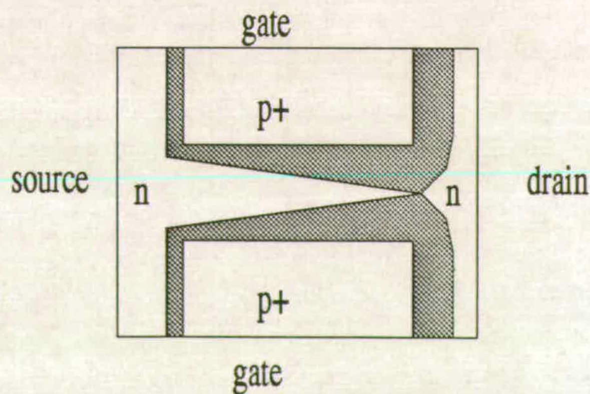


Figure 3.11. Pinch-off In JFET.

This voltage is known as the **pinch-off voltage**, since beyond this value no increase in current is seen. At this point the current is said to **saturate**. The reason for this, is that the extension of the depletion region is largely into the drain region rather than equally into the gate region (the higher of the two concentrations), and the increased potential is dropped across the depletion region, thus limiting the potential at the source end to be equal to the pinch-off voltage. This is analogous to the condition of the abrupt step junction discussed in §3.3. Alternatively the current flow can be modulated by applying a high enough voltage to the gate regions, again extending the depletion regions causing a reduction of the channel width and a 'pinching' of the channel. This has the effect of reducing the 'starting width' of the channel, before increasing the drain voltage to the point of saturation. This condition is shown in Figure 3.12

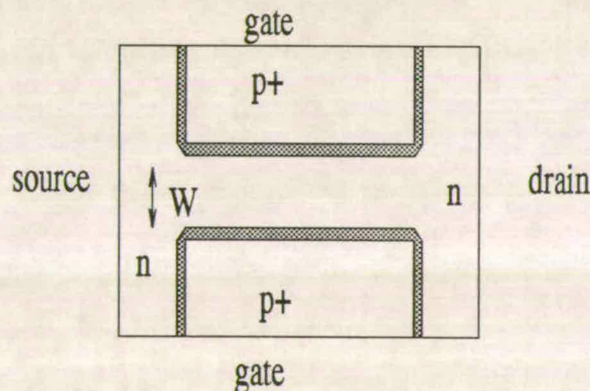


Figure 3.12. Gate Modulation Of JFET.

3.4.2. The IGFET

The basis structure of the igfet is shown in figure 3.13. A semiconductor wafer has two regions of opposite polarity diffused into it to form the source and drain regions. An insulator is formed on the surface between these two regions. The third electrode is placed on top of the insulator, through which a controlling electric field, perpendicular to the channel connecting the source-drain regions, can be applied.

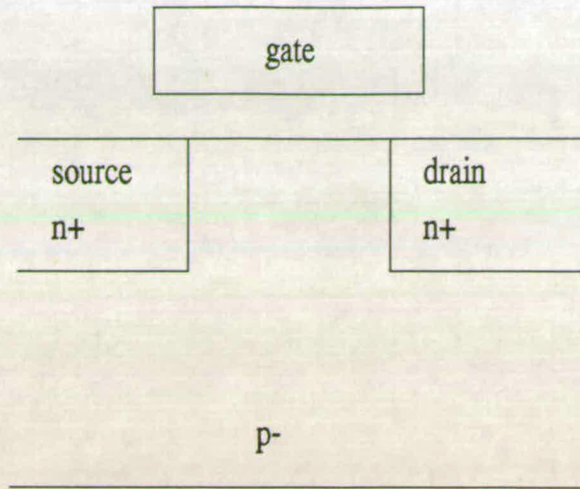


Figure 3.13

This device has several common names currently in use, the origin of which often lies with the technology available at the time: the metal-insulator-semiconductor transistor (*MISFET, MIST*); the metal-oxide-semiconductor transistor (*MOSFET, MOST*); the metal-nitride-oxide-semiconductor (*MNOSFET, MNOST*); In the case of the MOSFET, the insulator is silicon dioxide. Early fabrication processes used metal as the gate electrode material. The introduction of polysilicon as the gate material did not warrant a name-change, and the term MOSFET is perhaps the most widely used (even if the device is actually an MNOSFET). The most general term is the igfet. However the focus of this thesis is on (poly gate)MOSFETs and this terminology will be used hereafter.

3.4.3. The 2 Terminal MOSFET.

The MOSFET can be studied by examining the component structures separately. In the first instance, the 2 terminal structure, or the MOS capacitor, while existing as a separate entity within integrated circuits, forms a functional block within the MOS transistor. The MOS capacitor has been the subject of extensive research [6] and has a history of being a problematic fabrication process [7] . This has arisen because of

the complexity of the semiconductor surfaces and thus the capacitor has proved a valuable tool in analysing these surfaces and interfaces. Figure 3.14 shows the schematic of the structure.

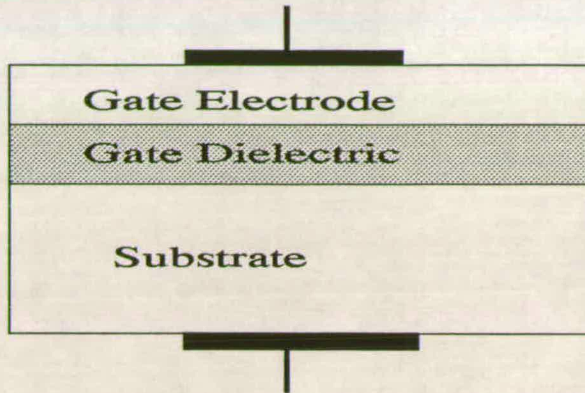


Figure 3.14. The 2 Terminal MOSFET.

The capacitor acts as a switch between the source and drain by modifying the surface type accordingly. The capacitance of the structure is composed of the oxide capacitance plus a contribution from the substrate. The substrate contribution will depend on the voltage applied to the upper gate electrode. Before this is investigated, though, it is necessary to consider the nature of the materials used to fabricate the capacitor.

Work Functions

The work function of a material is the energy required to remove an electron from the fermi level of the material to a vacuum. This in itself is very difficult to measure. However if two materials are brought together, the much more useful parameter, the work function difference, can be used. If the material (in this case a metal), were connected to the substrate then charge would flow till the potential built up would counterbalance the difference in work functions. This term when applied to the MOS device is given the label ϕ_{ms}. Table 3.1 shows some relevant work function differences used in MOS processing. By applying a suitable voltage to the gate

Gate material to Substrate material	ϕ_{ms} $N = 10^{14} cm^{-3}$	ϕ_{ms} $N = 10^{18} cm^{-3}$
Al to n-	-0.36	-0.15
$n+$ to n-	-0.52	-0.35
Al to p-	-0.81	-0.95
$n+$ to p-	-0.96	-1.20

Table 3.1: Work Function Differences in IC Processes.

electrode, the energy bands can be aligned across the insulting gate dielectric. The bending of the bands, which corresponds to a spatial concentration distribution at the surface as opposed to the uniform bulk, is then reduced, until the bands at the surface are brought into alignment with those in the substrate. This voltage, known as the **flat-band voltage,** V_{FB}, is related to the difference of the work functions and fermi levels such that

$$V_{FB} = \phi_m - \phi_s = \phi_{ms} \quad . \tag{3.56}$$

or since $\phi = E/q$

$$V_G = \frac{E_{Fm} - E_{Fs}}{q} \tag{3.57}$$

Oxide Charges

Oxide charges also contribute to the characteristics of the MOS capacitor. These charges can be considered to consist of four components.[6] .

(i) **Oxide fixed charge, Q_f.**

This exists very close to the $Si - SiO_2$ interface due to the mechanism of oxide formation and is found to be independent of oxide thickness or substrate dopant type or concentration.

(ii) **Oxide trapped charge, Q_{ot}.**

This can exist throughout the oxide, but usually at either of the interfaces and can be acquired through radiation, photoemission or by high-energy carrier injection from the substrate.

(iii) **Mobile ionic charge, Q_m.**

Contamination of the oxide in the form of small mobile usually positive ions (commonly sodium or potassium) plagued the early fabrication processes. This has to a large degree been controlled by a cleaner process and by a technique known as *gettering*, whereby the introduction of oppositely charged ions, (notably chlorine or phosphorus) immobilise these charges.

(iv) **Interface trap charge, Q_{it}.**

Also known as fast surface states, these charges exist at the interface and are caused by incomplete bonding of the substrate silicon with the thermally grown silicon dioxide. For this reason, the crystal orientation <100> is often preferred in the substrate silicon as this minimises the concentration of atoms present at the surface to form the silicon dioxide.

The nett effect of these oxide charges on the MOS capacitor can be aggregated in a term Q_o, assumed to exist at the oxide-silicon interface. The number of these effective interface charges is around $5 \times 10^{10}\ cm^{-2}$ in modern controlled processes and is always positive. This is equivalent to a charge density of $0.8 \mu C\ cm^{-2}$

The 2-terminal MOS device can then be viewed as the sum of two series capacitors: the oxide capacitance and the substrate capacitance. Figure 3.15 shows

these schematically.

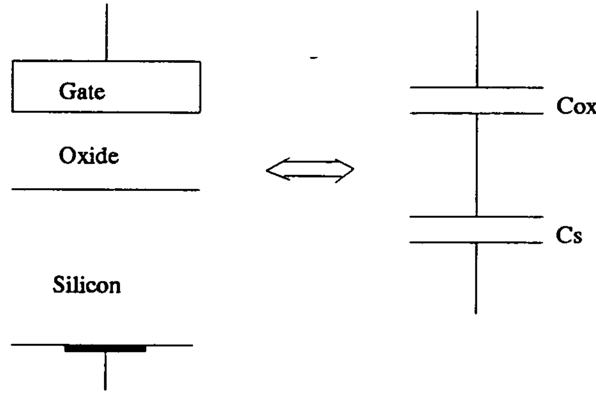


Figure 3.15. Capacitance Model Of 2 Terminal MOSFET.

The oxide capacitance C_{ox}, in the case of a parallel plate capacitor, is given by

$$C_{ox} = \frac{\varepsilon A}{t_{ox}} \tag{3.58}$$

where t_{ox} is the oxide thickness. Defining C' as the capacitance per unit area we can write,

$$C'_{ox} = \frac{\varepsilon}{t_{ox}} \tag{3.59}$$

The parameter C'_s is dependent on the charge distribution in the silicon substrate, which in turn is a function of the applied field. To evaluate this we must solve Poisson's equation for the system (§ 3.4.4).

Consideration of the energy bands, however, allow us to qualitatively examine the substrate conditions, classifying the results into 5 distinct categories. These are shown in the energy band diagrams of figure 3.16 where for the sake of clarity a p-type substrate is assumed. Deep in the bulk the Fermi level is shifted towards the valence band according to (3.23)

$$E_F - E_V = k_B T \ln\left(\frac{N_V}{N_A}\right)$$

The influence of the effective oxide charge Q_o, is to deplete the surface of p-type

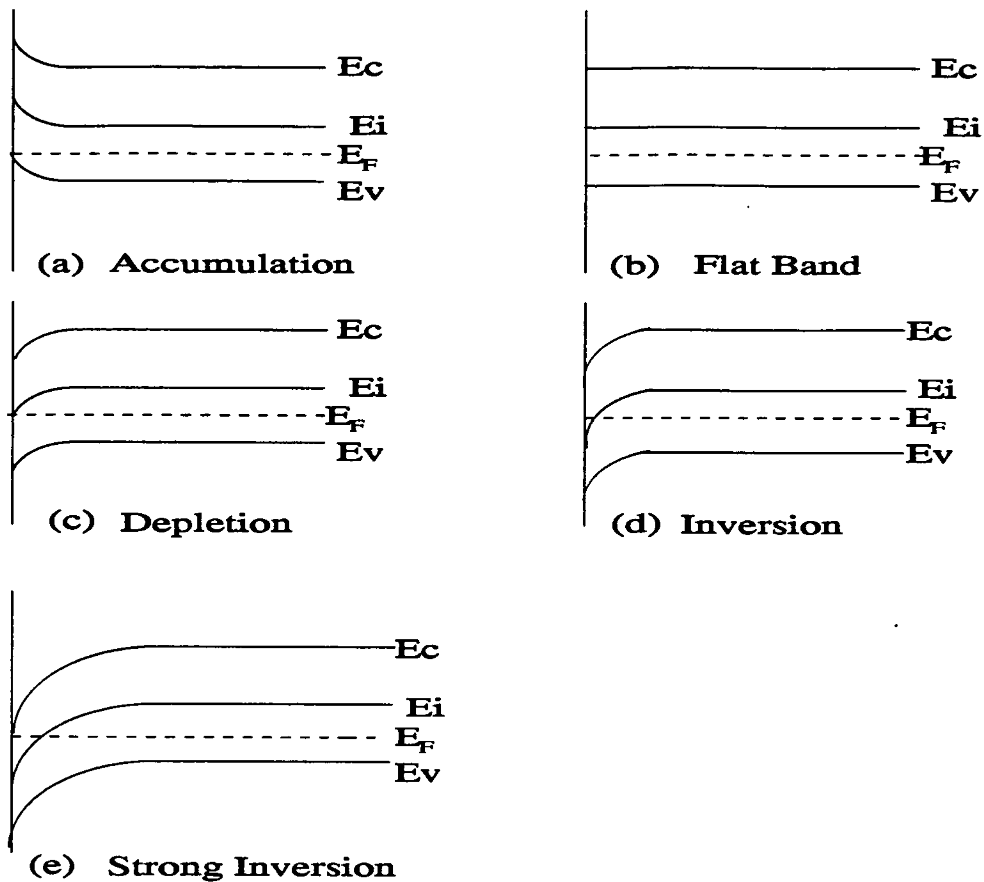


Figure 3.16. Energy Band Diagrams For 2 Terminal MOSFET.

carriers, since the oxide charge is positive. This effect is localised at the surface. Therefore, this depletion is seen as a bending of the energy bands relative to the fermi level, reflecting the reduced concentration of holes in this region. The energy bands of the two electrodes are further misaligned because of the work function difference.

Consider now the 'ideal' case i.e. where there exist no charges in the oxide in the work function difference between the gate and the bulk is zero. The voltage on the gate is then swept from negative to positive. The reaction of the energy bands in the substrate can be easily arrived at.

(i) Accumulation

When the gate voltage is negative, electrons will be driven away from the

surface in favour of the accumulation of holes. Thus the valence band at the surface will bend toward the fermi level again in accordance with (3.23) and shown in 3.16(a)

(ii) Flat band

As the gate voltage swings positive, the opposite effect is seen. In the first instance, the bands bend the other way, until at some point in the transition, the **flat-band condition** is seen as shown in figure 3.16(b), where the flat band voltage is reduced by $\frac{Q_o}{C'_{ox}}$ from that given in (3.56).

(iii) Depletion

Beyond the flat band condition, the surface starts to become depleted of holes as the applied gate voltage becomes more positive. This depletion regime exists while $0 < \psi_s < \phi_F$ where ψ_s is the surface potential (Figure 3.16(c)).

(iv) Inversion

When the intrinsic energy level crosses the Fermi level at the surface due to the applied positive gate voltage, then the onset of weak inversion is said to occur, as the Fermi level is then shifted closer to the conduction band than the valence band, shown by figure 3.16(d). When the intrinsic level crosses the Fermi level by the same amount, ϕ_F, as the bulk then strong inversion† is said to occur. This is shown in Figure 3.16(e).

†This definition is somewhat arbitrary and may vary according to source. Further discussion takes place in Chapter 4, §4.2.

3.4.4. General Analysis Of The Two Terminal MOSFET

A more rigorous analysis of the MOS capacitor can be performed using Poisson's equation. This allows us, through charge conservation, to solve for the potential and electric field as a function of depth in the substrate. This makes use of the equations developed in §3.2.5. viz.

$$n(y) = n_o e^{(E_F - E_i)/k_B T}$$

and

$$p(y) = p_o e^{(E_i - E_F)/k_B T}$$

Using the potential in the bulk as the reference, and defining the **Fermi potential**, ϕ_F, as

$$\phi_F = \frac{E_i - E_F}{q} \tag{3.60}$$

and introducing β,

$$\beta = \frac{q}{k_B T} \tag{3.61}$$

allows the concentrations n and p to be written as

$$n = n_o e^{-\beta\psi(y)} \tag{3.62}$$

$$p = p_o e^{\beta\psi(y)} \tag{3.63}$$

where n_o and p_o, are the bulk concentrations and $\psi(y)$ is referenced to the bulk. The charge density at a distance y in the substrate, including holes, electrons and ionized acceptor ions (for a p-type substrate) is given by

$$\rho(y) = q[p(y) - n(y) - N_A] \tag{3.64}$$

In the bulk the charge density is zero, so $p_o - n_o = N_A$.
Substituting this into (3.64) and using (3.62) and (3.63) gives

$$\rho(y) = q[p_o e^{-\beta\psi(y)} - p_o - n_o e^{\beta\psi(y)} + n_o] \tag{3.65}$$

which reduces then to

$$\rho(y) = q[p_o(e^{-\beta\psi(y)} - 1) - n_o(e^{\beta\psi(y)} - 1)] \tag{3.66}$$

Poisson's equation in 1D states

$$\frac{d^2\psi}{dy^2} = -\frac{\rho(y)}{\varepsilon_s} \tag{3.67}$$

Substituting for the charge density

$$\frac{d^2\psi}{dy^2} = -\frac{q}{\varepsilon_s}[p_o(e^{-\beta\psi(y)} - 1) - n_o(e^{\beta\psi(y)} - 1)] \tag{3.68}$$

In a p-type material $p_o \approx N_A$ and from the Law of Mass Action, $n_o \approx n_i^2/N_A$.

$$\frac{d^2\psi}{dy^2} = -\frac{q}{\varepsilon_s}[N_A(e^{-\beta\psi(y)} - 1) - \frac{n_i^2}{N_A}(e^{\beta\psi(y)} - 1)] \tag{3.69}$$

Equation (3.25) states

$$p = n_i e^{(E_i - E_F)/k_B T}$$

which on substitution of (3.60) and (3.61) becomes

$$p = n_i e^{\beta\phi_F} \tag{3.70}$$

In the bulk,

$$p \approx N_A \tag{3.71}$$

Therefore

$$N_A = n_i e^{\beta\phi_F} \tag{3.72}$$

and so

$$n_i^2 = N_A^2 e^{-2\beta\phi_F} \tag{3.73}$$

Eliminating n_i from (3.69) yields

$$\frac{d^2\psi}{dy^2} = -\frac{q}{\varepsilon_s}[N_A(e^{-\beta\psi(y)} - 1) - \frac{N_A^2}{N_A}(e^{-2\beta\phi_F}(e^{\beta\psi(y)} - 1)] \tag{3.74}$$

which reduces to

$$\frac{d^2\psi}{dy^2} = -\frac{q}{\varepsilon_s} N_A[e^{-\beta\psi(y)} - 1 - e^{-2\beta\phi_F}(e^{\beta\psi(y)} - 1)] \tag{3.75}$$

Multiplying both sides by 2 $d\psi/dy$ and recognising that

$$2\frac{d\psi}{dy}\frac{d^2\psi}{dy^2} = \frac{d}{dy}\left(\frac{d\psi}{dy}\right)^2 \tag{3.76}$$

produces

$$\frac{d}{dy}\left(\frac{d\psi}{dy}\right)^2 = -\frac{q}{\varepsilon_s} N_A[e^{-\beta\psi(y)} - 1 - e^{-2\beta\phi_F}(e^{\beta\psi(y)} - 1)]2\frac{d\psi}{dy} \tag{3.77}$$

Integrating from a point deep in the bulk where where $\psi = 0$, and $\frac{d\psi}{dy} = 0$ to a point y,

$$LHS = \int_{\infty}^{y} \frac{d}{dy}\left(\frac{d\psi}{dy}\right)^2 dy$$

$$= \left[\left(\frac{d\psi}{dy}\right)^2\right]_{\infty}^{y}$$

$$= \left[E^2\right]_{\infty}^{y}$$

$$= E^2(y) - E^2(\infty)$$

$$= E^2(y)$$

and

$$RHS = -\frac{q}{\varepsilon_s} N_A \int_{\infty}^{y}[e^{-\beta\psi(y)} - 1 - e^{-2\beta\phi_F}(e^{\beta\psi(y)} - 1)]2\frac{d\psi}{dy}dy$$

Now since $\psi = 0$ at $y = \infty$, and $\psi = \psi(y)$ at y, introducing the dummy variable, ψ' gives

$$RHS = -2\frac{q}{\varepsilon_s} N_A \int_{0}^{\psi(y)} [e^{-\beta\psi'(y)} - 1 - e^{-2\beta\phi_F}(e^{\beta\psi'(y)} - 1)]d\psi'$$

$$= -2\frac{q}{\varepsilon_s}N_A\left[-\beta^{-1}e^{-\beta\psi'(y)} - \psi'(y) - e^{-2\beta\phi_F}\left(\beta^{-1}e^{\beta\psi'(y)} - \psi'(y)\right)\right]_0^{\psi(y)}$$

$$= -2\frac{q}{\varepsilon_s}N_A\Big[-\beta^{-1}e^{-\beta\psi(y)} - \psi(y) - e^{-2\beta\phi_F}\left(\beta^{-1}e^{\beta\psi(y)} - \psi(y)\right)$$

$$+\beta^{-1} + 0 + e^{-2\beta\phi_F}\left(\beta^{-1} - 0\right)\Big]$$

$$= -2\frac{q}{\varepsilon_s}N_A\left[\beta^{-1}(1 - e^{-\beta\psi(y)}) - \psi(y) - e^{-2\beta\phi_F}\left(\beta^{-1}e^{\beta\psi(y)} - \psi(y) - \beta^{-1}\right)\right]$$

$$= -2\frac{q}{\varepsilon_s}N_A\left[-\beta^{-1}e^{-\beta\psi(y)} - \psi(y) + \beta^{-1} - e^{-2\beta\phi_F}\left(\beta^{-1}e^{\beta\psi(y)} - \psi(y) - \beta^{-1}\right)\right]$$

$$= 2\frac{q}{\varepsilon_s}N_A\left[\beta^{-1}e^{-\beta\psi(y)} + \psi(y) - \beta^{-1} + e^{-2\beta\phi_F}\left(\beta^{-1}e^{\beta\psi(y)} - \psi(y) - \beta^{-1}\right)\right]$$

whence

$$E = -\frac{d\psi}{dy}$$

$$= \mp\left(2\frac{q}{\varepsilon_s}N_A\right)^{1/2}\Big[\beta^{-1}e^{-\beta\psi(y)} + \psi(y) - \beta^{-1}$$

$$+e^{2\beta\phi_F}\left(\beta^{-1}e^{\beta\psi(y)} - \psi(y) - \beta^{-1}\right)\Big]^{1/2} \tag{3.78}$$

The electric field deep in the bulk equals zero. Therefore, assigning a value of Q_C' to the total semiconductor charge per unit area, and E_s to the value of the electric field at the surface permits us to write

$$E_s = \frac{Q_C'}{\varepsilon_s} \tag{3.79}$$

E_s an be obtained from (3.79) by substituting ψ_s for $\psi(y)$ to produce

$$Q_C' = \mp\sqrt{2q\varepsilon_s N_A}\left[\beta^{-1}e^{-\beta\psi_s} + \psi_s - \beta^{-1} + e^{-2\beta\phi_F}\left(\beta^{-1}e^{\beta\psi_s} - \psi_s - \beta^{-1}\right)\right]^{1/2} \tag{3.80}$$

This then lets us view the total semiconductor charge as a function of surface potential.† This is plotted in figure 3.17

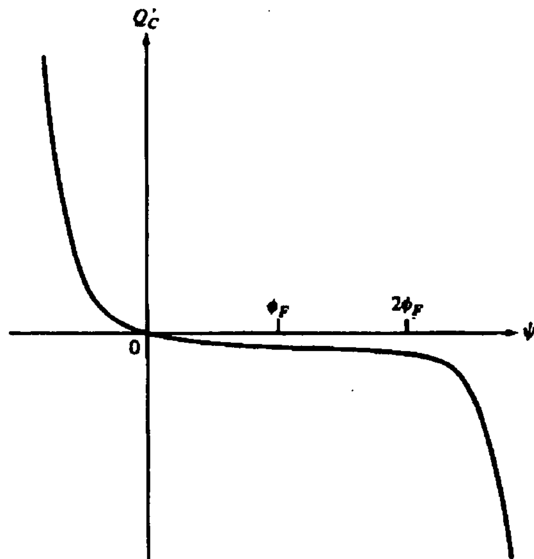


Figure 3.17. Total Charge *vs* Surface Potential

This is valid in all regimes (accumulation, depletion and inversion). The variation of the potential $\psi(y)$ as a function of y can be obtained by numerical integration of (3.69) and then direct substitution of this into (3.62), (3.63) and (3.79) yields the dependence of n, p, and E as a function of y. Useful results can be obtained by approximating for the specific terms in (3.79) according to the substrate condition as follows.

Referring to figure 3.16, we can see that values for surface potential relative to the bulk can be used for the 5 regions shown. The most important condition to consider here is the inversion region. The surface becomes inverted when the intrinsic energy level at the surface crosses the fermi level. i.e. $|\psi_s| > \phi_F$. For commonly used substrate doping levels $(10^{12} - 10^{16} cm^{-3})$, the Fermi potentials are

† It should be noted here that this is only valid in 1D. i.e. this assumes a lateral uniformity. Further discussion of this takes place in Chapter4.

approximately between $9/\beta$ and $16/\beta$. i.e. $18/\beta < 2\phi_F < 32/\beta$. Neglecting the negative exponential terms in (3.79) gives

$$Q_C' = -\sqrt{2qN_A\varepsilon_s}\left[\psi_s - \beta^{-1} + \beta^{-1}e^{\beta(\psi_s - 2\phi_F)}\right]^{1/2} \tag{3.81}$$

The total charge per unit area below the oxide is the sum of the charge in the inversion layer per unit area, Q_I', plus that in the depletion region Q_B'.

$$Q_C' = Q_B' + Q_I' \tag{3.82}$$

Values for Q_I' and Q_B' can be obtained by integrating the expressions for carrier concentrations obtained from equations (3.24) and (3.25). A convenient and widely used approximation however is the **charge sheet model**, whereby the inversion charge is assumed to be contained in an infinitesimally thin region below the gate oxide. The bulk charge then can be found by assuming that $n = 0$ in formulating Q_C', thus eliminating the fourth term within the square bracket in (3.79), and again omitting the negative exponential we obtain

$$Q_B' = -\sqrt{2qN_A\varepsilon_s}(\psi_s - \beta^{-1})^{1/2} \tag{3.83}$$

whereupon

$$Q_I' = -\sqrt{2qN_A\varepsilon_s}\left[\left(\psi_s - \beta^{-1} + \beta^{-1}e^{\beta(\psi_s - 2\phi_F)}\right)^{1/2} - (\psi_s - \phi_F)^{1/2}\right] \tag{3.84}$$

Often the β^{-1} term is omitted with little loss of accuracy. It it easy to see that the presence of the exponential term in (3.84) will dominate as soon as $\psi_s > 2\phi_F$. Therefore *a large change in the inversion charge is needed to cause a small change in the surface potential.* This relationship is shown in figure 3.18. Ideally we wish to relate the gate voltage to the substrate conditions. To do this we must consider the charges in the system and their corresponding potentials. Consider the system shown in figure 3.19. The gate potential, V_{GB}, is dropped across the oxide and the silicon. Therefore.

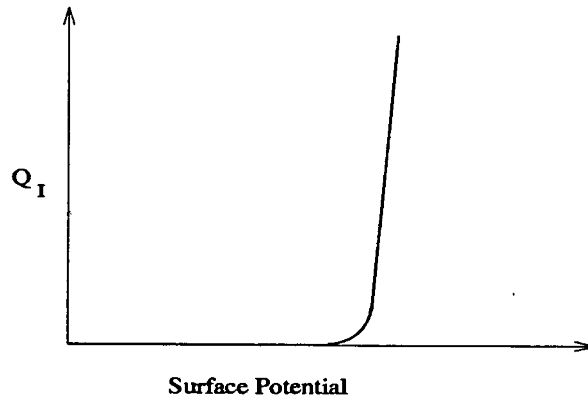


Figure 3.18. Inversion Charge *vs* Surface Potential.

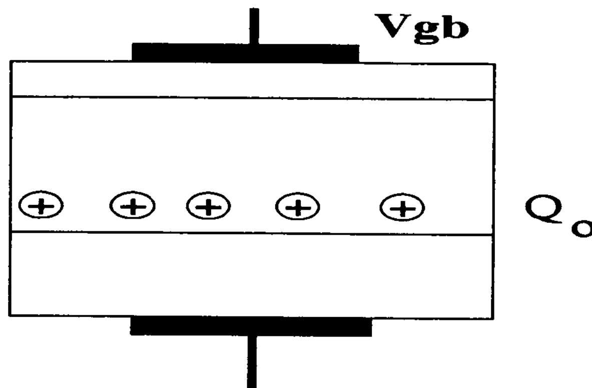


Figure 3.19. Oxide Charges In MOS Capacitor

$$V_{GB} = \phi_{ms} + \psi_{ox} + \psi_s \qquad (3.85)$$

Also by charge conservation

$$Q_G^{'} + Q_o^{'} + Q_B^{'} + Q_I^{'} = 0 \qquad (3.86)$$

and

$$Q_G^{'} = C_{ox}^{'} \psi_{ox} \qquad (3.87)$$

where

$Q_o^{'}$ is the aggregate charge per unit area in the oxide

C'_{ox} is the oxide capacitance per unit area

Recognizing that

$$V_{FB} = \phi_{ms} - \frac{Q'_o}{C'_{ox}} \tag{3.88}$$

then

$$V_{GB} = V_{FB} + \psi_s + \gamma\left(\psi_s + \beta^{-1} e^{\beta(\psi_s - 2\phi_F)}\right)^{1/2} \tag{3.89}$$

Where the term γ, is defined as

$$\gamma = \left[\frac{2q\varepsilon N_A}{C'_{ox}}\right]^{1/2} \tag{3.90}$$

and is known as the **body effect coefficient**. Applying this to (3.90), neglecting the exponential term whenever it becomes negative yields

$$V_{GB} = V_{FB} + \psi_s + \gamma\sqrt{\psi_s} \tag{3.91}$$

By defining a value of ψ_s, a value $V_{t0} = V_{GB}$, can be also defined. Where the value of $\psi_s > 2\phi_F$, the exponential term will dominate. However, figure 3.18 demonstrates that small changes in ψ_s cause large changes in the inversion charge. Stating the converse, large changes in the inversion charge are necessary for a small change in the surface potential. Therefore, the surface potential is relatively insensitive to changes in V_{GB} beyond the a critical value of ψ_s. Equations (3.84) and (3.85) cannot be solved explicitly to express Q'_I as a function of V_{GB}. A numerical substitution [8] however yields a plot as shown in figure 3.20. From this it can be seen that the transition to strong inversion is not a sharp function. V_{t0} is therefore known as the **extrapolated threshold voltage**, though the value ψ_s is undefined other than being greater than $2\phi_F$.

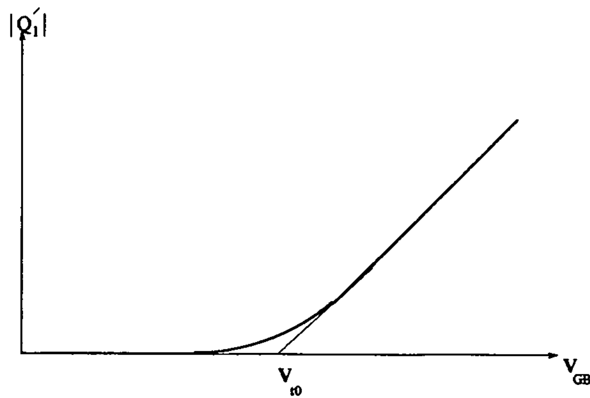


Figure 3.20. Inversion Layer Charge *vs* Gate Voltage.

3.4.5. The Three Terminal Device

Modifying the two-terminal device, by adding an $n+$ region at one end creates the structure shown in figure 3.21.

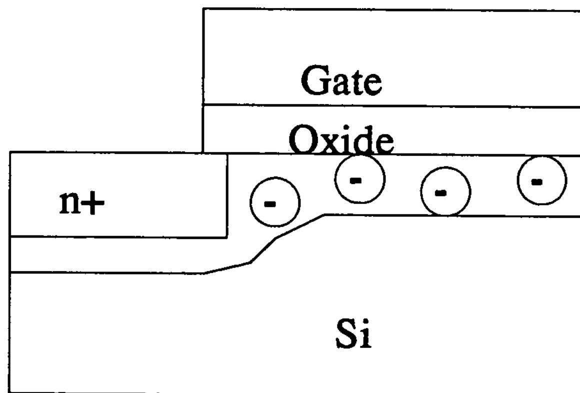


Figure 3.21. The Three Terminal Device.

Here the $n+$ drain region is connected to the inversion region formed by the potential, V_{GB}, applied to the gate relative to the bulk, causing the channel to invert. If the drain region is now biased positively, attracting electrons out of the structure, then the tendency is for the inversion layer to shrink. If the drain bias with respect to bulk, V_{DB}, is big enough the channel will disappear completely. To restore the inversion layer an increased potential, $V_{GB} + V_{DB}$ must be applied. Therefore the electron concentration in the inversion layer is not controlled solely by ψ_s, but by $\psi_s - V_{DB}$.

Therefore (3.55) can be re-written as

$$n(y) = n_o e^{\beta(\psi(y) - V_{DB})} \tag{3.92}$$

The assumption is made that holes do not contribute to the current flow to allow (3.63) to remain as is. Poisson's equation then can be written as

$$\frac{d^2\psi}{dy^2} = -\frac{q}{\varepsilon_s}[p_o(e^{-\beta\psi(y)} - 1) - n_o(e^{\beta\psi(y)} - 1)] \tag{3.93}$$

In a similar fashion to the above section discussing the two terminal structure, the solution yields the following:

$$E(y) = \pm\left(2\frac{q}{\varepsilon_s}N_A\right)^{1/2}\Big[\beta^{-1}e^{-\beta^{-1}\psi(y)} + \psi(y) - \beta^{-1} +$$

$$e^{2\beta\phi_F}\left(\beta^{-1}e^{\beta\psi(y) - V_{DB}} - \psi(y) - \beta^{-1}e^{-\beta V_{DB}}\right)\Big]^{1/2} \tag{3.94}$$

$$Q_C'(y) = \mp(2q\varepsilon_s N_A)^{1/2}$$

$$\left[\beta^{-1}e^{-\beta^{-1}\psi(y)} + \psi(y) - \beta^{-1} + e^{2\beta\phi_F}\left(\beta^{-1}e^{\beta\psi(y) - V_{DB}} - \psi(y) - \beta^{-1}e^{-\beta V_{DB}}\right)\right]^{1/2} \tag{3.95}$$

The bulk charge can be approximated to

$$Q_B' = -\sqrt{2q\varepsilon N_A}\sqrt{\psi_s} \tag{3.96}$$

In the inversion region, given by $Q_C' - Q_B'$,

$$Q_I' = -\sqrt{2q\varepsilon_s N_A}\left[\left(\psi_s + \beta^{-1}e^{\psi_s - \beta(2\phi_F + V_{DB})}\right)^{1/2} - \psi_s^{1/2}\right] \tag{3.97}$$

The gate voltages can also be written as

$$V_{GB} = V_{FB} + \psi_s + \gamma\left[\psi_s + \beta^{-1}e^{\beta(\psi_s - 2(\phi_F + V_{DB}))}\right]^{1/2} \tag{3.98}$$

3.4.6. The 4 Terminal Device.

With the above results we can now derive the equation for the drain current in the MOST. Consider an incremental 'slice' of the region of flow as shown in figure 3.22

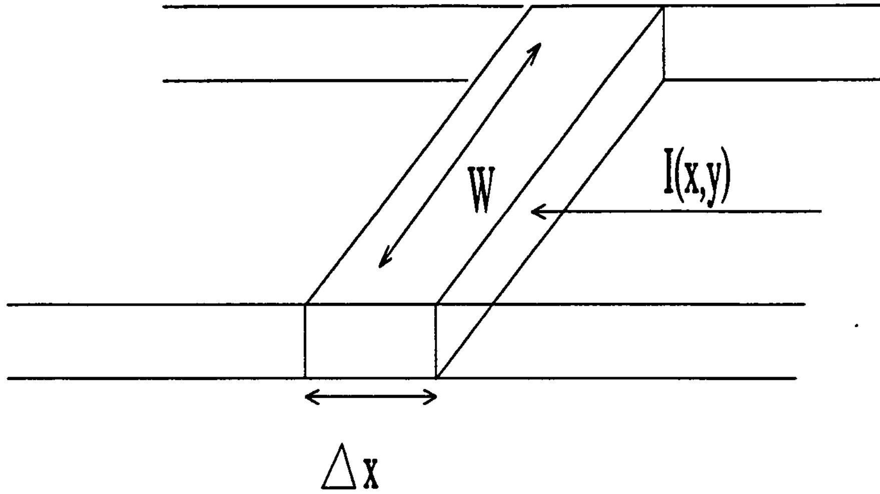


Figure 3.22. Drain Current Derivation.

The component due to one slice, dI_D, is

$$dI_D = dI_{drift}(x,y) + dI_{diff}(x,y) \tag{3.99}$$

To evaluate the drift component, (3.32) can be used where the following substitutions are made.

$b \rightarrow W$

$V \rightarrow \partial\psi_s$

$a \rightarrow \partial x$

$c \rightarrow dy$

The drift component is then

$$dI_{drift}(x,y) = (Wdy)q\mu n(x,y)\frac{\partial\psi(x,y)}{\partial x} \tag{3.100}$$

and the diffusion component from (3.34), where, defining, $\phi_T = k_B T/q$, then $D = \phi_T \mu$, and

$$d_{I_{diff}}(x,y) = -(Wdy)q\mu\phi_t \frac{\partial n(x,y)}{\partial x} \tag{3.101}$$

In a similar fashion to (3.62), at an arbitrary point in the channel n will be given by

$$n(x,y) = n_o e^{\beta(\psi(x,y)-V(x))} \tag{3.102}$$

where $V(0) = V_{SB}$ and $V(L) = V_{DB}$ and $V(x)$ is assumed independent of y. Therefore, on differentiating (3.93)

$$\frac{\partial n(x,y)}{\partial x} = \beta n(x,y)\left[\frac{\partial \psi(x,y)}{\partial x} - \frac{dV(x)}{dx}\right] \tag{3.103}$$

whereby

$$dI_{diff}(x,y) = -(Wdy)q\mu n(x,y)\left[\frac{\partial \psi(x,y)}{\partial x} - \frac{dV(x)}{dx}\right] \tag{3.104}$$

Therefore we can eliminate the $\frac{\partial \psi}{\partial x}$ term to give

$$dI_D = (Wdy)q\mu n(x,y)\frac{dV(x)}{dx} \tag{3.105}$$

To find the total drain current, we must integrate the 'slices' of current, from the surface to the point beyond which n is negligible. i.e.

$$I_D = W\mu \frac{dV(x)}{dx} q \int_{y_s}^{y_c} n(x,y)dy \tag{3.106}$$

This integral then is simply the inversion charge, Q'_I, at a point x in the channel. Therefore

$$I_D = \mu W(-Q'_I)\frac{dV(x)}{dx} \tag{3.107}$$

and then

$$\int_0^L I_D dx = \int_{V_{SB}}^{V_{DB}} \mu W(-Q'_I)\frac{dV(x)}{dx} \tag{3.108}$$

The left hand side = $I_D L$ and so

$$I_D = \frac{W}{L} \int_{V_{SB}}^{V_{DB}} \mu(-Q'_I) dV(x) \tag{3.109}$$

This is a general expression valid over all regions of operation. Different formulae are possible according to the evaluation of Q'_I. To obtain a working model, the condition of strong inversion is examined. As with the two terminal structure, certain relationships are valid; viz.

$$V_{GB} = V_{FB} + \psi_s \tag{3.110}$$

$$Q'_G + Q'_O + Q'_I + Q'_B = 0 \tag{3.111}$$

$$Q'_G = C'_{ox} \psi_{ox} \tag{3.112}$$

$$Q'_B = -\gamma C'_{ox} \sqrt{\psi_s} \tag{3.113}$$

$$Q'_I = -C'_{ox} \left(V_{GB} - V_{FB} - \psi_s - \gamma \sqrt{\psi_s} \right) \tag{3.114}$$

We define a parameter, $V_{CB}(x)$,

$$V_{CB}(x) = \psi_s(x) - \phi_B \tag{3.115}$$

where ϕ_B is the arbitrarily defined potential in inversion. Therefore

$$dV_{CB}(x) = d\psi_s \tag{3.116}$$

and (3.113) can be written as

$$Q'_B = -\gamma C'_{ox} (\phi_B + V_{BC})^{1/2} \tag{3.117}$$

This can be substituted directly into (3.109), recognizing that $V_{CB}(0) = V_S$, and $V_{CB}(L) = V_{DB}$.

$$I_D = \frac{W}{L} \mu C'_{ox} \int_{\psi_s(x=0)}^{\psi_s(x=L)} \left(V_{GB} - V_{FB} - \psi_s - \gamma \sqrt{\psi_s} \right) d\psi_s \tag{3.118}$$

Integrating this and using $\psi_s(x=0) = V_{SB} + \phi_B$ and $\psi_s(x=L) = V_{DB} + \phi_B$ yields

$$I_D = \frac{W}{L}\mu C'_{ox}\Big[(V_{GB} - V_{FB} - \phi_B)(V_{DB} - V_{SB}) - \frac{1}{2}\left(V_{DB}^2 - V_{SB}^2\right)$$

$$-\frac{2}{3}\gamma(\phi_B + V_{DB})^{3/2} - (\phi_B + V_{SB})^{3/2}\Big] \qquad (3.119)$$

This can be further reduced using the identities $V_{GB} = V_{GS} + V_{SB}$ and $V_{DB} = V_{DS} + V_{SB}$ to

$$I_D = \frac{W}{L}\mu C'_{ox}\Big[(V_{GB} - V_{FB} - \phi_B)V_{DS} - \frac{1}{2}V_{DS}^2$$

$$-\frac{2}{3}\gamma(\phi_B + V_{SB} + V_{DS})^{3/2} - (\phi_B + V_{SB})^{3/2}\Big] \qquad (3.120)$$

This expression, although providing reasonable accuracy, is often too complicated for circuit simulation. An approximation can be made by taking a Taylor series expansion of (3.120) around the point $V_{CB} = V_{SB}$ [8] to give

$$I_D = \frac{W}{L}\mu C'_{ox}\left[(V_{GS} - V_t)V_{DS} - \frac{V_{DS}^2}{2}\right] \qquad (3.121)$$

where

$$V_t = V_{FB} + \phi_B + \gamma\sqrt{\phi_B + V_{SB}} \qquad (3.122)$$

3.5. Ion Implanted Channels

Frequently the substrate between source and drain is implanted with dopant to adjust the concentration in the substrate primarily to fix the threshold voltage. This is the preferred option to raising the substrate doping which would have the same effect, though would also increase the source and drain junction capacitances. The general effect of these has been investigated by [9, 10] for a variety of dopant profiles where the integral of the dopant must be evaluated.

3.6. Summary

Chapter 3 has discussed the basic device physics necessary to understand the operation of the MOST. Attention has been paid to the threshold voltage derivation, which is one of the fundamental parameters in device characterization. The assumptions and approximations used in the derivation have been highlighted, and will be seen again in the results presented in chapters 5 and 6. The current equation for the MOST has been derived and approximations made to produce a commonly used working model.

References

1. C. Kittel, *Introduction to Solid State Physics,* Wiley & Sons, New Yprk, 1976.

2. S. M. Sze, *Physics Of Semiconductor Devices,* Wiley & Sons, New York, 1981.

3. S. Selberherr, "Process and Device Modelling For VLSI", *Microelectronics Reliability*, vol. 24, no. 2, pp. 225-257, 1984.

4. E. Conwell and V. F. Weisskopf, "Theory Of Impurity Scattering In Semiconductors", *Physics Review*, no. 77, p. 358, 1950.

5. B. E. Deal and A. S. Grove, "General Relationship for the Thermal Oxidation of Silicon", *Journal of Applied Physics*, vol. 36, no. 12, pp. 3770-3778, 1965.

6. E. H. Nichollian and J. Brews, *MOS (Metal Oxide Semiconductor) Physics And Technology,* Wiley & Sons, 1982.

7. M. M. Atalla, *US Patent 3,206,670*, 1960.

8. Y. P. Tsividis, *Operation And Modelling Of The MOS Transistor,* McGraw-Hill, New York, 1987.

9. J. R. Brews, "Threshold Shifts Due To Nonuniform Doping Profiles In Surface Channel MOSFETs", *IEEE Trans. Elec. Dev.*, vol. ED-26, no. 11, pp. 1696-1710, Nov. 1979.

10. R. R. Troutman, "Ion Implanted Threshold Tailoring For Insulated Gate Field-Effect Transistors", *IEEE Trans. Elec. Dev.*, vol. ED-24, no. 3, pp. 182-192, Mar. 1977.

Chapter 4

Device Width

4.1. Introduction

Chapter 3 dealt with the basic physics of the MOS device and related structures. These were idealised cases, which are seldom, if ever, achieved in VLSI process fabrication. The sharp edges which delineate the features are, in reality, imprecise, ambiguous and open to a variety of interpretations. This arises through process variations, fundamental physical limits and originally in ill-defined terminology which is itself a consequence of approximations in model derivations. The attempt to close the loop between the drawn dimension and the final values of the circuit parameters has forced model refinement and a greater understanding of the physical processes as a consequence of the fabrication techniques. Current circuit simulators require process determined values as input to predict the circuit response. The popular program SPICE, for example can require up to 19 parameters for the level 3 model [1] such as threshold voltage, mobilities, oxide thickness, surface inversion potential, junction depth, overlap capacitances, width and length effects on threshold voltages etc. These parameters are model specific and their values must be derived from measurements on fabricated devices. Interactions between parameters make the extraction technique at best difficult and at worst ambiguous. Consequently it is possible to extract parameters which are not physically meaningful even though they accurately model the device characteristics. The trade-off lies in encoding simple models for ease and speed of operation whilst also making them sufficiently precise to achieve accurate results. The method used to extract parameters then, is critical in

achieving these objects. Equally so is the definition of the parameters in the first place.

4.2. Threshold Voltage.

The primary, and most ambiguous parameter within the SPICE range is the threshold voltage. The previous chapter described the origin of the 'backwards extrapolation' technique for the threshold voltage. This is but one definition which has the advantage that it is easily measured. Akers [2] has reviewed six of the most common definitions. These are

(1) the inversion charge density is equal to zero

(2) the channel current is linearly extrapolated to zero for the linear region of the MOSFET

(3) the channel current is quadratically extrapolated to zero for the saturation of the MOSFET.

(4) the channel current is a small constant value normalised by the length to width ratio

(5) a specified mobile charge level at the virtual cathode

(6) the surface potential ψ_s is equal to $2\phi_F$, where ϕ_F is the bulk Fermi potential.

Widely different values can be obtained according to the definition used. The inherent assumption in all of the above is the topography of a flat device. This is perhaps most easily seen in the sixth of Akers' definitions. The assumption here is that the surface potential is constant throughout the width of the device. The edges of the device however are not fabricated as they are drawn. For example, the transition

from active area to field region is determined by the interaction of the doping profiles, oxide thickness and topographic structure of the device itself. Whilst the influence of the first of these is self-evident, the last has proved to be an area of confusion. Initial publications showed a reduction of threshold voltage with width [3, 4] . Later the opposite effect was seen [5] . This anomaly was resolved when the width dependence of the threshold voltage was investigated, in relation to the isolation technique employed during fabrication [6, 7] .

In some instances, the threshold has been defined as the gate voltage necessary to produce 1 μA regardless of width [8] . However it is obvious that doubling the width of a device will double the 'small' current. Asenov [9] redefined the 'small current' as current per unit micron of device width in two of his V_t definitions. However the definition of the device width is itself ambiguous.

4.2.1. Width Dependence Of Threshold Voltage.

The influence of device width on threshold voltage has been studied [3, 4, 9-11] and the threshold voltage has been observed to increase with decreasing device width. To understand this phenomena, it is necessary to view the device in a 2D cross-section as shown in figure 4.1. In the ideal case the depletion region is limited to the area under the gate oxide. In reality this is not so. The reason for this is that field lines emanating from the edges of the gate electrode terminate on ionized substrate atoms outwith the region directly below the gate. Thus the voltage needed to invert the region is higher than that predicted in chapter 3. The degree to which this fringing field affects the threshold voltage is determined by three factors:

(a) the oxide thickness gradation

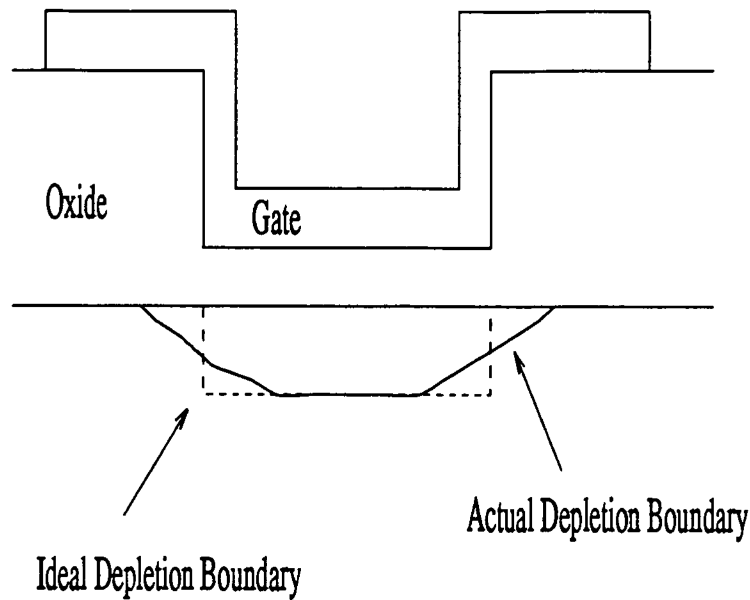


Figure 4.1. Depletion region in cross-section

(b) the doping profile transition

(c) the surface silicon topographic profile

If the device is very wide, then the percentage effect of the fringing field is small and little effect is seen. If the device is narrow, the edge effect plays a proportionately higher part in the operation of the device. A device is defined as narrow [2] if the width of the depletion region is approximately equal to or greater than the width of the device. In this case, the portion of the depletion region outwith the 'flat' gate contains a quantity of charge which must be incorporated into the depletion charge used in the V_t calculation (3.91). Increasing the gate voltage causes a further spread of the depletion charge out from the gate region. The shape containing this 'extra' charge has been approximated by Akers [2] as a triangle, quadrant and square, of which the last gave the best fit to experimental data. All of [3, 4, 9-11] however, assumed uniform substrate doping. Further refinement was seen with analytical

models including tapered oxide and a field doping encroachment assuming a linear concentration gradient toward the channel [12] . By calculating the charge in this extension, and the capacitance of the structure, the threshold voltage shift was obtained by dQ/C.

The analytic approach is limited to approximations of the three factors stated above. The advantage however is in the development of simpler models for SPICE implementation. e.g. the level 3 model makes the assumption that the 'charge shapes' are ellipsoids where the ratio of the minor to major axes is given by the fitting parameter δ [13] .

The assumption of constant surface potential in the charge extension models limited their usefulness to large gate voltages. Numerical modelling removed many of the inherent approximations of the analytic approach. In this case the surface potential is allowed to vary according to lateral position [14] and in [10] was not pinned at $2\phi_F$. The inversion region could then be examined at low gate voltages, including the surface potential variation.

The variation of surface potential when considered by Sugino [15] threw up an the anomalous threshold variation according to the last of the three parameters named above viz; the silicon topography. Sugino predicted a decrease of the threshold voltage with width reduction, according to the slope angle of the recessed oxide. This reduction of threshold voltage with width was called the 'Inverse Narrow Width Effect', and was further investigated by Akers [5, 16] An analytic expression was obtained by considering the effect of the converging electric field lines on the sidewall of the recessed structure. Thus it was seen that a sharp corner on the substrate could act in the same way as a lightning rod, providing a focus for electric field lines. Therefore, the definition of surface potential as a threshold voltage necessitated a surface position at which the potential is quoted. This anomalous behaviour also showed that the device edges conducted current *before* the rest of the channel [17] .

This is the opposite to that found for non-recessed structures. The intermediate case of the semi-recessed device could show either result according to the slope angle of the transition region. This of course meant that the definition of threshold voltage which relies on the surface potential, could no longer be used for narrow devices, where the range of uniform surface potential was vanishingly small.

4.2.2. The Parasitic Field Transistor

As discussed in section 4.2 of Chapter 2, the maximum packing density is obtained by a trench or box isolation process. Cross-talk between parasitics can occur, however along the trench sidewalls, where high interface charge may exist as the crystal orientation changes from the lower surface concentration normally chosen to reduce this effect. This is seen to a lesser extent with semi-recessed processes, but is difficult to quantify due the very small physical dimensions over which the leakage paths exist [18] . The trend towards steeper wall angles to increase the packing density meant that in the extreme, an angle close to $\pi/2$ could delineate the field from active area. As discussed above, this can lead to the INWE because of field lines crowding at the top corner of the device. Around the same time that this phenomenon was discovered, the same reasoning concluded the existence of the opposite effect in the thick field parasitic device.

Between two unrelated diffused junctions, there exists the field oxide, across which may pass a section of interconnect, creating an unwanted MOS transistor. To maintain isolation, the field oxide is made thick enough, (normally at least 10 times the gate oxide) and the substrate further doped to prevent inversion in the channel. The structure of this device is shown in figure 4.2 As can be seen the gate electrode must invert the region between the source and drain. The electric field lines emanating from the gate, *diverge* at the silicon corners at either edge of the channel.

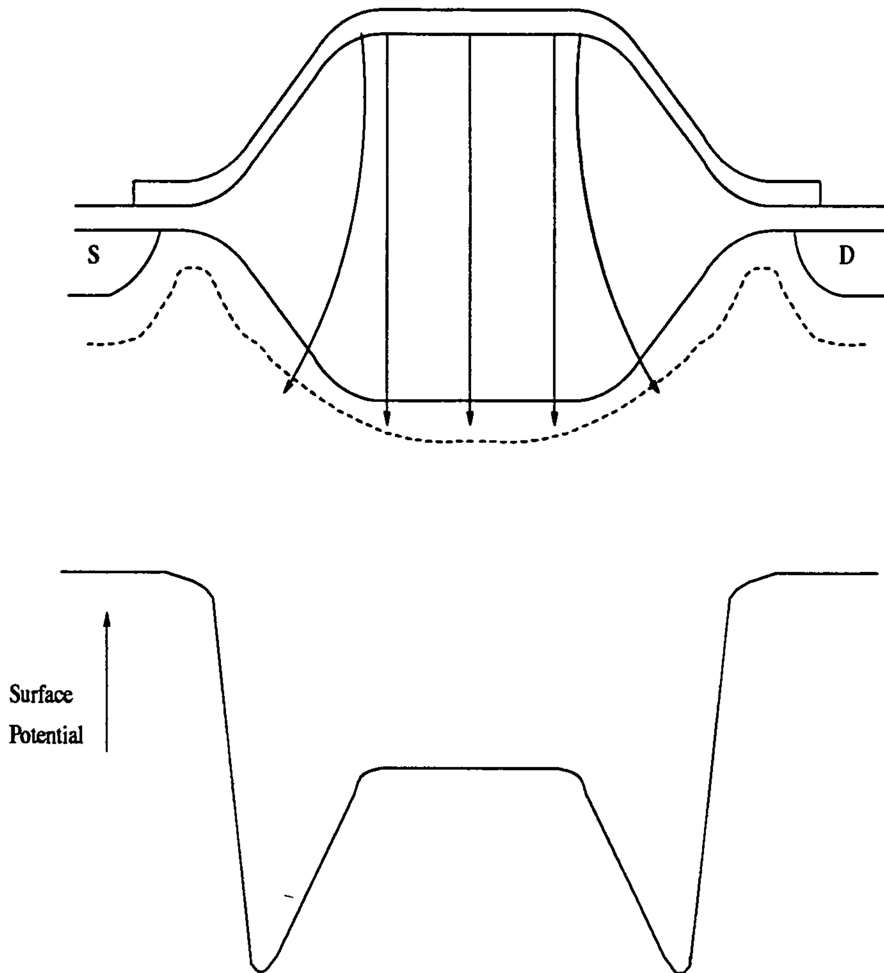


Figure 4.2. Electric Field and Potential Across Device Width.

This effect was studied by Goodwin [19] who showed that the diverging field lines cause a thinner depletion layer under the corners, where the depletion layer is more spread out and contains less charge than in the centre of the channel. Thus the potential at these corners is weaker than the rest of the channel. This effect is only seen when the junctions are either removed from the corners laterally or are shallower than the field oxide below the silicon surface. These potential 'low spots' invert last, after the channel centre and because they are *in series* with the rest of the device, they prevent conduction occurring in the channel till the gate voltage is beyond that needed to invert the channel centre. Thus they raise the thick field threshold voltage.

This contrasts with the opposite effect in the thin gate oxide device where the corners are in parallel with the channel centre, and can turn on first, providing the sub-threshold leakage current. The topographical structure of these two devices is seen then to play a governing role in the current characteristics. The dependency of the field isolation on the recessed silicon has been studied in 2D simulations [20, 21] but both the corner and the recessed oxide have been difficult to quantify experimentally [22, 23] .

To fully incorporate the width variation of the actual device into a predicted current model requires 3D simulation. Thurner, [24] has shown that the shape of the transition region is a major contributor to the drain current. This was done by assuming analytic shape functions of the birds beak profiles in the LOCOS process, characterised by a parameter, *BEAKL*, the distance between the gate oxide and the full field oxide. The gradient of the slope was shown to be mirrored in the potential distribution down the transition region, although the study did not incorporate the field implant.

Finally the doping profile can have as great an impact on the edge characteristics as the oxide thickness and the topographic profile. However, the concentration profile and the topographic profile are invariably linked, particularly in the LOCOS process. Asenov [9] used an analytic model for the surface topography and numerical modelling of the dopant distribution to study enhancement and depletion devices, and extract the voltage dependencies each case. The voltage was defined in terms of the current which was normalised to the drawn width of the device.

4.3. Device Width

Intuitively, the device width is the lateral extent of the conducting channel, and will usually be different from the drawn dimension. However, it may also be different

from the final measured dimension on the wafer. By definition, the width of the conducting channel can only be measured electrically. To do this, the channel characteristics must be interpreted according to equations which describe the $I-V$ relationship.

4.3.1. The Series Resistance.

To interpret the measured data in terms of the $I-V$ equations, the parasitic resistance external to the transistor must first be eliminated. Equation (3.121) in Chapter 3 relates the drain current to drain voltage.

$$I_D = \frac{W}{L}\mu C'_{ox}\left[(V_{GS}-V_t)V_{DS}-\frac{V_{DS}{}^2}{2}\right] \tag{4.1}$$

This can be approximated when the drain voltage is low such that $\frac{V_{DS}}{2} \ll V_{GS}-V_t$ to

$$I_D = \frac{W}{L}\mu C'_{ox}(V_{GS}-V_t)V_{DS} \tag{4.2}$$

The drain voltage has been assumed to act over the channel region only. However, as shown in figure 4.3, there are other resistances present. In series with the channel resistance there are resistances arising from

(i) External wiring

(ii) Silicon contacts

(iii) Source/drain sheet resistances

(iv) Current crowding from/to the inversion layer

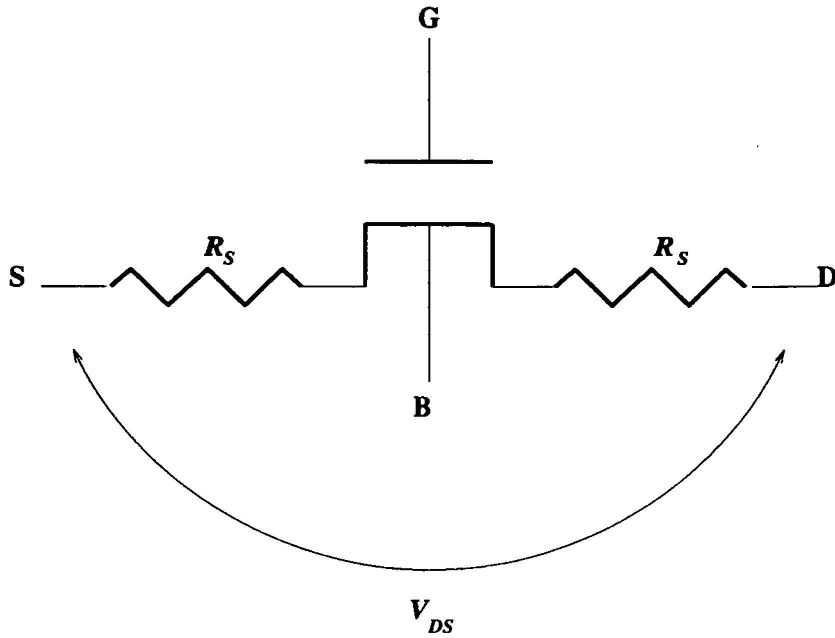


Figure 4.3. Source and Drain Series Resistances.

If the collective resistance of these is denoted as R_s, on either side, then,

$$V_{DS} = \hat{V}_{DS} + 2R_s I_D \tag{4.3}$$

where $\hat{V}_{DS}$ is the voltage dropped across the channel only. The drain equation then becomes

$$I_D = \frac{W}{L}\mu C'_{ox}(V_{GS} - V_t)(V_{DS} - 2I_D R_s) \tag{4.4}$$

The measured resistance is given by

$$R_m = \frac{V_{DS}}{I_D} \tag{4.5}$$

and from (4.4)

$$R_m = 2R_s + \frac{L}{W\mu C'_{ox}V_{gt}} \tag{4.6}$$

Re-arranging this and substituting R_x for $2R_s$,

$$R_x = R_m - \frac{L}{W\mu C'_{ox}V_{gt}} \tag{4.7}$$

This equation provides a means of extracting the series resistance, by plotting V_{gt}^{-1} against R_m. Other methods are described in [25-27] . Then the $I-V$ characteristics of the device itself can be obtained.

4.3.2. Mobility Modulation

A further modification is also made to (4.1) whereby the mobility term in the equation is replaced by an **effective surface mobility**. The surface mobility is different from the bulk mobility due to interaction with the interface where the surface roughness can cause scattering, and the interface traps can hold and release carriers [8] . Further, the presence of a perpendicular electric field, in the form of the applied gate voltage can modulate the surface mobility. The surface mobility, μ_s is given by [28]

$$\mu_s = \frac{\mu_o}{1+\theta E_{eff}} \tag{4.8}$$

where

μ_o is the low field surface mobility

θ is the experimentally determined mobility degradation coefficient.

E_{eff} is proportional to the perpendicular electric field and can be approximated as V_{gt}.

Incorporating this into (4.7) yields

$$R_x = R_m - \frac{L(1+\theta V_{gt})}{W \mu_o C'_{ox} V_{gt}} \tag{4.9}$$

and therefore

$$R_x = R_m - \frac{A}{WV_{gt}} - \frac{A\theta}{W} \tag{4.10}$$

where

$$A = \frac{L}{\mu_o C'_{ox}}$$

The effect of including the mobility modulation term has a large impact on the parameter extraction. It should be noted that relating R_m to V_{gt}^{-1} does not explicitly yield the series resistance, R_x and thus invalidates the technique described from the previous section.

4.3.3. Width Definitions

As with threshold voltage there are several definitions of device width. Ideally only two would be expected: the drawn width and the final width. In reality there are 4 possible modes of operation within which to define different width parameters. These are defined as follows.

The Drawn Width.

This is simply the reference as used by the designer.

The Topographic 'Flat' Width.

This was a measure of device width considered by Asenov [9] . Here the width was defined as the extent of the gate oxide region, before the onset of the birds beak encroachment. This was rejected as a possibility because of the difficulty in measurement and the destructive nature of the test. This method, however, is employed in length measurement by Karnett [29] whereby the sample was cleaved using scribe marks to locate the area of interest and snapping the wafer between a plexiglass sandwich. Agreement to within 0.1 μm with different electrical tests was obtained, and highlighted the necessity of calibration against the approximations in the device equations used in the extraction routines.

The Electrically Extracted Width

This technique uses device models which, through judicious measurements, allow the effects of the discrete parameters to be extracted, and through this, values for these parameters themselves. Ma and Wang [11] proposed a technique based on the assumption that the drain current was proportional to the device width. By plotting one against the other, the intercept on the width axis, corresponding to zero

drain current, gave the width loss, ΔW. This relationship is seen in (3.121) in chapter 3, and established a technique to extract the relationship between the drawn dimension and the effective electrical width. Several devices of constant length and varying widths meant that all parameters but the width in (3.121) could be held constant thus permitting a unique I_{ds} *vs* W relationship. Equation (3.121) is valid only for the transistor operating in the linear regime. The same technique was used by Chung [13] who developed an *I-V* model based on semi-empirical parameters extracted from experimental data.

The Modelled Width

Voss & Engl [30] extended the work of Ji & Sah [14] to consider the dependence of the depletion region under the gate on the applied gate voltage for two different isolation techniques. Here the width was defined as the point along the channel where the electron concentration was equal to the doping (the start of inversion). The use of a 2D process simulator provided another meaning for the width of a transistor [31] where, an effective width, W_{eff}, was defined as

$$W_{eff} = \int \frac{n(x)dx}{n_c} \tag{4.11}$$

where

n is the carrier concentration in the channel area

n_c is the uniform) carrier concentration in the centre of the channel.

and the integral is performed from one end to the other end of the gate along the channel width. The outcome was an applied voltage-above-threshold, V_{gt}, dependence of effective width of the form shown in below in fig 4.4 It can be seen

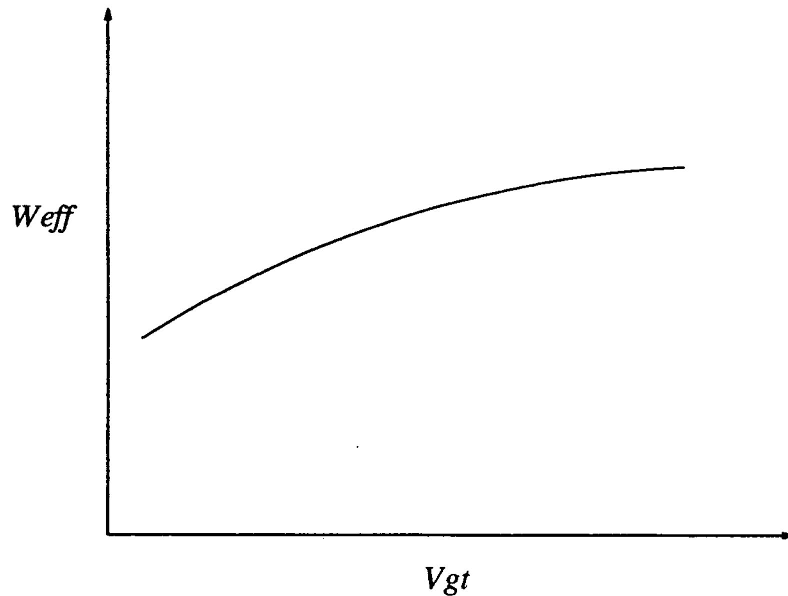


Figure 4.4 The $W_{eff} - V_{gt}$ relationship

that the effective width is predicted as being most sensitive to V_{GS} variations at values just above threshold, where V_{gt} is just above zero. This was explained as the extension of the depletion region down into the birds beak, but limited by both the field implant and the increased oxide thickness of the birds beak. The same dependence has not been observed in the case of the trench isolated device where no lateral extent of the depletion region could occur.

4.3.4. Electrical Extraction Techniques

The V_{gt} dependence of W_{eff} was confirmed by Chia and Hu [32] who measured the effect using an electrical extraction technique. The same form of dependence was seen despite no allowance being made for the external series resistance which was assumed independent of W_{eff}. This is contrary to the work done by Wan *etal* [33] and Arora [34] who showed the necessity of extracting the series resistance to produce a single value of W_{eff}, and by definition showed no variation with gate voltage. The latter used the SEM calibration technique similar to that of Karnett [29] as detailed

above though no junction staining technique was necessary. In this case the width obtained by the SEM value is simply that of the thin gate oxide region. The authors plot the measured resistance, R_m against V_{gt}^{-1}, nominally in the range 1- 4 volts and extract the slope. From equation (4.9),

$$R_{chan} = \frac{L_{eff}(1 + \theta V_{gt})}{\mu_o C'_{ox} W_{eff} V_{gt}} \tag{4.12}$$

where the length and width have been replaced by their effective values and the mobility modulation has been accommodated. The total resistance is given by

$$R_m = R_{chan} + R_x \tag{4.13}$$

where R_x is the external series resistance. The slope value is then given by m where

$$m = \frac{L_{eff}}{\mu_o C'_{ox} W_{eff}} \tag{4.14}$$

The slope is then multiplied by the width and the result plotted against the slope according to

$$mW_D = \Delta W m + \frac{L_{eff}}{\mu_o C'_{ox}} \tag{4.15}$$

where

$$W_{eff} = W_D - \Delta W$$

Thus the width loss can be obtained from the slope of the new graph.

The work of Deen & Zuo [35] differed from previous researchers in that zero effective width was *not* defined at the point of zero current. The authors ignored the series resistance values other than that associated with the transition to and from the channel inversion region. This technique allowed a value of current to be attributed to the transition region in the width direction. Deen & Zuo present no physical model† for this but it is the only definition which distinguishes an "edge width" as opposed to the others which incorporate all parameters into an effective width. Again using a form of equation (4.9), the series resistance is eliminated and a plot of channel conductance against drawn width for different values of V_{gt}. The value of the series resistance then, impacts the further parameter extraction process. At this point it is appropriate to consider, in some detail, the work of three sets of authors in the field.

Their approaches are correlated, from their initial I_D equation to the final equations used in the extraction process. Firstly Suciu & Johnson [36] presented their experimental technique to extract the series resistance. Their initial equation (described in terminology consistent with this thesis) is

$$I_D = \frac{(\mu_o C'_{ox} W_{eff}/L_{eff}) V_{gt} V_{DS}}{1 + (\theta + \mu_o C'_{ox} W_{eff}/L_{eff} R_x) V_{gt}} \tag{4.16}$$

Therefore

$$\frac{V_{DS}}{I_D} = \frac{L_{eff}}{\mu_o C'_{ox} W_{eff} V_{gt}} + \frac{\theta L_{eff}}{\mu_o C'_{ox} W_{eff}} + R_x \tag{4.17}$$

Thus using the approach of Suciu & Johnston the following can be derived

† To be presented [35] .

$$R_m = \frac{A}{W_{eff}V_{gt}} + \frac{A\theta}{W_{eff}} + R_{chan} \tag{4.18}$$

Arora *et al* [34] use as their initial I_D equation

$$I_D = \frac{W_{eff}\mu_o C'_{ox}}{L_{eff}(1+\theta V_{gt})} V_{gt}\hat{V}_{DS} \tag{4.19}$$

where again

$$\hat{V}_{DS} = V_{DS} - R_x$$

Substituting then for $\hat{V}_{DS}$,

$$I_D = \frac{W_{eff}V_{gt}V_{DS}}{A(1+\theta V_{gt})} - \frac{W_{eff}V_{gt}}{A(1+\theta V_{gt})} R_x I_D \tag{4.20}$$

and so

$$I_D\left[1 + \frac{W_{eff}V_{gt}R_x}{A(1+\theta V_{gt})}\right] = \frac{W_{eff}V_{gt}V_{DS}}{A(1+\theta V_{gt})} \tag{4.21}$$

Rearranging produces

$$\frac{V_{DS}}{I_D} = \frac{A(1+\theta V_{gt})}{W_{eff}V_{gt}}\left[1 + \frac{W_{eff}R_xV_{gt}}{A(1+\theta V_{gt})}\right] \tag{4.22}$$

$$= \frac{A(1+\theta V_{gt})}{W_{eff} V_{gt}} + R_x \tag{4.23}$$

And so from the method of Arora *et al*, the final equation for R_m is

$$R_m = \frac{A}{W_{eff} V_{gt}} + \left[\frac{A\theta}{W_{eff}} + R_x \right] \tag{4.24}$$

This equation is identical to that used by Suciu & Johnson.

Deen & Zuo, however, pair their initial I_D equations for two different device widths to extract the series resistance.

For the sake of clarity the subscript has been omitted from the effective width terms. The subscript D has been used to denote the drawn dimension but otherwise in equations (4.25) to (4.36) the quoted widths are the effective widths.

The initial I_D equation is

$$I_1 = KV_{gt} W_1 \left[V_{DS} - I_1 \frac{R_{P0}}{W_1} \right] + 2G_P \left[V_{DS} - I_1 \frac{R_{P0}}{W_1} \right] \tag{4.25}$$

where

$$K = \frac{1}{A(1+\theta V_{gt})}$$

$$R_{P0} = R_x \times W$$

Using a different device width a similar equation can be written

$$I_2 = KV_{gt}W_2\left[V_{DS} - I_2\frac{R_{P0}}{W_2}\right] + 2G_P\left[V_{DS} - I_2\frac{R_{P0}}{W_2}\right] \tag{4.26}$$

Subtracting (4.26) - (4.25) gives

$$I_2 - I_1 = KV_{gt}V_{DS}(W_2 - W_1) - KV_{gt}R_{P0}(I_2 - I_1) + 2G_PR_{P0}\left[\frac{I_2}{W_2} - \frac{I_1}{W_1}\right] \tag{4.27}$$

producing

$$KV_{gt}V_{DS}(W_2 - W_1) = (I_2 - I_1) + KV_{gt}R_{P0}(I_2 - I_1) - 2G_PR_{P0}\left[\frac{I_2}{W_2} - \frac{I_1}{W_1}\right] \tag{4.28}$$

and so

$$\frac{V_{DS}}{(I_2 - I_1)} = \frac{1}{KV_{gt}(W_2 - W_1)} + \frac{R_{P0}}{(W_2 - W_1)} - \frac{2G_PR_{P0}}{KV_{gt}(W_2 - W_1)(I_2 - I_1)}\left[\frac{I_2}{W_2} - \frac{I_1}{W_1}\right] \tag{4.29}$$

$$\frac{V_{DS}}{(I_2-I_1)}=\frac{1}{KV_{gt}(W_2-W_1)}+\frac{R_{P0}}{(W_2-W_1)}-$$

$$\frac{2G_PR_{P0}}{KV_{gt}(W_2-W_1)}\left[\frac{1}{I_2-I_1}\right]\left[\frac{I_2}{W_2}-\frac{I_1}{W_1}\right] \tag{4.30}$$

Defining $\left[\frac{1}{I_2-I_1}\right]\left[\frac{I_2}{W_2}-\frac{I_1}{W_1}\right]=\Xi$ and substituting for K

$$\frac{V_{DS}}{I_2-I_1}=\frac{A(1+\theta V_{gt})}{V_{gt}(W_2-W_1)}+\frac{R_{P0}}{(W_2-W_1)}-\frac{A(1+\theta V_{gt})2G_PR_{P0}}{V_{gt}(W_2-W_1)}\Xi \tag{4.31}$$

$$\frac{V_{DS}}{I_2-I_1}=\frac{A}{(W_2-W_1)}\frac{1}{V_{gt}}+\left[\frac{A\theta+R_{P0}}{W_2-W_1)}\right]-\frac{A(1+\theta V_{gt})2G_PR_{P0}}{V_{gt}(W_2-W_1)}\Xi \tag{4.32}$$

The term Ξ can be justifiably neglected since the ratio of the current to the effective width is approximately constant and thus the second term in Ξ approximates to zero. If $\Xi=0$ then, equation (4.32) reduces to

$$\frac{V_{DS}}{I_2-I_1}=\frac{A}{(W_2-W_1)}\frac{1}{V_{gt}}+\left[\frac{A\theta+R_{P0}}{(W_2-W_1)}\right] \tag{4.33}$$

Again this is the same form as equation (4.24), that used by Arora *et al* and Suciu & Johnson. This equation can be derived from that quoted in [35] where

$$\frac{V_{DS}}{I_{D2}-I_{D1}}=\frac{1+KV_{gt}R_{P0}}{KV_{gt}(W_2-W_1)} \tag{4.34}$$

where substituting for K gives

$$\frac{V_{DS}}{I_{D2}-I_{D1}}=\left[\frac{A}{W_2-W_1}\right]\frac{1}{V_{gt}}+\left[\frac{R_{P0}+A\theta}{W_2-W_1}\right] \tag{4.35}$$

The approach taken by Deen & Zuo is to graph $V_{DS}(I_{D2}-I_{D1})^{-1}$ vs $V_{gt}{}^{-1}$. Here the intercept is quoted as $R_{P0}(W_2-W_1)^{-1}$. It is seen from equation (4.35), that the intercept in fact includes the term $A\theta(W_2-W_1)^{-1}$. If θ were sufficiently small then this term may be neglected. (In fact Deen & Zuo quote a non-negligible value of value of $0.1V^{-1}$). A further complication arises in computing the value of $\hat{V}_{DS}$ given by a form of equation (4.3)

$$\hat{V}_{DS}=V_{DS}-I_D(R_{P0}W_{eff}{}^{-1}) \tag{4.36}$$

However, this requires W_{eff} to be known, unless this is approximated as the drawn dimension, limiting the accuracy at small widths. Once the series resistance has been extracted, graphs of channel conductance are plotted against drawn widths for different values of V_{gt}. 5 values are chosen around an integer value of V_{gt} to produce 5 lines. The intersection of these 5 lines is defined as the point of zero effective width, This intersection does not occur at $G_t=0$ and thus a 'residual' or edge conductance is defined. The results of Deen & Zuo produce a relationship between ΔW and V_{gt} such that ΔW reduces as V_{gt} increases although the sensitivity of ΔW to V_{gt} increases as V_{gt} increases, contrary to [30] . The graph produced is of the form shown in figure 4.5

From this the value of ΔW is extrapolated at $V_{gt}=0$ and is used as a definition of 'intrinsic' channel width and 'intrinsic' ΔW, denoted as ΔW_0. Using this definition equation (4.21) can be written as

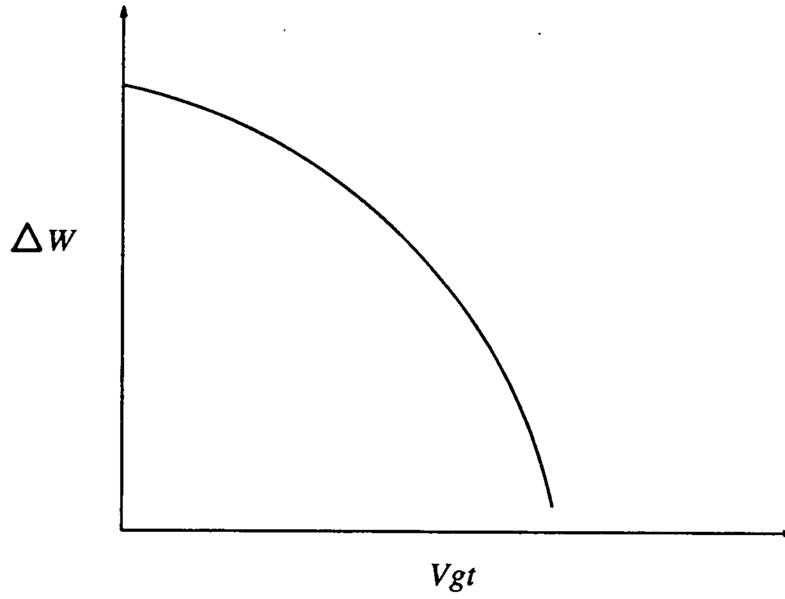


Figure 4.5. $\Delta W - V_{gt}$ relationship[35]

$$I_D = \frac{W_{eff}\,\mu_o C'_{ox}}{L_{eff}(1+\theta V_{gt})} V_{gt}\hat{V}_{DS} + 2I_{edge} \tag{4.37}$$

According to these definitions, edge currents are extracted, where the edge of the device is defined by the effective width extracted at $V_{gt} = 0$. Three sets of devices of different drawn widths such that the sum of the two equals the third, allows the edge current to be extracted according to

$$2I_{edge} = I_{D1} + I_{D2} - I_{D3} \tag{4.38}$$

For this equation to be valid, the device widths must be defined such that

$$W_{D1} + W_{D2} = W_{D3} + \Delta W \tag{4.39}$$

This requires the ΔW value to be known before the extraction process is used.

Table 4.1 is a summary of the current techniques used and measurements obtained in the literature.

Author	Extraction	Modelled	SEM	$W = W(Vgt)$	Edge Currents
Wan et al [33]	X				
Chia & Hu [32]	X	X		X	
Ma & Wang [11]	X				
Voss & Engl [30]	X	X		X	
Arora et al [34]	X		X		
Deen & Zuo [35]	X			X	X

Table 4.1. Current Techniques and Measurements

Deen & Zuo quoted an edge current, based on an extrapolated width definition. This incorporates the topographical and doping conditions present. To identify these definitions with the fabricated device it it necessary to examine the process sequence.

4.3.5. Width Formulation

The topographical influences of the process occur primarily in the choice of isolation technique as discussed in Chapter 2. The dopant profiles are impacted by every high temperature step and may be further influenced by those processes which change the topography. Using a standard LOCOS isolation, within a silicon gate process, certain critical steps can be identified.

Step A: Active Area Photo

A difference between drawn and final dimension is expected to occur in the photolithography and etch used to delineate the feature. The definition of the nitride layer is usually very good, since the film thickness is much smaller than the drawn linewidth. However, this photo defines the active area, which is also the first definition of the gate width, and thus is the drawn dimension on the mask. Typically, the combined thickness of the pad oxide and nitride layers is around 0.3 μm. This is usually dry etched, thus retaining the drawn dimension to better than 0.1 μm. The option here is to leave the pad oxide or remove it. This has little effect on the topographic width but must be considered when implanting the channel stop implant.

Step B: Field Implant

Used to prevent unwanted parasitic crosstalk, as discussed in §4.2.2, this also restricts the lateral extent of the conducting channel which is one of the definitions of device width. A higher dose implant encroaches further into the active area, and a higher energy implant removes the concentration from the surface [9] . This latter condition is preferred in the case of a boron implant, since the impurity segregation depletes the silicon of the implant, negating the effect of using a higher dose. Figure 4.6 shows the impurity profiles as obtained by SUPREM3 for two different implant energies after a field oxidation resulting in a 1μm thick oxide.

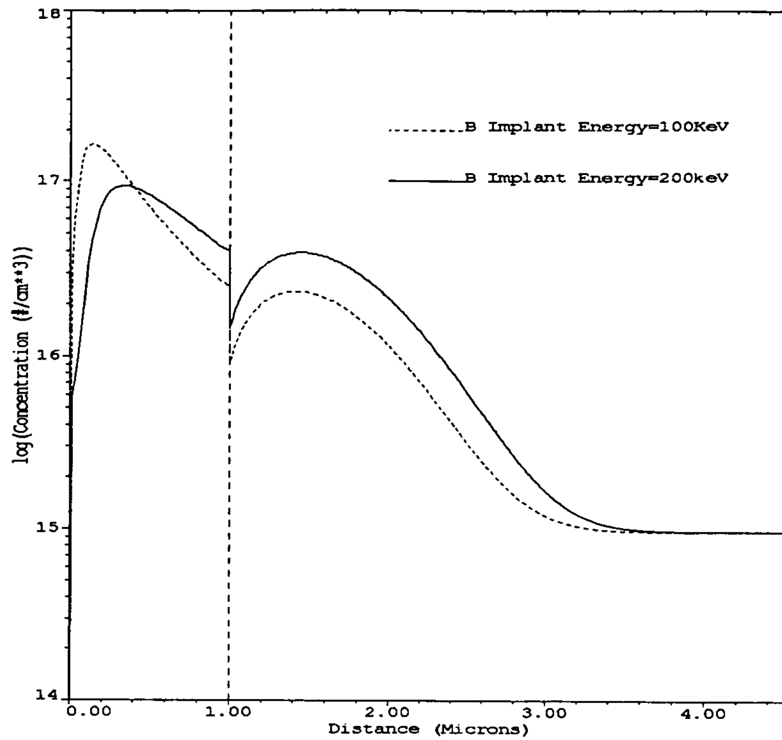


Figure 4.6. SUPREM3 Simulation of field implants.

Step C: LOCOS Process.

The isolation technique is recognised as the major influence on channel width variation. During this process several parameters can alter the device width, or affect the final electrical measurement.

(i) The LOCOS profile is impacted by the process parameters as discussed in chapter 2.

(ii) The substrate dopant, particularly the field implant, is determined by the oxidation cycle as stated in Step B. The amount of dopant present under the birds beak towards the active area is determined by the degree of lateral diffusion and the amount of oxidation i.e. the length of the birds beak.

(ii) Non-uniform birds beak lengths have been reported [37-39] according to the width of the nitride mask. As the width reduces, the birds beak increases because of the increased resistance to bending of the nitride 'beam'. This impacts the electrical extraction technique which assumes constant offset for all drawn widths.

(iv) Enhanced diffusion is known to occur due to the high stress at the mask edge causing vacancy injection during oxidation. This diffuses the implant both into the substrate and towards the channel.

(v) Section 4.2.2 discussed the effect of the transition angle. Thus the shape of the profile may impact the electrical measurements. Guillemot [40] studied the profiles over a range of processing parameters and classified these into two shapes. These shapes were determined according to the nitride and pad oxide thickness combinations used. Figures 4.7 and 4.8 shows the two shapes seen experimentally. The shapes are distinguished by the presence of a step in the silicon substrate, only seen when the ratio of nitride to pad oxide thicknesses is large.

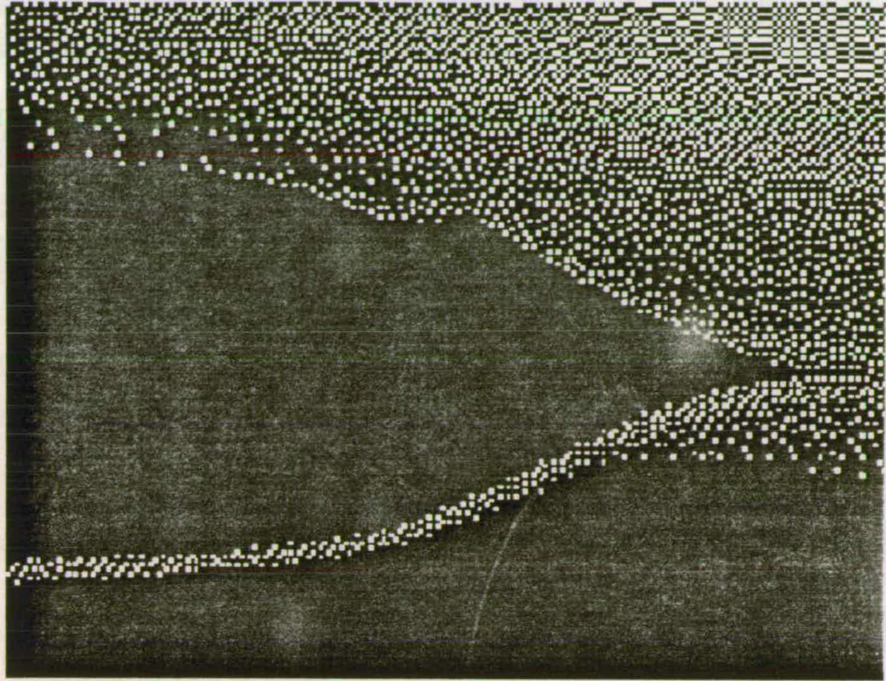


Figure 4.7. Bird's Beak - *Tn/Tpad* small

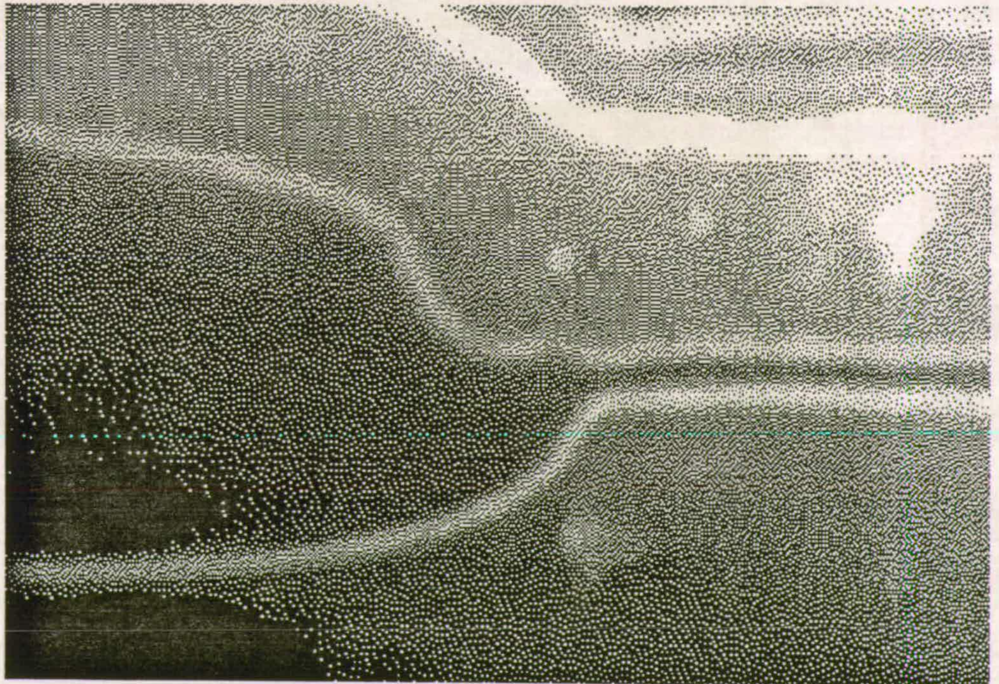


Figure 4.8. Bird's Beak - *Tn/Tpad* large

Step D: Etchback

After field oxide, the nitride is removed and the pad oxide is also etched back prior to gate oxidation. This etchback is done using an isotropic etch. Therefore, overetching can cause an etchback profile as seen in figure 4.9 where the field oxide has been removed down the birds beak. This effect will always occur [7] since the birds beak thickness is not uniform across the wafer and is smaller for small widths. To ensure an etchback to bare silicon for the larger widths, the small devices are overetched. This exposes the surface of the silicon, which had previously been the tip of the birds beak. In reality, since the etchback is done to ensure complete removal of the pad oxide, usually all devices suffer from this, with the small geometries suffering more.

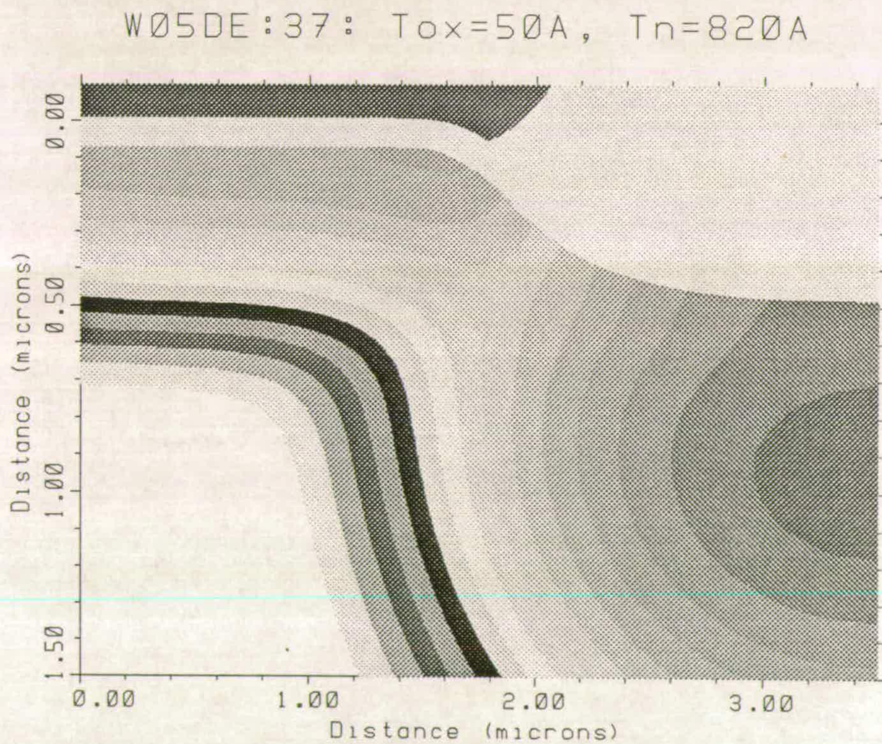


Figure 4.9. Bird's Beak showing overetch condition

Step E: Reoxidation.

This is usually a two stage process, as discussed in Chapter 2. The final oxidation is the gate oxide. This again reduces the dopant at the surface, due to suckout, and rounds off any exposed corners [41] . Oxidation of a sharp corner progresses at a slower pace [42-44] than a planar surface, so an oxide thinning is be seen. However, the smoothing of this sharp feature mitigates the effect of the enhanced electric field discussed in §4.2.2. If the transition is sufficiently sharp, the Inverse Narrow Width Effect is possible according to the degree of overetch and reoxidation as long as the gate oxide extends beyond the topographically flat device [7] .

Step F: Threshold Adjust Implant

The threshold adjust implant which fixes the doping concentration in the centre of the device, is stopped by the field oxide at the channel edge. The gradual transition of gate oxide to field oxide means that the implant reaches the silicon in a gradually reducing amount. The interaction of this and the diffused field implant determines the variation of the threshold voltage according to the width [9] . A further, as yet unreported, effect can be seen in the doping profile in the silicon substrate in figure 4.9. In the case of the overetched device, the implant has penetrated the oxide down the sidewall. Subsequent diffusion leaves a highspot of concentration, since the diffusion of the dopant at the corner is restricted by equal concentrations at the surface and down the sidewall. The result is to leave a concentration peak at the point where the electric field lines will concentrate as shown previously in figure 4.2 This high concentration spot opposes the creation of the charge peaks as discussed in [45] .

Step G: Thermal Budget

No further changes in topography occur after this stage, since the gate width is sealed by the polysilicon bar crossing the active area. Heat treatments further diffuse

the implants, drawing the two together.

In summary the device width, as defined by the lateral extent of the conducting channel, is determined by the interaction of the threshold adjust and field implants, and by the lateral extent of the gate oxide. The edge effects may be expected to be dominated by the concentrations at the edge of the topographic profile as defined by the gate oxide. Some information can be obtained on the topographic width by consideration of the standard processing, though device simulation is necessary to study the influence on electrical width. The impact of the processing steps A to G on the topographic and electrical widths is shown in table 4.2.

	Electrical	Topographical
Step A	↓	↓
Step B	↓	→
Step C	↓	↓
Step D	↑	↑
Step E	↓↑	↓↑
Step F	↓↑	→
Step G	↓↑	→

Table 4.2

↑ width increase

↓ width decrease

↑↓ indeterminate

→ no difference

4.4. Summary

This chapter has reviewed the meaning of device width from several viewpoints and discussed the measurement techniques. Parasitic effects on thick field devices has shown a trade-off situation whereby optimizing packing density can produce unwanted subthreshold currents. The transition region therefore has shown itself as being crucial in determining characteristics of both devices. Edge effects in active devices are influenced by both the topographic structure and the substrate dopant profiles. To quantify these effects, a range of profiles under different dopant conditions is necessary. To this end, an experiment was run and devices modelled, electrically tested and measured topographically. This is described in Chapter 5.

References

1. P. Antognetti and G. Massobrio, *Semiconductor Device Modelling with SPICE*.

2. L. A. Akers and J. J. Sanchez, "Threshold Voltage Models Of Short, Narrow And Small Geometry MOSFETs: A Review", *Solid-State Electronics*, vol. 25, no. 7, pp. 621-641, 1982.

3. K. O. Jeppson, "Influence Of The Channel Width On The Threshold Voltage Modulation In MOSFETs", *Electronics Letters*, vol. 11, no. 14, 10 July 1975.

4. K. E. Kroell and G. K. Ackermann, "Threshold Voltage Of Narrow Channel Field Effect Transistors", *Solid State Electronics*, vol. 19, pp. 77-81, 1976.

5. L. A. Akers, "The Inverse-Narrow-Width Effect", *IEEE Electron Device Letters*, vol. EDL-7, no. 7, pp. 419-421, July 1986.

6. L. A. Akers, M. Sugino, and J. M. Ford, "Characterisation Of The Inverse-Narrow-Width Effect", *IEEE Trans. Elec. Dev.*, vol. ED-34, no. 12, pp. 2476-2484, Dec. 1987.

7. F. M. Klaassen, P. A. van der Plas, R. J. W. Debets, N. A. H. Wils, and M. G. Pitt, "Narrow Width Effects In Submicron MOS ICs", in *ESSDERC*, 1989.

8. Ong, in *Modern MOS Technology*, McGraw-Hill, New York, 1984.

9. A. M. Asenov, E. N. Stefanov, and B. Z. Antov, "Numerical Study Of The Effect Of The Doping Profile On The Threshold Voltage Of Narrow-Channel MOS Transistor", *Solid-State Electronics*, vol. 30, no. 12, pp. 1305-1315, 1987.

10. H-K. An and A. H. Zemanian, "Computationally Efficient Determination Of Threshold Voltages In Narrow-Channel MOSFETS Including Fringing And Inversion Effects", *IEEE Trans. Elec. Dev.*, vol. 36, no. 3, pp. 534-541, Mar 1989.

11. Y-R. Ma and K. L. Wang, "A New Method To Electrically Determine Effective MOSFET Channel Width", *IEEE Trans. Elec. Dev.*, vol. ED-29, no. 12, pp. 1825-1827, Dec 1982.

12. L. A. Akers, M. M. E. Beguwala, and F. Z. Custode, "A Model Of A Narrow-Width MOSFET Including Tapered Oxide And Doping Encroachment", *IEEE Trans. Elec. Dev.*, vol. ED-28, no. 12, pp. 1490-1495, Dec 1981.

13. S. S. Chung, "A Complete Model Of The I-V Characteristics For Narrow-Gate MOSFETs", *IEEE Trans. Elec. Dev.*, vol. 37, no. 4, pp. 1020-1030, Apr 1990.

14. C-R. Ji and C-T. Sah, "Two-Dimensional Numerical Analysis Of The Narrow Gate Effect In MOSFET", *IEEE Trans. Elec. Dev.*, vol. ED-30, no. 6, pp. 635-647, June 1983.

15. M. Sugino and L. A. Akers, "Subthreshold Current In Oxide Isolated Structures", *IEEE Electron Device Letters*, vol. EDL-4, no. 4, Apr. 1983.

16. *L. A. Akers.* Private Communication

17. N. Shigyo, S. Fukuda, T. Wada, K. Hieda, T. Hamamoto, H. Watanabe, K. Sunouchi, and H. Tango, "Three-Dimensional Analysis Of Subthreshold Swing And Transconductance For Fully Recessed Oxide (Trench) Isolated 1/4 Micron Width MOSFETs", *IEEE Trans. Elec. Dev.*, vol. 35, no. 7, pp. 945-951, July 1988.

18. J.-C. Marchetaux, B. S. Doyle, and A. Boudou, "Interface States Under LOCOS Birds Beak Region", *Solid-State Electronics*, vol. 30, no. 7, pp. 745-753, 1987.

19. S. H. Goodwin and J. D. Plummer, *Electrical Performance And Physics Of Isolation Region Structures For VLSI*, ED-31, pp. 861-872, July 1984.

20. S. W. Sun and K. C. Weng, *Solid State Electronics*, vol. 32, no. 4, pp. 333-336, 1989.

21. J. W. Lutze and J. P. Krusius, "Electrical Limitations Of Advanced LOCOS Isolation For Deep Submicrometer CMOS", *IEEE Trans. Elec. Dev.*, vol. 38, no. 2, pp. 242-245, Feb. 1991.

22. E. Dubois, J-L. Coppee, B. Baccus, and D. Collard, "A Study Of The Electrical Performance Of Isolation Structures", *IEEE Trans. Elec. Dev.*, vol. 37, no. 6, pp. 1477-1486, June 1990.

23. J-L. Coppee and F. Van De Wiele, "Experimental Study of the Electrical Performance of Isolation Structures", *IEEE Electron Device Letters*, vol. 9, pp. 536-538, Oct. 1988.

24. M. Thurner and S. Selberherr, "Three-Dimensional Effetcs Due To The Field Oxide In MOS Devices Analyzed With MINIMOS 5", *IEEE Trans. Comp-Aid. Des*, vol. 9, no. 8, pp. 856-867, Aug 1990.

25. C. Coifi, M. Macucci, and B. Pellegrini, "A New Measurement Method Of MOS Transistor Parameters", *Solid-State Electronics*, vol. 33, no. 8, pp. 1065-1069, 1990.

26. J. Scarpulla and J. P. Krusius, "Improved Statistical Method For Extraction Of MOSFET Effective Channel Length And Resistance", *IEEE Trans. Elec. Dev.*, vol. ED-34, no. 6, pp. 1354-1359, June 1987.

27. B. Ricco, L. Selmi, and E. Sangiorgi, "A Novel Method To Determine The Source And Drain Resistances Of Individual MOSFETS", in *IEDM*, pp. 122-125, 1988.

28. J. Mavor, M. A. Jack, and P. B. Denyer, in *Introduction to MOS LSI Design*, Addison-Wesley, 1983.

29. M. P. Karnett and R. E. Dunham, "An Improved SEM Technique For Accurate Determination OF MOSFET Channel Length", *SPIE Integrated Circuit Metrology, Inspection, and Process Control*, vol. 921, pp. 440-447, 1988.

30. H. J. Voss and W. L. Engl, "The Influence Of Isolation Technique On MOS Transistor Characteristics, Measurements And Simulations", *Solid-State Electronics*, vol. 33, pp. 419-425, 1990.

31. G. J. Hu, C. Chang, and Y-T Chia, "Gate-Voltage-Dependent Effective Channel Length And Series Resistance of LDD MOSFETs", *IEEE Trans. Elec. Dev.*, vol. ED-34, no. 12, pp. 2469-2475, Dec. 1987.

32. Y-T Chia and G. J. Hu, "A Method to Extract Gate-Bias-Dependent MOSFET's Effective Channel Width", *IEEE Trans. Elec. Dev.*, vol. 38, no. 2, pp. 424-427, Feb 1991.

33. C-P. Wan, H. Yang, and B. J. Sheu, "A New Method To Determine Effective Channel Widths of MOS Transistors for VLSI Device Design", in *Proceedings of the 1990 International Conference on Microelectronic Test Structures*, pp. 217-220.

34. N. D. Arora, L. A. Bair, and L. M. Richardson, "A New Method To Determine The MOSFET Effective Channel Width", *IEEE Trans. Elec. Dev.*, vol. 37, no. 3, pp. 811-814, Mar 1990.

35. M. J. Deen and Z. P. Zuo, "Edge-Effects In Narrow-Width MOSFETs", *IEEE Trans. Elec. Dev.*, vol. 38, no. 8, pp. 1815-1819, Aug 1991.

36. P. I. Suciu and R. L. Johnston, "Experimental Derivation of the Source And Drain Resistance of MOS Transistors", *IEEE Trans. Elec. Dev.*, vol. ED-27, no. 9, pp. 1846-1848, Sep 1980.

37. P. A. van der Plas, N. A. H. Wils, and R. de Werdt, "Geometry Dependent Birds Beak Formation For Submicron LOCOS Isolation", in *ESSDERC*, pp. 131-134, 1989.

38. H. Umimoto and S. Odanaka, "Three-Dimensional Numerical Simulation Of Local Oxidation Of Silicon", *IEEE Trans. Elec. Dev.*, vol. 38, no. 3, pp. 505-511, Mar 1991.

39. T. Mizuno, S. Sawada, S. Maeda, and S. Shinozaki, "Oxidation Rate Reduction In The Submicrometer LOCOS Process", *IEEE Transactions On Electron Devices*, vol. ED-34, no. 11, Nov 1987.

40. N. Guillemot, G. Pananakakis, and P. Chenevier, "A New Analytical Model Of The Birds Beak", *IEEE Transactions On Electron Devices*, vol. ED-34, no. 5, pp. 1033-1038, May 1987.

41. K. Yamabe and K. Imai, "Nonplanar Oxidation and Reduction of Oxide Leakage Currents at Silicon Corners by Rounding-Off Oxidation", *IEEE Trans. Elec. Dev.*, vol. ED-34, no. 8, Aug 1987.

42. T. Shankoff, T. T. Sheng, R. B. Marcus, S. E. Haszko, and T. E. Smith, "Bird's Beak Configuration And ELimination Of Gate Oxide Thinning Produced During Selective Oxidation", *Electrochemical Society: Solid State Science And Technology*, pp. 216-222, 1980.

43. Y. Sakina, T. Ohno, and S. Matsumoto, "Two-Dimensional Analysis of Thermal Oxidation of Silicon", *Japanese Journal of Aplied Physics*, vol. 22, no. 8, pp. L514-L516, 1983.

44. D-B. Kao, J. P. McVittie, W. D. Nix, and K. C. Saraswat, "Two-Dimensional Thermal Oxidation of Silicon - I. Experiments", *IEEE Trans. Elec. Dev.*, vol. ED-34, no. 5, May 1987.

45. F. M. Klaassen and W. C. J. De Groot, "Modelling Of Scaled-Down Transistors", *Solid-State Electronics*, vol. 23, pp. 237-242, 1980.

Chapter 5

SEM and Modelled Results

5.1. Introduction

The parameters of interest in the device width investigation are the topographic profiles at the device edges. To assess their impact, a systematic approach is required. These edge profiles are determined by the isolation technique and the *LOCOS* process has been selected for study since it is currently the industry workhorse. Adjustment of the *LOCOS* process parameters, however, can produce the extremes of topography which characterise other isolation techniques. The items of interest along the transition edge are

(i) the existence of a corner or step in the silicon

(ii) the length of the bird's beak

(iii) the silicon slope angle from field to active area.

Others parameters such as the gate and field oxide thicknesses, were held constant. To achieve this, two process parameters were varied: the pad oxide and the masking nitride.

5.2. Fabrication

A test chip of transistors of varying lengths and widths was used as a vehicle for this study. The layout utilised a $2 \times N$ pad array, with a 960 μm pitch for ease of

probing, in a standard in-house design for SPICE parameter extraction. Three pad connections were used to contact source, gate and drain, with sets of three identical devices grouped together using the layout is shown in figure 5.1.

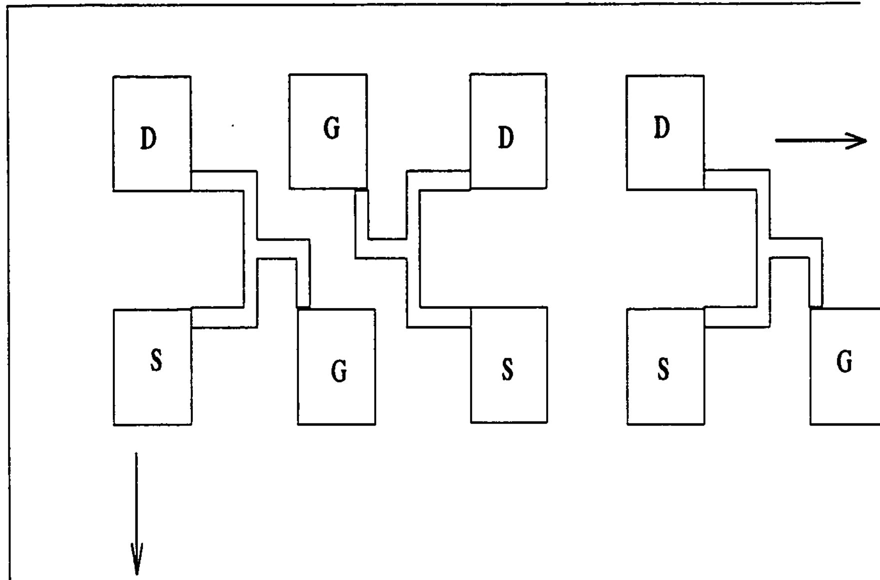


Figure 5.1. Device Layout.

The minimum drawn width was 1.0 μm, which was approximately the same dimension as the nominal bird's beak length. The transistor dimensions are shown in table 5.1.

Length (μm)	Width (μm)									
	1.0	1.5	2.5	5.0	7.0	10.0	12.0	15.0	20.0	25.0
5.0	+	+	+	+	+	+	+	+	+	+
10.0		+	+	+		+				

Table 5.1. Drawn Device Dimensions.

The process was a standard silicon gate NMOS. The full process run is documented in Appendix A. The drawn width defines the active area at first photo. Therefore the variable parameters are at first photo and thereafter the processing remains standard. A matrix of pad oxide and nitride thicknesses were chosen. Oxide thicknesses of 0.005, 0.014, 0.025, 0.037, 0.050, 0.063, 0.080, 0.093 μm were matched with 6 different nitride thicknesses of 0.037, 0.082, 0.170, 0.330, 0.440, 0.55 μm, to produce 48 samples in total. The gate oxide was .08μm and the field 1.0μm. The boron field implant was $2e13$ at an energy of $100keV$ and the V_T adjust $4e11$ at $40keV$. For reference the samples were identified as shown in table 5.2

T_{pad} (Angstroms)	T_n (Angstroms)					
	370	820	1700	3300	4400	5500
50	3D	5D	11D	15D	19D	23D
140	3C	5C	11C	15C	19C	23C
250	3B	5B	11B	15B	19B	23B
370	3A	5A	11A	15A	19A	23A
500	1D	7D	9D	13D	17D	21D
630	1C	7C	9C	13C	17C	21C
800	1B	7B	9B	13B	17B	21B
930	1A	7A	9A	13A	17A	21A

Table 5.2

5.2.1. The Process Variations

The pad oxide is usually not removed before the field oxidation and therefore the field implant will normally be through this layer. However in this case, the pad oxide thicknesses were variable. Thus, to ensure a common implant energy, and therefore a common implant profile, this oxide was etched prior to implant.

After field oxidation, the nitride and pad oxide were removed in an isotropic etch. This is one instance of the etchback process, as discussed in step D in Chapter 4, §4.3.5. In the fabrication process, the white ribbon effect was also present and so a sacrificial oxidation was performed. A further etchback followed, prior to gate oxidation. The combined effect of these etches is to extend the gate dimension down the bird's beak as discussed in the previous chapter. This effect is examined in detail in the next two sections.

5.3. Results 1: Topography

The preparation of SEM samples is a laborious and time consuming process. The method of [1] is difficult and improvement was desirable. The problem demanded the analysis of 'real' devices as opposed to conveniently sized test structures and so the following procedure was developed. The wafers were positioned on a dicing saw wherein alignment was sufficient to scribe the wafers to within a few microns of the desired location. If the scribe were too far from the feature, polishing time became excessive and if too close, then the feature would be lost in the subsequent snapping of the sample.

The wafers snapped easily along the line of scribe and this ensured a good perpendicular cleave. The samples were then mounted by wax adhesive to a holder which was then placed edge-on to a grinding/polishing surface. The sample was polished to a smooth finish. Care must be taken to maintain the angle of cleave,

otherwise extension of the profile proportional to the cosine of the cleave angle occurs. Once polished, about 0.1 μm of oxide was removed by dipping the sample edge into HF. The vertical oxide profile was protected from the HF etch by the polysilicon gate. This then created a step to improve the SEM image and delineate the bird's beak. The process permitted examination of discrete features smaller than 0.1 μm.

5.3.1. SEM Profiles

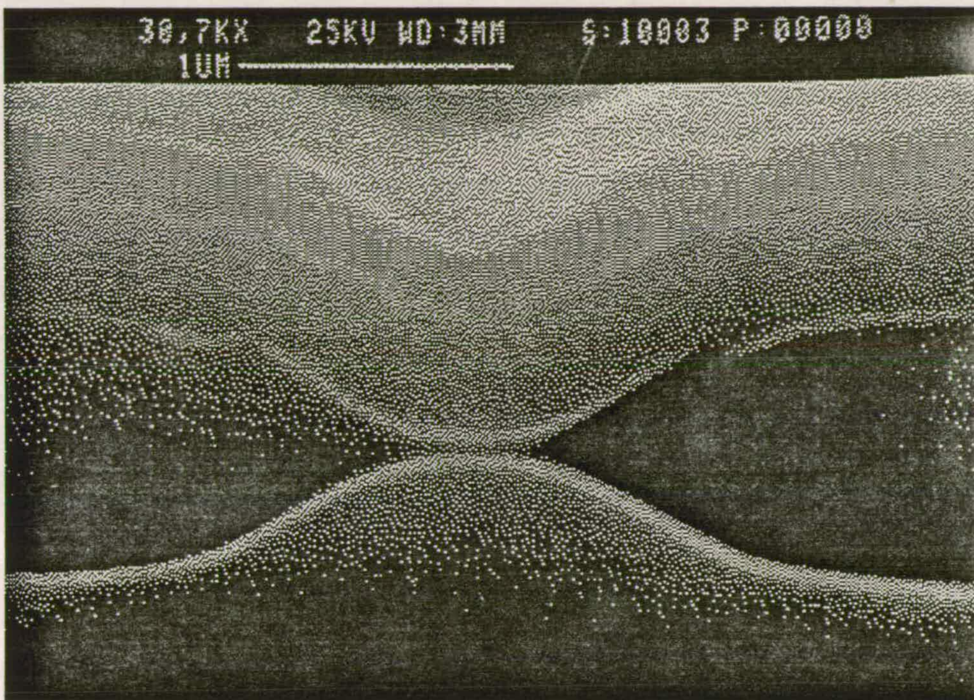


Figure 5.2. Minimum device.

Figure 5.2 shows a transistor where the thin gate region has just formed and the bird's beaks from either side are almost in contact.

Figure 5.3 shows a final profile where the etchback down the bird's beak has extended the channel. The silicon surface topography is similar to figure 5.2, though the isotropic etches have produced a curved device cross-section.

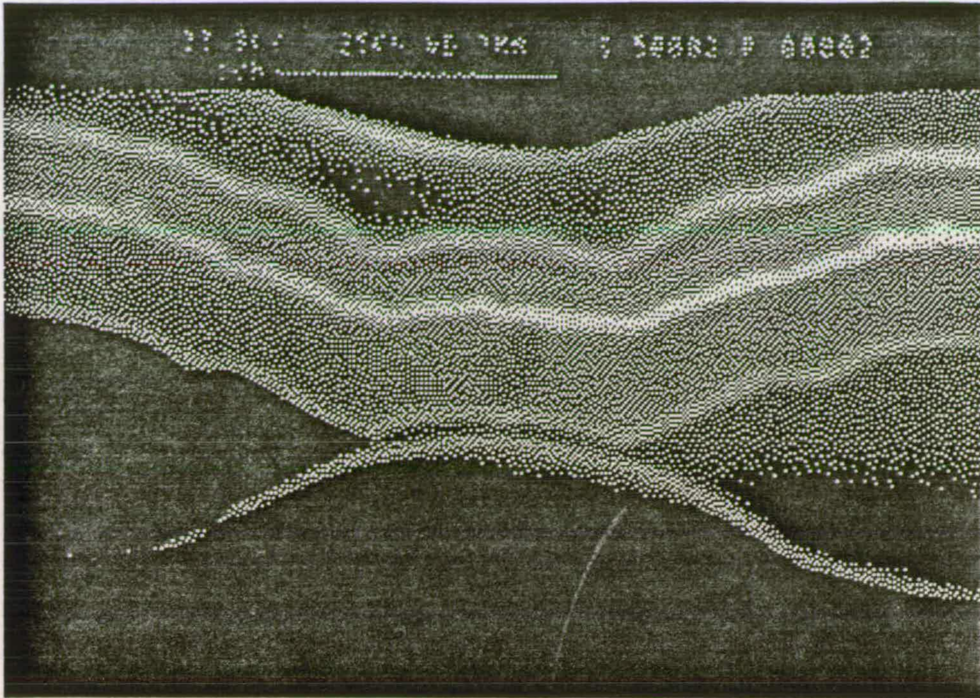


Figure 5.3. Channel extension due to etchback.

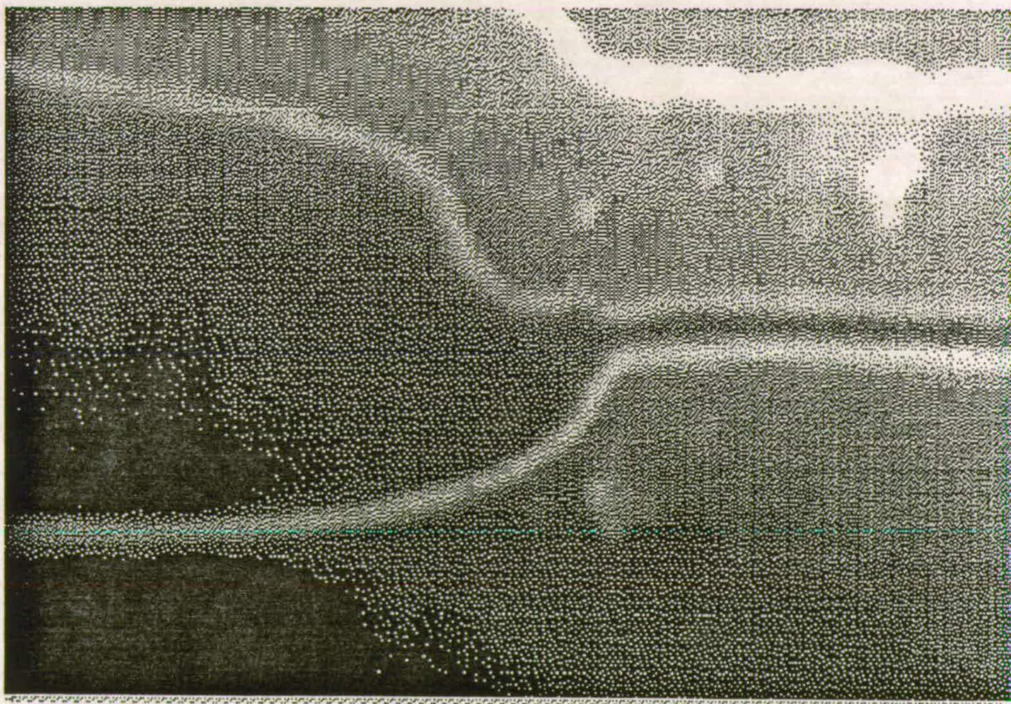


Figure 5.4. Suppression of bird's beak.

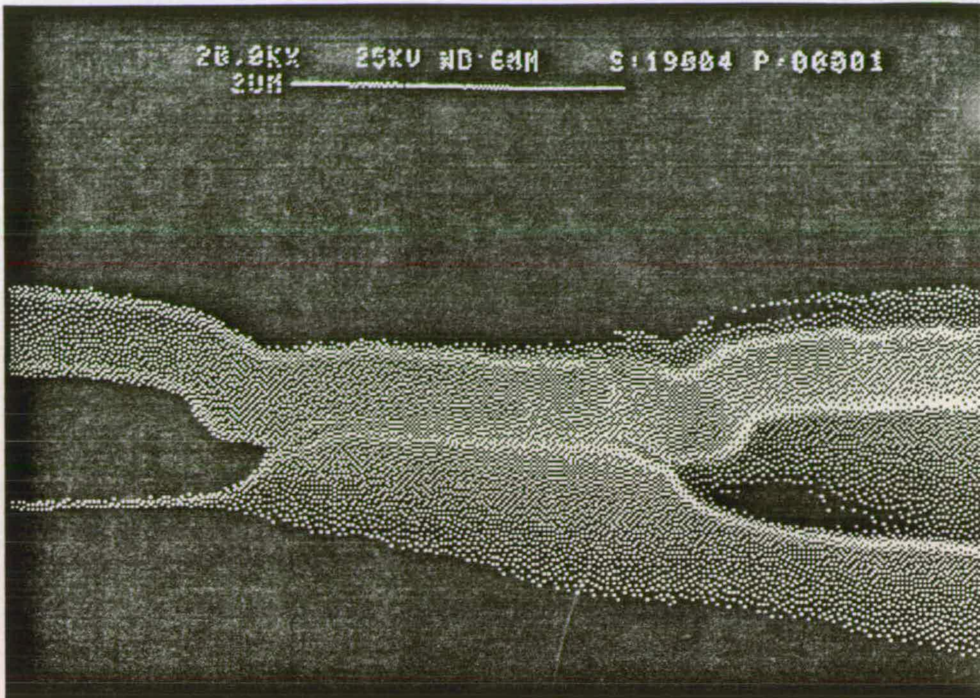


Figure 5.5. Overetch on suppressed bird's beak.

Figure 5.4 shows the suppression of the bird's beak due to the thicker nitride mask and existence of the corner can be clearly seen.

Figure 5.5 shows a suppression of the bird's beak and an extension of the topographic channel due to the etchback process.

The final bird's beak dimensions as measured by SEM are shown in figure 5.6. The contour plot shows the range of L_{bb} values as a function of T_{pad} and T_n. The maximum L_{bb} is seen to be about 1.25 μm. Therefore a width loss of 2.5 μm is expected and the minimum device the $W_D = 5$ μm transistor. The minimum L_{bb} is more difficult to assess. Etchback down the sidewalls can in extreme cases, extend the 'topographical transistor' to a width greater than the drawn width.

The shapes as defined by [2] are shown in figure 5.7 where shape 2 devices exhibit the corner feature and shape 1 devices show the smooth topography.

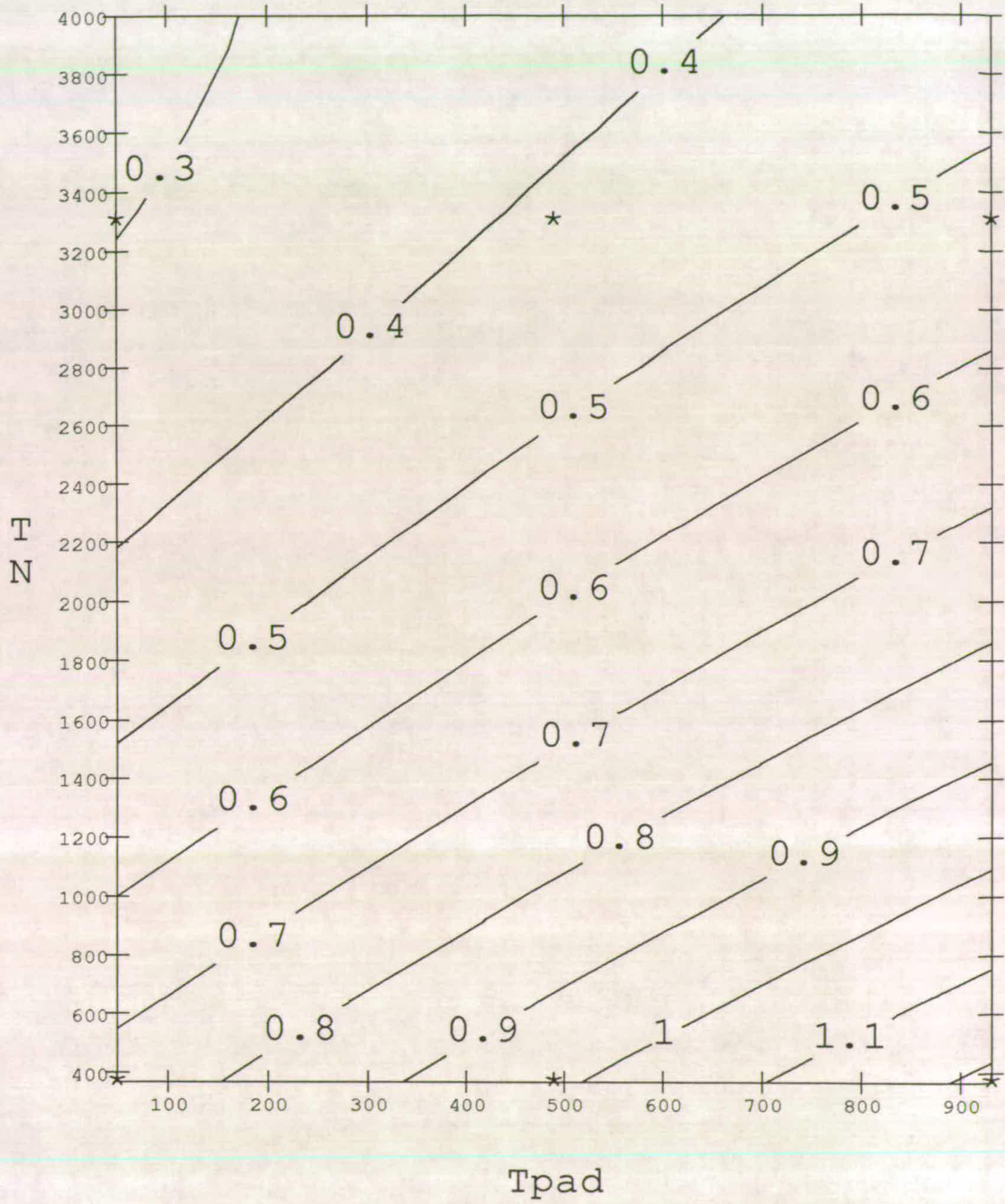


Figure 5.6. Bird's beak values at end of process.

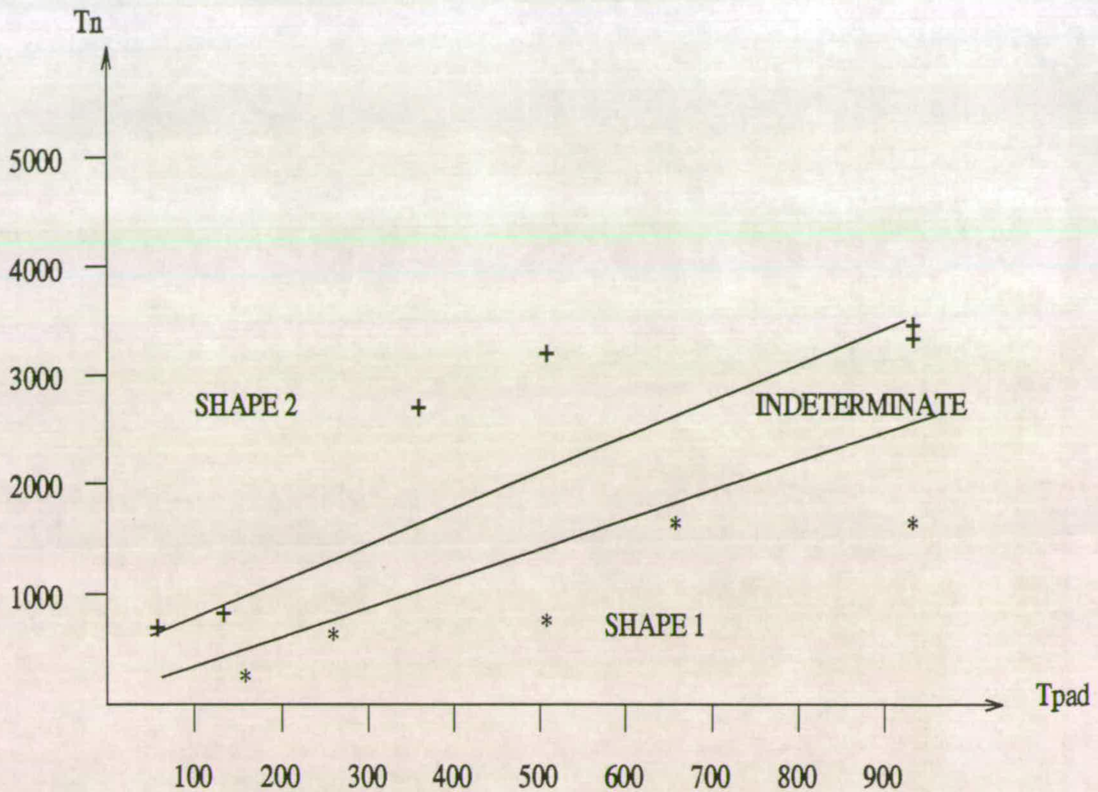


Figure 5.7. Shape categorisation for bird's beak.

5.4. Results 2: Simulation

The process was modelled using the 2D process simulator *TSUPREM*4 [3] . The results of this were used as input to the 2D device simulator *PISCES* [4] .

*TSUPREM*4 is a process simulator, which can model oxidation, diffusion and implantation in 2D. There are several model options available to the user for each of these process steps. Whilst some models sacrifice CPU time for accuracy, others reflect the lack of existing data. The processes of interest are oxidation and diffusion at the mask edge. There are 5 oxidation models available which can be divided into 2 analytical and 3 numerical models. The first analytical model uses the Deal-Grove Law [5] and a complementary error function to describe the decay of the oxidation rate under the nitride mask, akin to that of Wu [6] . Alternatively, a modified Guillemot-type [2] model may be used whereby the topographic profile is computed

as a function of the field oxide thickness, against Guillemot who measured the final shape as a function of the final field oxide thickness. As yet these have not been calibrated.

Of the three numerical models, the so-called *vertical* model is the simplest. Here the oxide growth is constrained to move in the y –direction only. Therefore it is unsuitable for the nonplanar 2D LOCOS process. The *compress* model simulates viscous flow during oxidation, but despite its name does not allow large deformations of the oxide to occur. It is considerably slower than any of the previous models. The *viscous* model differs from the compress model in that it also calculates stress in the oxide. The model fails at large stress values where there is significant compression of the oxide. At this point one of the Guillemot models must be used.

In addition to the choice of oxidation models, the diffusion models allow the simulation of point defects, which enhances the dopant diffusion under oxidising conditions. The full LOCOS process simulation then is very CPU intensive, but necessary to predict the results accurately. The range of pad oxide and nitride thicknesses dictated the choice of the numerical model (Shape 1) and the second Guillemot model (Shape 2). Despite the option of the Guillemot models, the full process simulation proved itself to be beyond the capability of the program. The etch model in *TSUPREM*4 is poor and cannot model the isotropic etching down the bird's beak.†

5.4.1. Process Modelling The Shape 1 profiles were modelled using the *compress* model. Figure 5.8 shows the final profiles of the sample 1A ($T_n = 0.037$ μm, $T_{pad} = 0.093$ μm).

---

† At the time of writing, the interface to the topography simulator *DEPICT* [7] was unavailable.

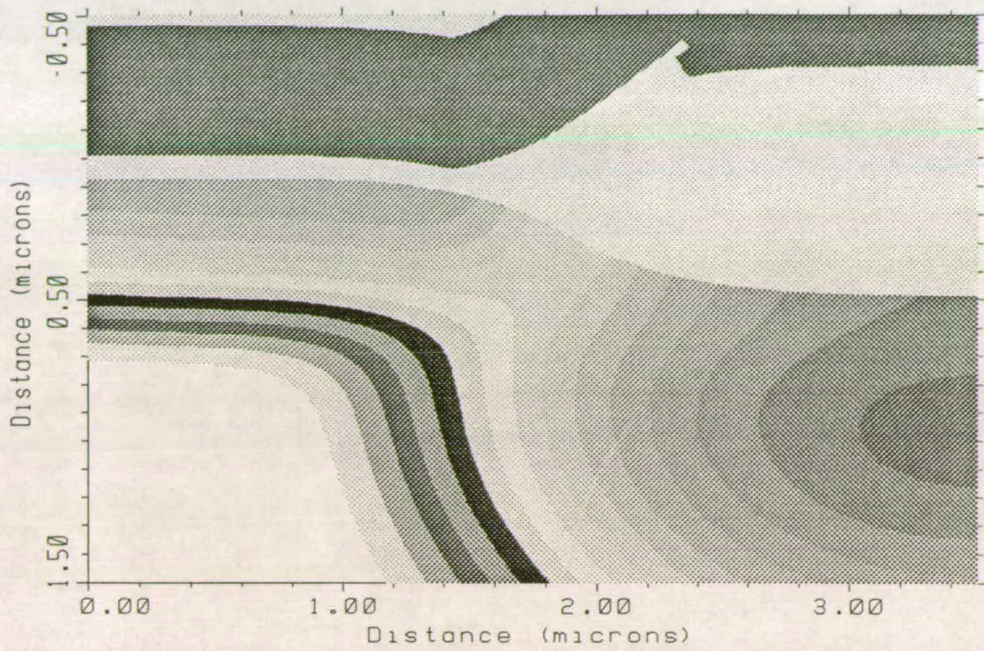


Figure 5.8. Bird's beak for 1AE.

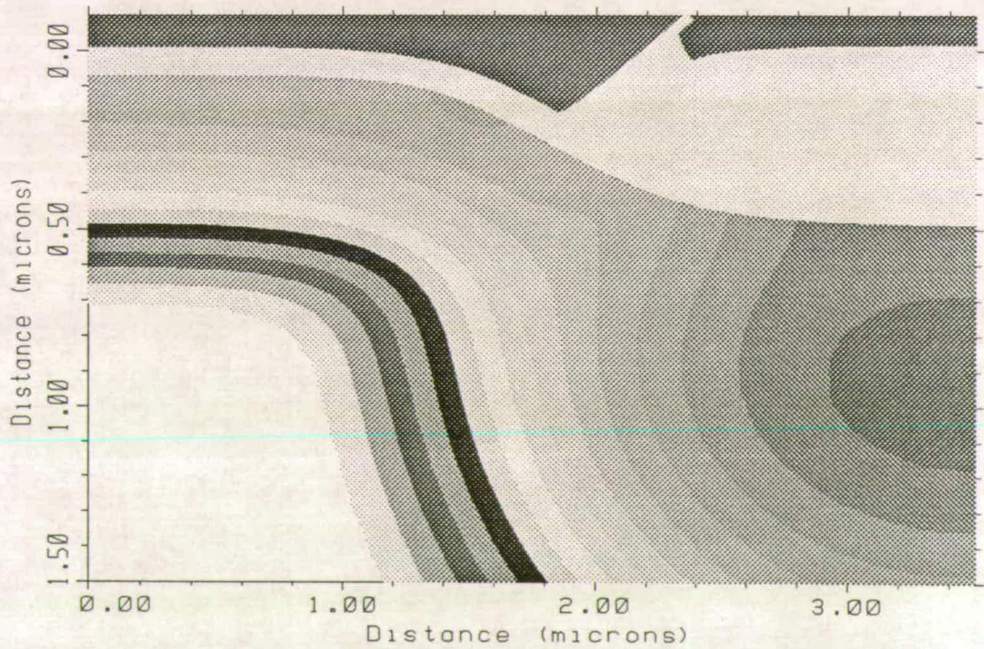


Figure 5.9 . Bird's beak showing overetch condition.

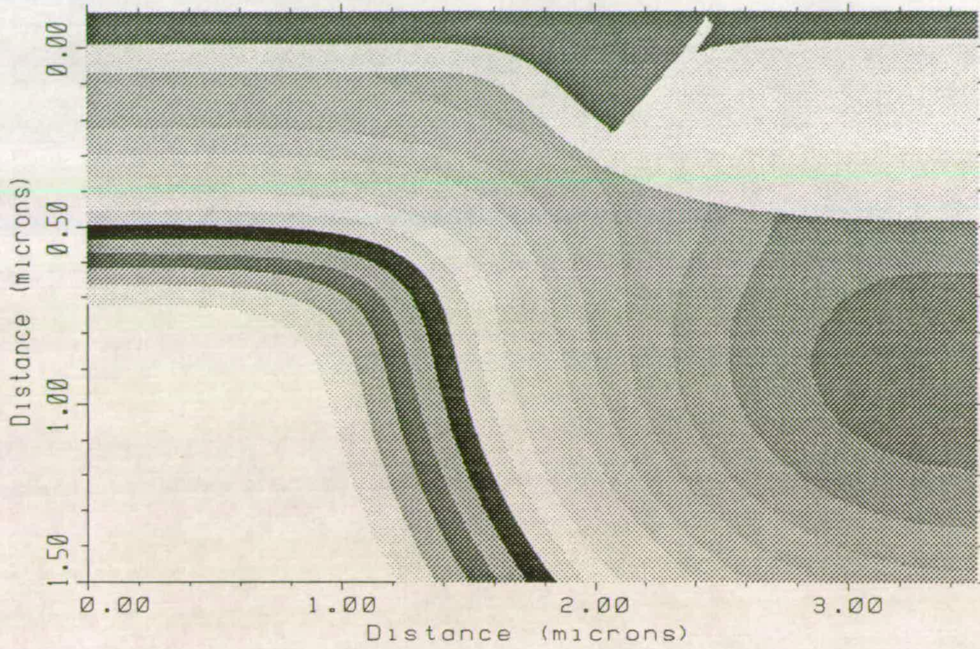


Figure 5.10. Bird's beak with large slope angle.

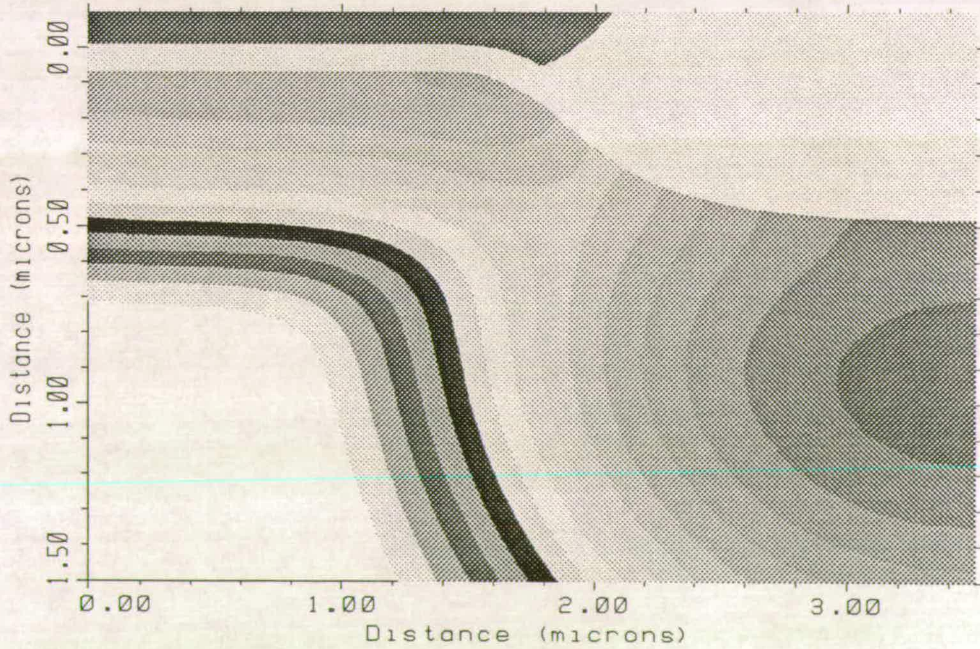


Figure 5.11. Bird's beak for thick nitride.

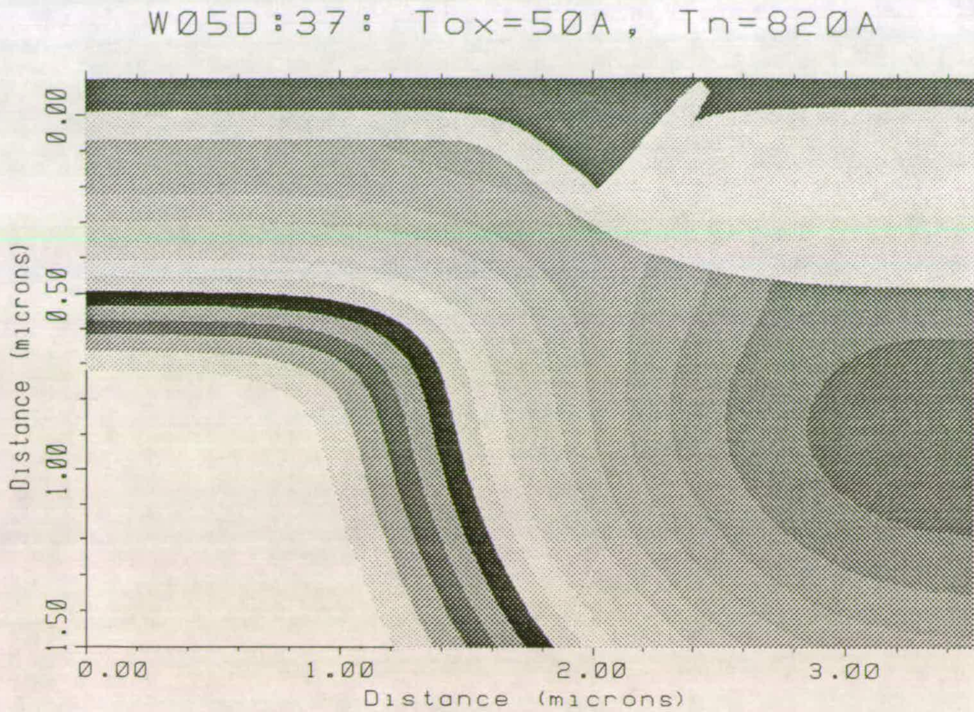


Figure 5.12. Bird's beak for thick nitride and overetch.

The characteristic bird's beak profile can be seen and the reduced concentration profile at the midpoint down the bird's beak. This is the area between the field and V_T adjust implants. Where the etchback process has removed the field oxide down the bird's beak, a concentration peak can be seen at the (soft) corner, as shown in figure 5.9

This effect is more pronounced as the slope angle of the bird's beak increases. For the same nitride thickness, but a lower oxide thickness($T_{pad} = 0.005\mu m$, sample 3D), the concentration peak as a result of overetching can be seen in figure 5.10.

For a thicker nitride the same high concentration effect can be seen, shown in figure 5.11. Also, where this sample is overetched a more uniform concentration profile is observed down the bird's beak, as shown in figure 5.12.

The reason for this can be seen in figure 5.13 where the diffusion at the corner has two components: one from the top surface and one from the sidewall.

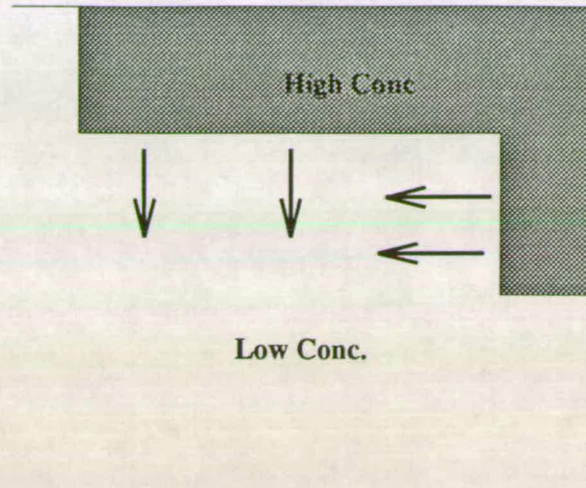


Figure 5.13. Corner Effect.

The diffusion away from the corner is inhibited until the diffusion down the sidewall reduces the concentration gradient in this direction. The presence of this concentration peak raises the local threshold voltage at this point, opposing the previously reported Inverse Narrow Width Effect. The varying surface concentration can be seen in figures 5.14 and 5.15. The larger bird's beak profiles obtained by use of thicker oxide show a greater dip in concentration profile down the silicon slope. This is reduced by overetching, which shortens the bird's beak, extending the topographical device width.

5.4.2. Device Models

The concentration profiles are, in themselves, not sufficient to predict the device widths, since the intervening gate oxide also varies in thickness. Therefore resulting data files were used as input to the device simulator *PISCES*. Keeping a substrate earth contact, the gate was biased incrementally and Poisson's equation solved each time to obtain the electron concentration throughout the device. The use of the device simulator enabled the substrate conditions to be examined during biasing.

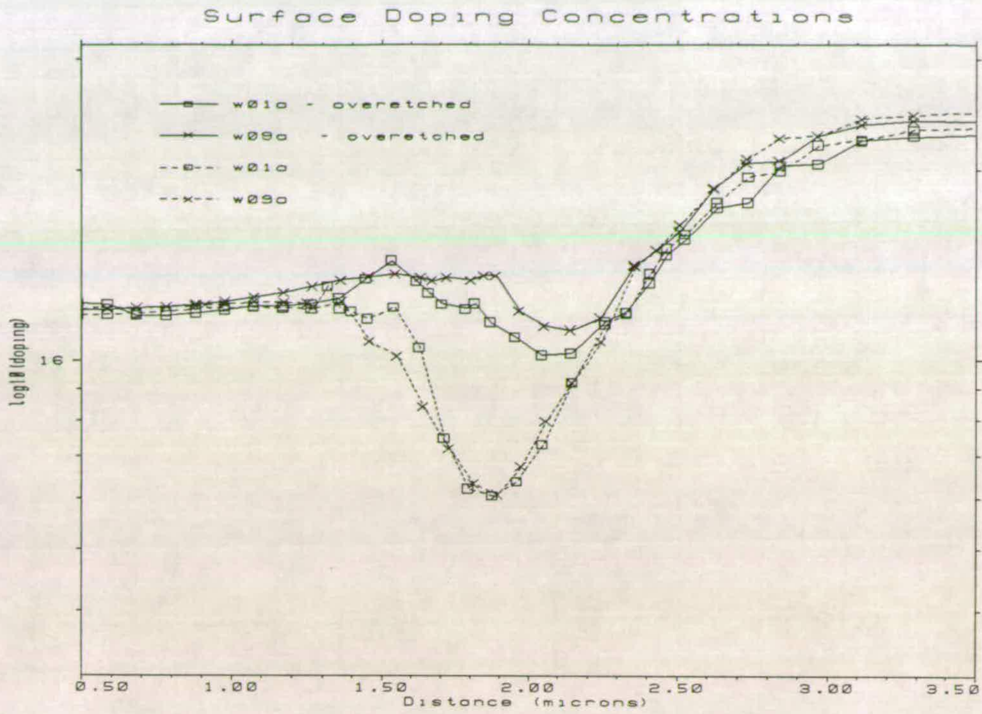


Figure 5.14. Surface concentrations for samples 1A and 9A.

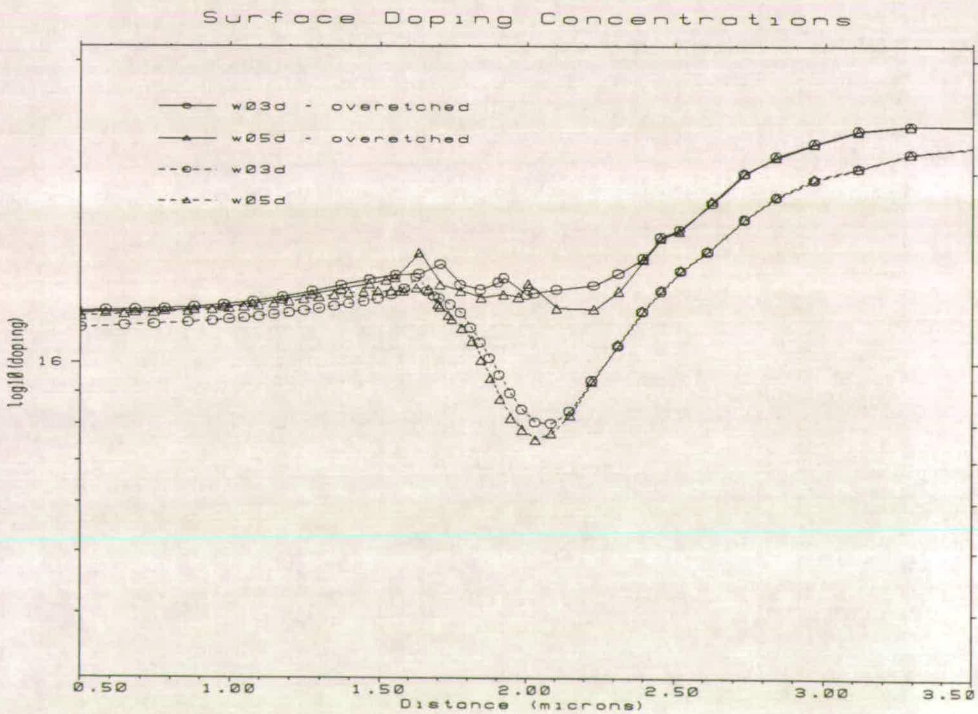


Figure 5.15. Surface concentrations for samples 3D and 5D.

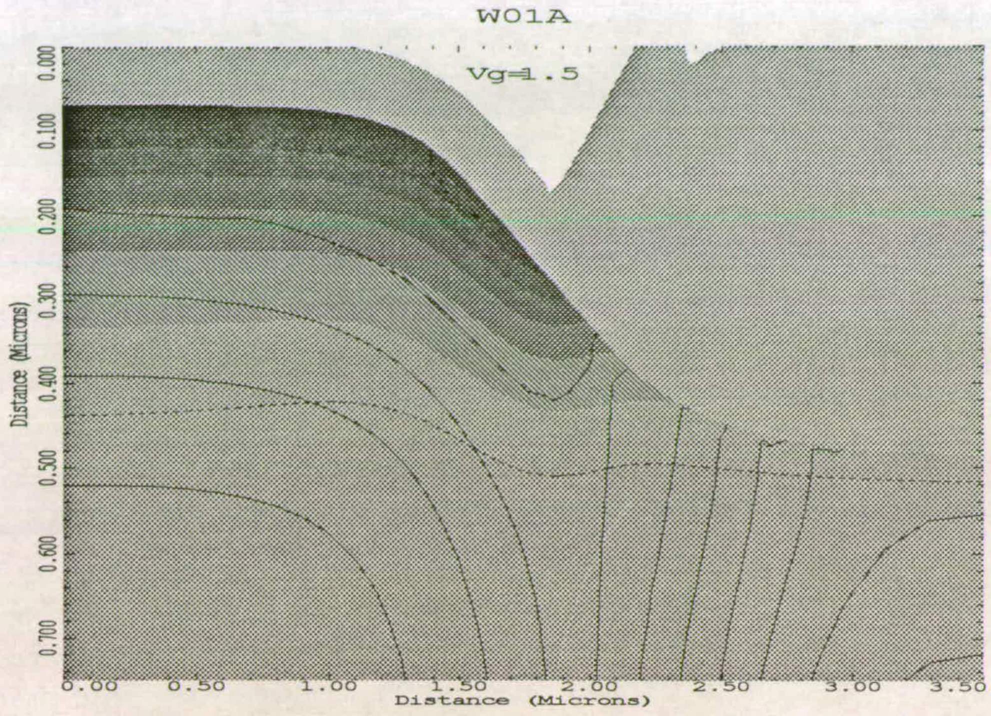


Figure 5.16. Vg=1.5 volts

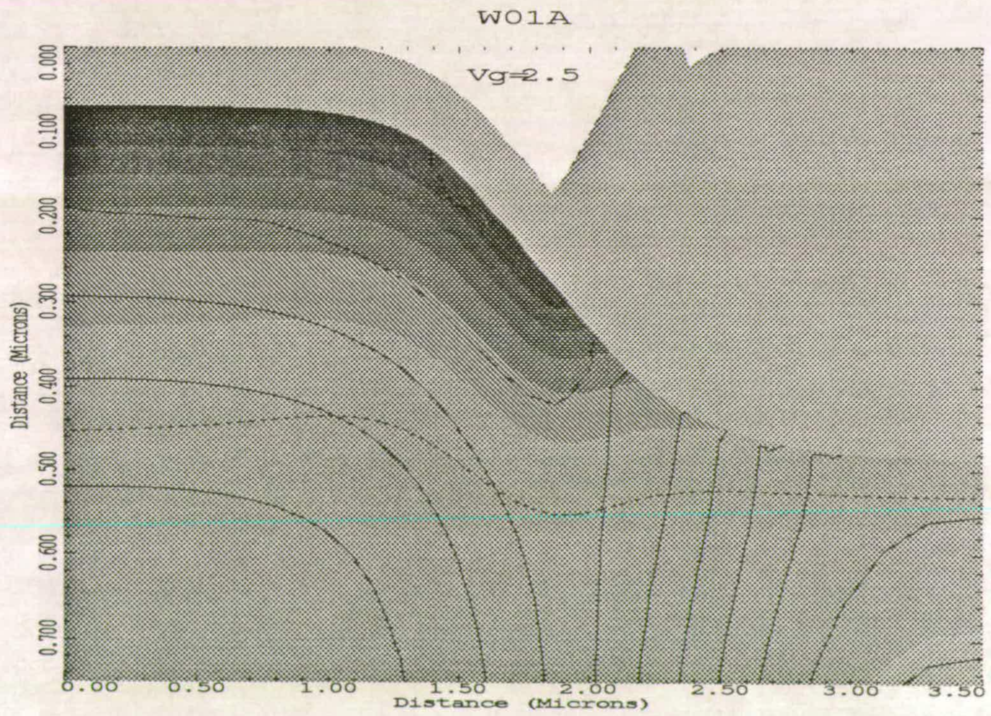


Figure 5.17. Vg=2.5 volts

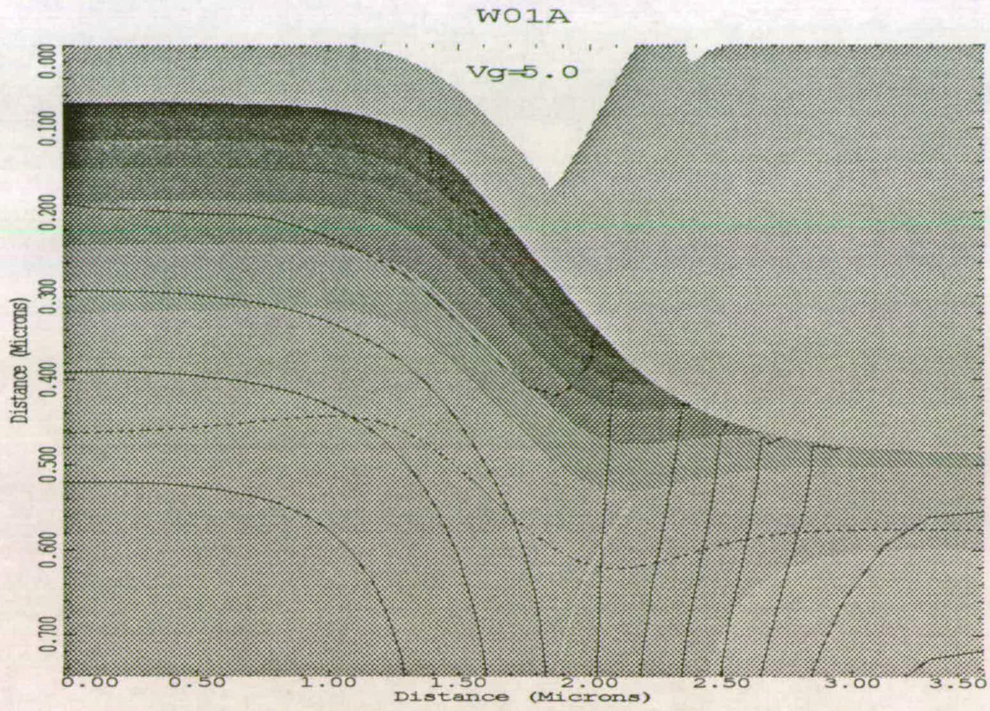


Figure 5.18. Vg=5.0 volts

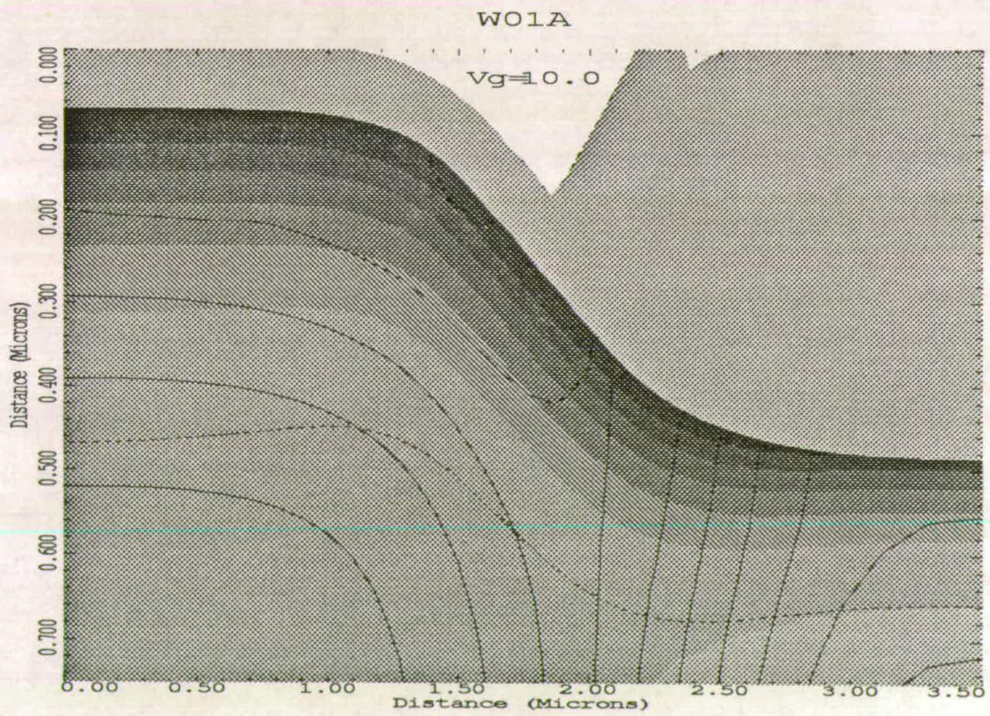


Figure 5.19. Vg=10 volts

Figures 5.16 to 5.19 show the electron concentration contours extending down the bird's beak as the gate voltage increases from 1.5V to 10V.

The determination of the the device width from this is dependent upon the definition. In their simulation work, Voss and Engl [8] defined the width at the point where the electron concentration was equal to the dopant concentration at the surface. However, as was seen in figures 5.14 and 5.15, the non-uniform surface concentration peaks at the edges of some devices, so a situation of localised inversion is possible. Alternatively, Chia and Hu [9] used the effective width definition of

$$W_{eff} = \frac{\int n(x)dx}{n_c} \tag{5.1}$$

where

n_c is the concentration in the centre of the channel

and the integration takes place across the entire surface.

Using this definition, the device simulation results were analysed to produce a graph of W_{eff} against V_{gt}. The results are shown in figure 5.20. The results show a diminishing dependence of W_{eff} against V_{gt}. This is contrary to the extracted data of Deen and Zuo, but agrees with the simulations of Chia and Hu.

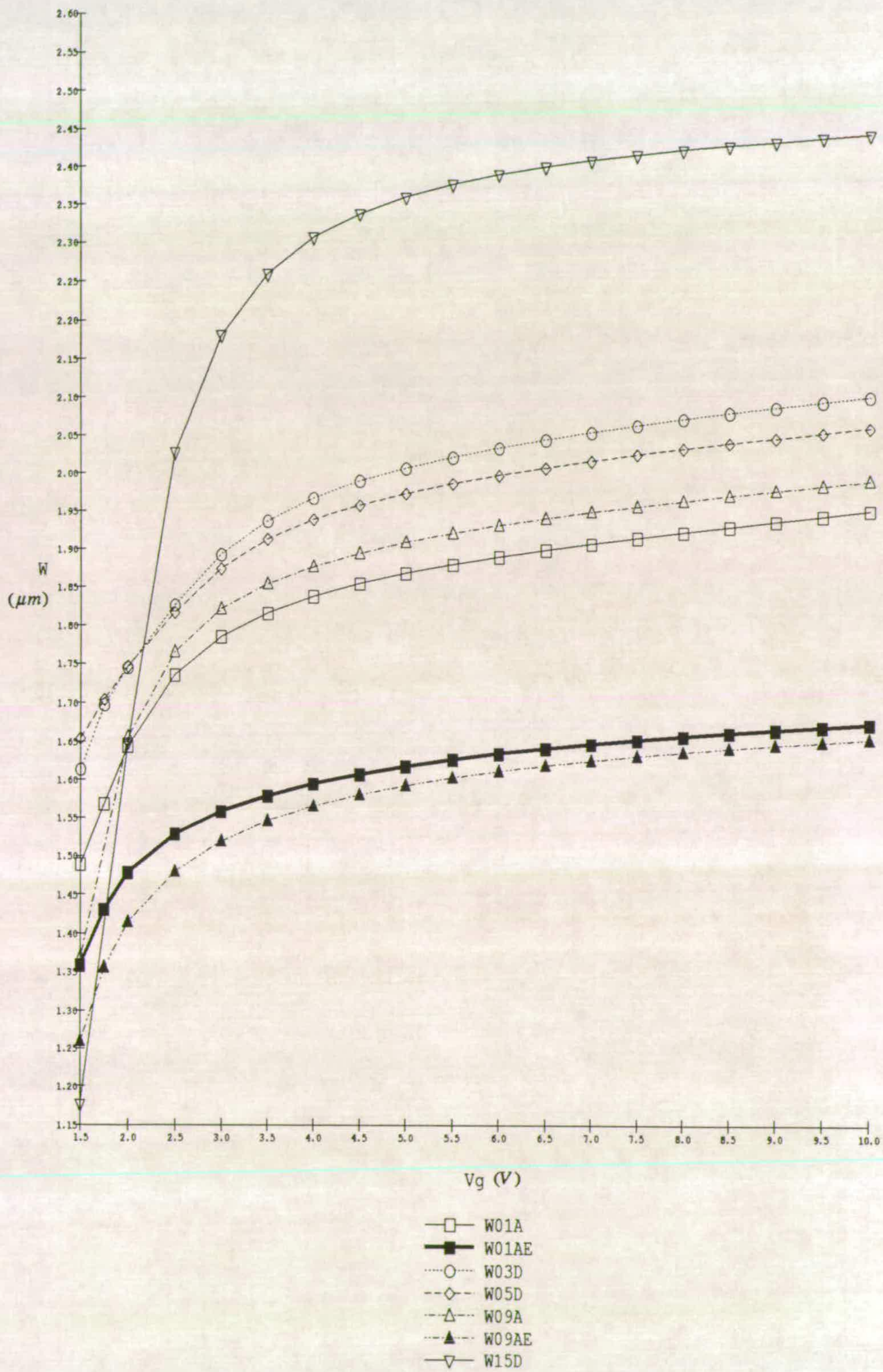


Figure 5.20. Simulation results for W_{eff}

5.5. Novel Extraction Strategy

The parameter extraction corollary of the definition in equation (5.1) is to relate the I_D value per unit width in the centre of the channel to the measured value. To this end a new method is proposed to examine the voltage dependency of effective gate width and extract meaningful edge currents. Using equation (4.7), the series resistance for each device width is extracted, by plotting V_{gt}^{-1} against R_m.

The channel resistance, R_{chan} is then given by

$$R_{chan} = R_m - R_x \tag{5.2}$$

Equation (4.4) can be re-written to produce G_t, the channel conductance, as

$$G_t = \frac{I_D}{\hat{V}_{DS}}$$

$$= \frac{W}{L} \mu C'_{ox} V_{gt} \tag{5.3}$$

The following procedure is then adopted.

(i) Extract the series resistance

(ii) Evaluate the channel conductance at each value of V_{gt}.

(iii) Compare two devices of different widths to extract G'_t, (G_t per unit width).

(iv) Divide the original value of G_t by G'_t, to obtain W_{eff}.

(v) ΔW is then given by $W_D - W_{eff}$

Step (iii) requires two devices, whereby the G_t values for the smaller device, ($G_t(W_1)$), are subtracted from the larger device, ($G_t(W_2)$). G'_t is then given by

$$G'_t = \frac{G_t(W_2) - G_t(W_1)}{W_2 - W_1} \tag{5.4}$$

To extract edge currents, the extrapolation technique of [10, 11] must be used to define a zero width when $I_{DS} \neq 0$. Figure 5.21 shows the rationale behind the edge current extraction.

Defining I_D' as the current per unit width flowing in the centre of the channel then

$$I_{D3} = (W_{D3} - \Delta W)I_D' + 2I_{edge} \tag{5.5}$$

$$I_{D2} = (W_{D2} - \Delta W)I_D' + 2I_{edge} \tag{5.6}$$

$$I_{D1} = (W_{D1} - \Delta W)I_D' + 2I_{edge} \tag{5.7}$$

Combining these three gives

$$I_{D3} - (I_{D2} + I_{D1}) = ID'(W_{D3} - (W_{D2} + W_{D1})) + \Delta W\ I_D' - 2I_{edge} \tag{5.8}$$

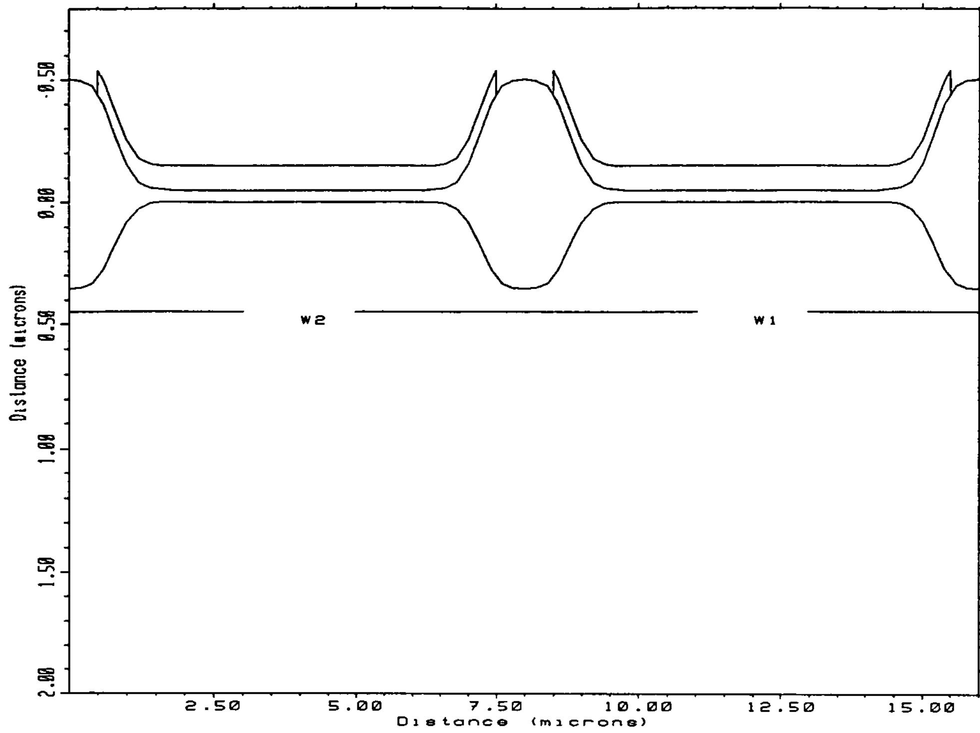
Choosing the mask widths such that

$$W_{D3} = W_{D2} + W_{D1} \tag{5.9}$$

then

$$I_{D3} - (I_{D2} + I_{D1}) = \Delta W I_D' - 2I_{edge} \tag{5.10}$$

This differs from the method of [10] in that the extracted width does not need to be known *a priori*. In the simplest case, mask widths can be chosen such that $W_{D3} = 2W_{D2}$.



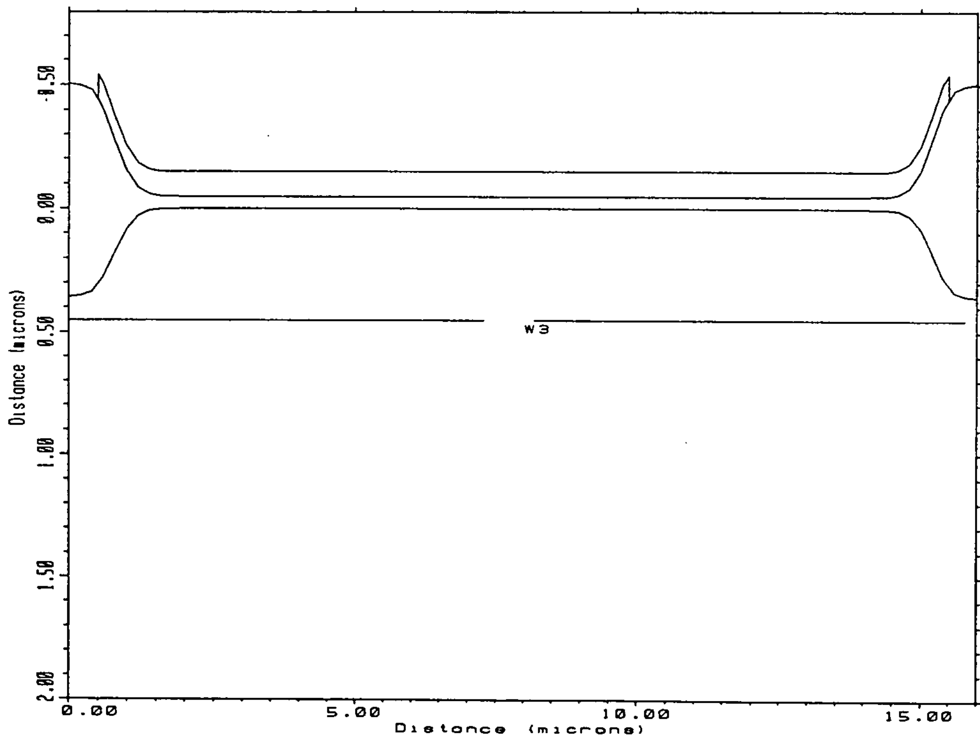


Figure 5.21. Edge current extraction.

5.6. Summary

This chapter has presented results of the topographic structures as measured by electron microscopy. The measurements have shown the variety of profiles possible and these have been classified into the two shape functions. 2D modelling has been performed to examine the electrical dependence of the device width on the applied gate voltage. A consistent relationship has been seen for the parameter range available with the simulators. A novel electrical extraction technique has been proposed to correlate the simulated and electrical techniques.

References

1. M. P. Karnett and R. E. Dunham, "An Improved SEM Technique For Accurate Determination OF MOSFET Channel Length", *SPIE Integrated Circuit Metrology, Inspection, and Process Control*, vol. 921, pp. 440-447, 1988.

2. N. Guillemot, G. Pananakakis, and P. Chenevier, "A New Analytical Model Of The Birds Beak", *IEEE Transactions On Electron Devices*, vol. ED-34, no. 5, pp. 1033-1038, May 1987.

3. *TSUPREM4*. Technology Modelling Associates

4. *PISCES*. Technology Modelling Associates

5. B. E. Deal and A. S. Grove, "General Relationship for the Thermal Oxidation of Silicon", *Journal of Applied Physics*, vol. 36, no. 12, pp. 3770-3778, 1965.

6. T. C. Wu, W. T. Stacy, and K. N. Ritz, "The Influence Of The LOCOS processing Parameters On The Shape Of The Birds Beak Structure", *Journal Of The Electrochemical Society: Solid State Science & Technology*, vol. 130, no. 7, pp. 1563-1556, July 1983.

7. *DEPICT*. Technology Modelling Associates

8. H. J. Voss and W. L. Engl, "The Influence Of Isolation Technique On MOS Transistor Characteristics, Measurements And Simulations", *Solid-State Electronics*, vol. 33, pp. 419-425, 1990.

9. G. J. Hu, C. Chang, and Y-T Chia, "Gate-Voltage-Dependent Effective Channel Length And Series Resistance of LDD MOSFETs", *IEEE Trans. Elec. Dev.*, vol. ED-34, no. 12, pp. 2469-2475, Dec. 1987.

10. M. J. Deen and Z. P. Zuo, "Edge-Effects In Narrow-Width MOSFETs", *IEEE Trans. Elec. Dev.*, vol. 38, no. 8, pp. 1815-1819, Aug 1991.

11. K. Terada and H. Muta, "A New Method To Determine Effective MOSFET Channel Length", *Japanese Journal Of Applied Physics*, vol. 18, no. 5, pp. 953-959, May, 1979.

Chapter 6

Electrical Extraction.

6.1. Introduction

Electrical extraction algorithms used to examine width effects can be classified into two categories. In the first of these all width effects are incorporated into one aggregate term, W_{eff} and no edge currents are defined i.e. $I_{edge} = 0$. In the second case, the W_{eff} term is split into two components to separate the current contribution from edge of the device i.e. $I_{edge} \neq 0$. Four methods of electrical extraction have been considered. In all cases the series resistance must be extracted first.

6.2. Series Resistance Extraction.

The series resistance was extracted using the method of Suciu and Johnson [1] . Using the notation of chapter 4, the method of extraction uses a form of equation (4.20).

$$R_m = \frac{A}{W_{eff} V_{gt}} + \left[\frac{A\theta}{W_{eff}} + R_x \right]$$

Multiplying both sides of the equation by V_{gt} and defining $b_o = \frac{A}{W_{eff}}$ produces

$$R_m V_{gt} = \left[b_o \theta + R_x \right] V_{gt} + b_o \qquad (6.1)$$

First a graph is made of $R_m V_{gt}$ *vs* V_{gt} for different device lengths. By defining $E = R_m V_{gt}$, the slope of the graph can be written as

$$\frac{dE}{dV_{gt}} = b_o\theta + R_x \qquad (6.2)$$

The intercept on the E axis is b_o. A subsequent extraction is then performed by plotting dE/dV_{gt} against b_o. The resulting slope yields θ and the intercept R_x.

Two values of drawn width were chosen, 5 μm and 10 μm. An external series resistance of 470Ω was added and the measurement repeated to verify the extraction. Typical resultant plots for three cases are shown in figures 6.1, 6.2 and 6.3. The extracted data for this sample, wafer 15, is shown table 6.1. Results for wafer 5 are shown in table 6.2. The product $R_x \times W_D$ is shown in column 7, and the correlation coefficient of the dE/dV_{gt} is shown in column 8.

This above procedure was repeated for all samples and the average value for $R_x \times W_D$ was found to be 2250 Ωμm with a standard deviation of 300 Ωμm.

6.3. Width Extraction I: $I_{edge} = 0$

Within the constraint $I_{edge} = 0$, two approaches to ΔWextraction were considered. A novel algorithm, the Channel Normalisation Method, was developed to provide a direct comparison with the simulation results presented in chapter 5. The second of these is the commonly used linear extrapolation technique[2] Each of these methods uses the channel conductance as the width dependent parameter for the extraction process. Once the series resistance, R_x, has been extracted, the channel conductance, G_t, can be found from

$$G_t = \frac{1}{R_m - R_x} \qquad (6.3)$$

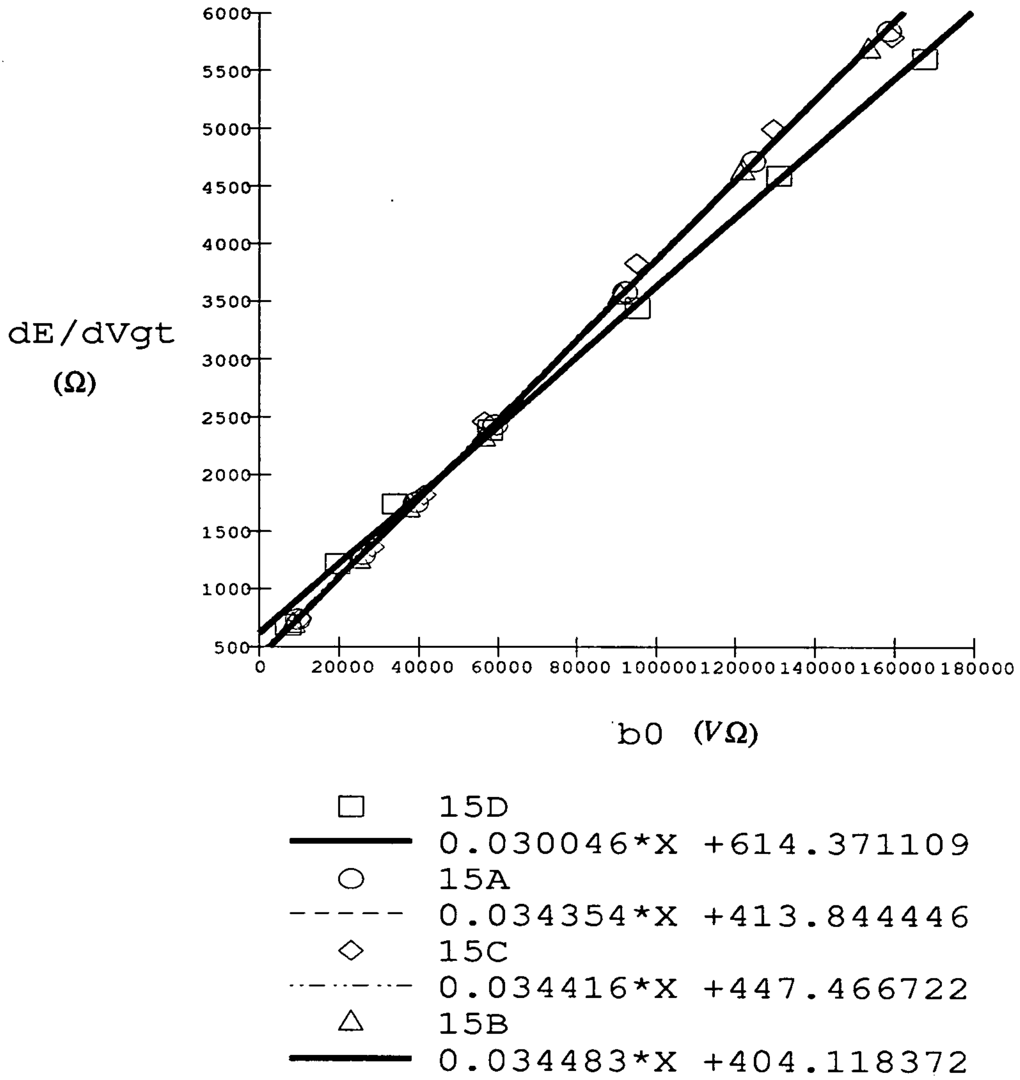


Figure 6.1. R_x extraction for $W_D = 5\ \mu m$.

6.3.1. Channel Normalisation

In this new technique the approach taken uses a width definition analogous to that of [3] as discussed in Chapter 5.

$$W_{eff} = \int \frac{n(x)dx}{n_c}$$

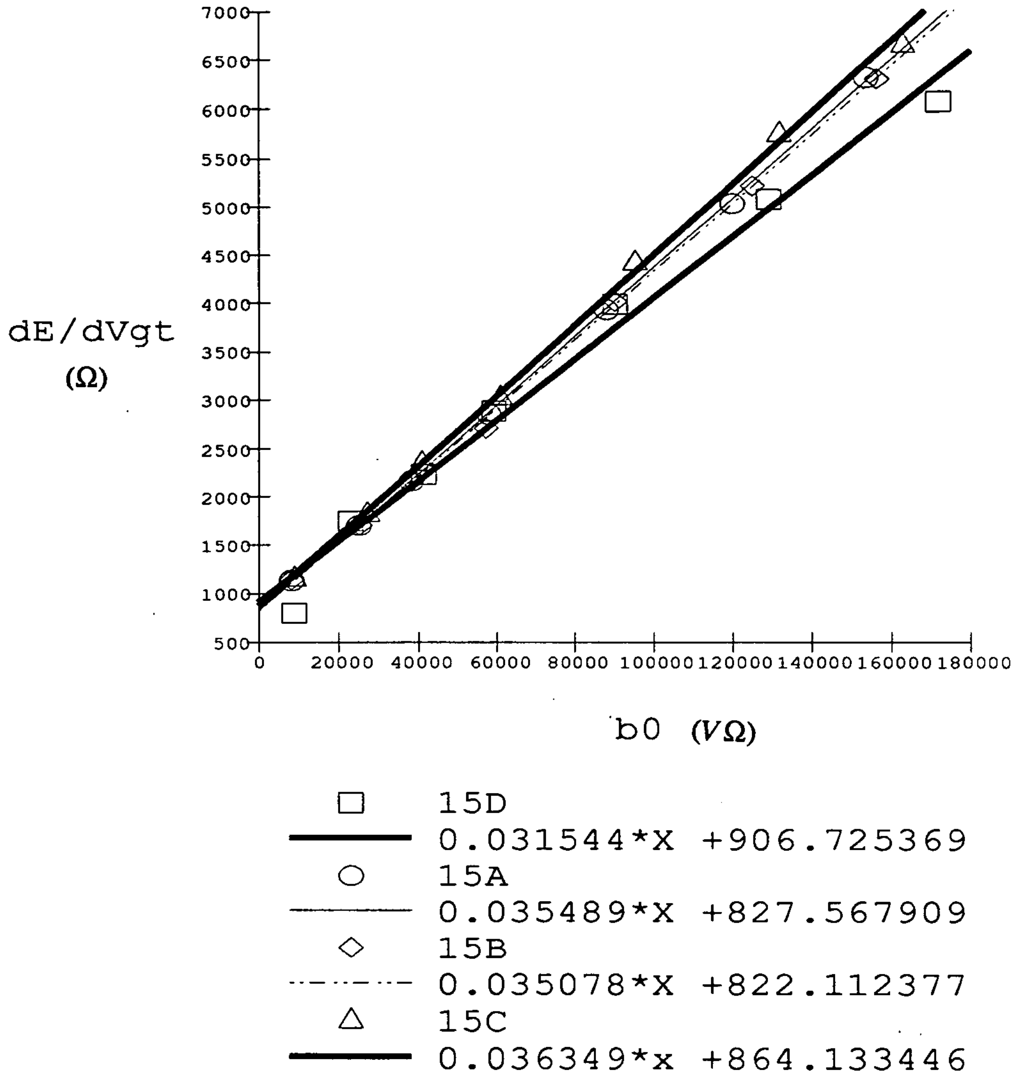


Figure 6.2. R_x extraction for $W_D = 5\mu m$. Added series resistance =470Ω.

This definition is the one used previously to produce figure 5.20 from simulated results.

Typical results are shown in figure 6.4 where ΔW is plotted against V_{gt} for three sets of paired devices used to calculate G_t'. The paired widths were : 10 & 5 μm, 12 & 7 μm and 25 &12 μm. The T_n value was 820A and the T_{pad} value was 370A. The

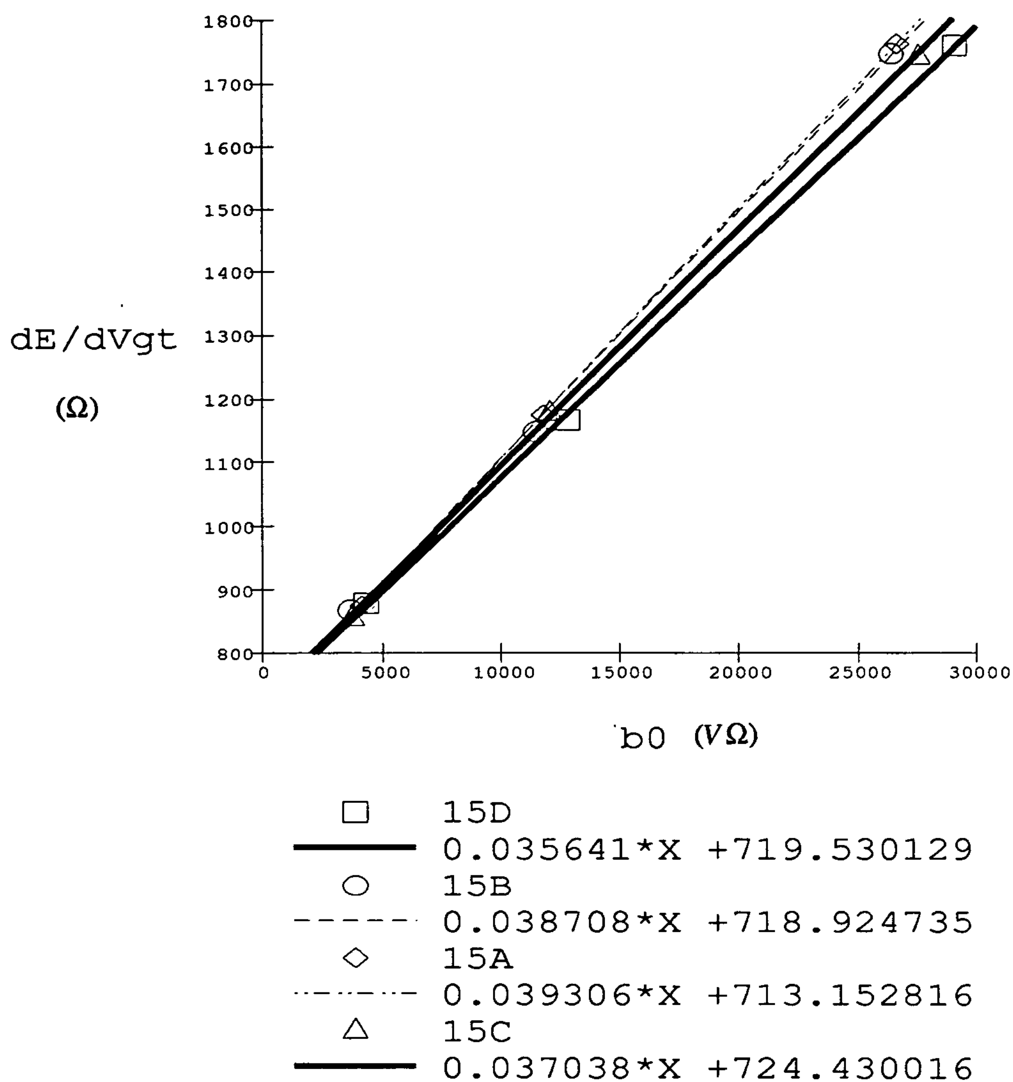


Figure 6.3. R_x extraction for $W_D = 10\mu m$. Added series resistance =470Ω.

resulting graphs show a decrease of ΔW with increasing V_{gt}, and a diminishing dependence of ΔW on V_{gt}. Whilst the trends are consistent, absolute values of ΔW vary, with large differences evident at low V_{gt}. The variability of ΔW at low V_{gt}, means that the overall change in ΔW as a function of V_{gt} cannot easily be quoted.

Sample	Drawn Width (Ω)	Added Resistance (Ω)	θ	Intercept	R_x	$R_x \times W_D$	Corr. Coeff.
15D	5	0	0.030	614	614	3070	0.998989
15C	5	0	0.034	447	447	2235	0.99889
15B	5	0	0.034	404	404	2020	0.99889
15A	5	0	0.034	413	413	2065	0.99998
15D	5	470	0.029	1123	653	3065	0.997
15C	5	470	0.036	864	394	1970	0.999
15B	5	470	0.035	822	350	1750	0.99969
15A	5	470	0.035	828	358	1790	0.999888
15D	10	470	0.036	720	250	2500	0.9998
15C	10	470	0.037	724	254	2540	0.9998
15B	10	470	0.039	719	249	2490	0.999
15A	10	470	0.039	713	243	2430	0.99999

Table 6.1. Wafer 15: Extracted Series Resistances.

6.3.2. Channel Conductance Extrapolation

An alternative technique to the channel normalisation, is to extrapolate the G_t vs W_D data for discrete values of V_{gt}. The approach uses the first two steps of the channel normalisation technique. viz.

Sample	Drawn Width (Ω)	Added Resistance (Ω)	θ	Intercept	R_x	$R_x \times W_D$	Corr. Coeff.
5D	5	0	0.032	466	466	2330	0.9999
5B	5	0	0.032	443	433	2215	0.9998
5C	5	0	0.031	519	519	2585	0.9996
5A	5	0	0.031	508	508	2540	0.9997
5D	10	0	0.038	244	244	2440	0.9998
5B	10	0	0.039	240	240	2400	0.9998
5C	10	0	0.039	250	250	2500	0.9996
5A	10	0	0.039	254	254	2540	0.9999

Table 6.2. Wafer 5: Extracted Series Resistances.

(i) Extract the series resistance

(ii) Evaluate the channel conductance at each V_{gt}

At this point, once the series resistance has been extracted, the approach of [3] is adopted.

(iii) Plot G_t against W_D for discrete values of V_{gt}

(iv) Extrapolate these to intersect on the W_D axis

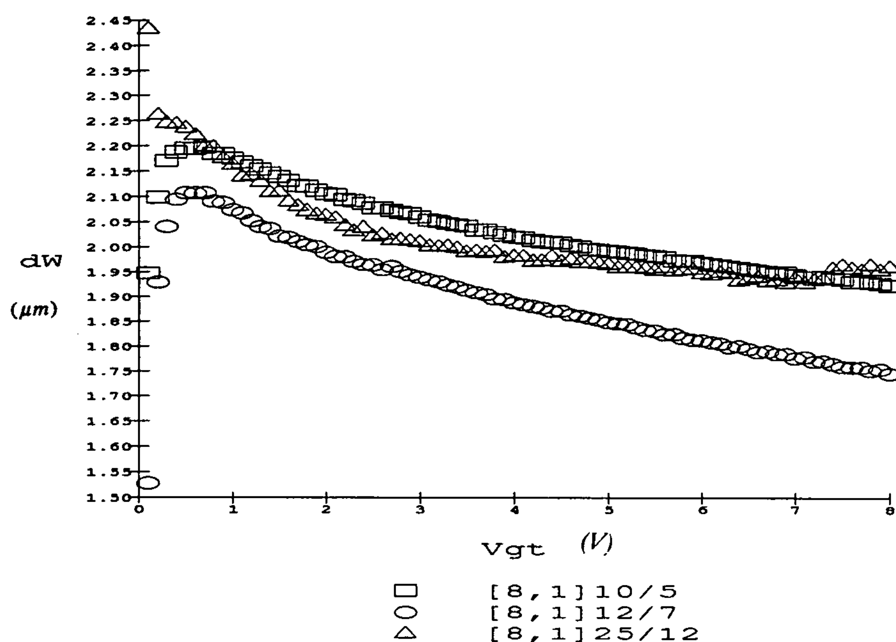


Figure 6.4. ΔW Extraction Using Channel Normalisation

These intersection points are then the ΔW values for the corresponding values of V_{gt}. Figure 6.5 shows a typical plot of G_t vs W_D. The least squares fit to each of these lines, other than $V_{gt} = 0.5$ volts, produced a correlation coefficient better than 0.999, with a coefficient of determination better than 0.999. In the case of $V_{gt} = 0.5V$, both the correlation coefficient and the coefficient of determination were better than 0.99. Figure 6.6 shows the graph of extracted ΔW *vs* V_{gt} for 4 samples, with different values of T_{pad}. The nitride thickness was 820A in all cases. Again a reduction of ΔW with increasing V_{gt} can be observed and, as with the previous technique, the spread in results is large, although the same trends can be seen:

(a) ΔW values at low V_{gt} show the largest variation.

(b) ΔW shows a diminishing dependence on increasing V_{gt}.

6.4. Width Extraction II. $I_{edge} \neq 0$

The previous extraction techniques assumed a value of zero for the edge current, I_{edge}, incorporating all width parasitics into one value, W_{eff}. To define an edge current, distinct from that flowing in the centre of the channel, it is necessary to define an existing current when the effective width of the device is zero. This approach was used by [4] and the method is similar in its approach to the extrapolation technique described in §6.3.2. The value of ΔW however is *not* given by the intercept on the W_D axis. Instead, integer values of V_{gt} were chosen and five values of G_t calculated for five values of V_{gt} within $\pm 0.2V$ of each integer value. The intercept of these five extrapolated lines is then used as a definition of ΔW, which is a function of V_{gt}. The lines are averaged to produce a single value for each V_{gt} and G_t. and a dual linear fitting technique employed to extract ΔW. The slopes and corresponding intercepts of the fitted V_{gt} - G_t lines are then plotted to extract ΔW as the negative slope.

A similar equation to (4.24) is used again as the basis for this technique. Subtracting the series resistance from the measured, yields the channel resistance, allowing (4.24) to be written as

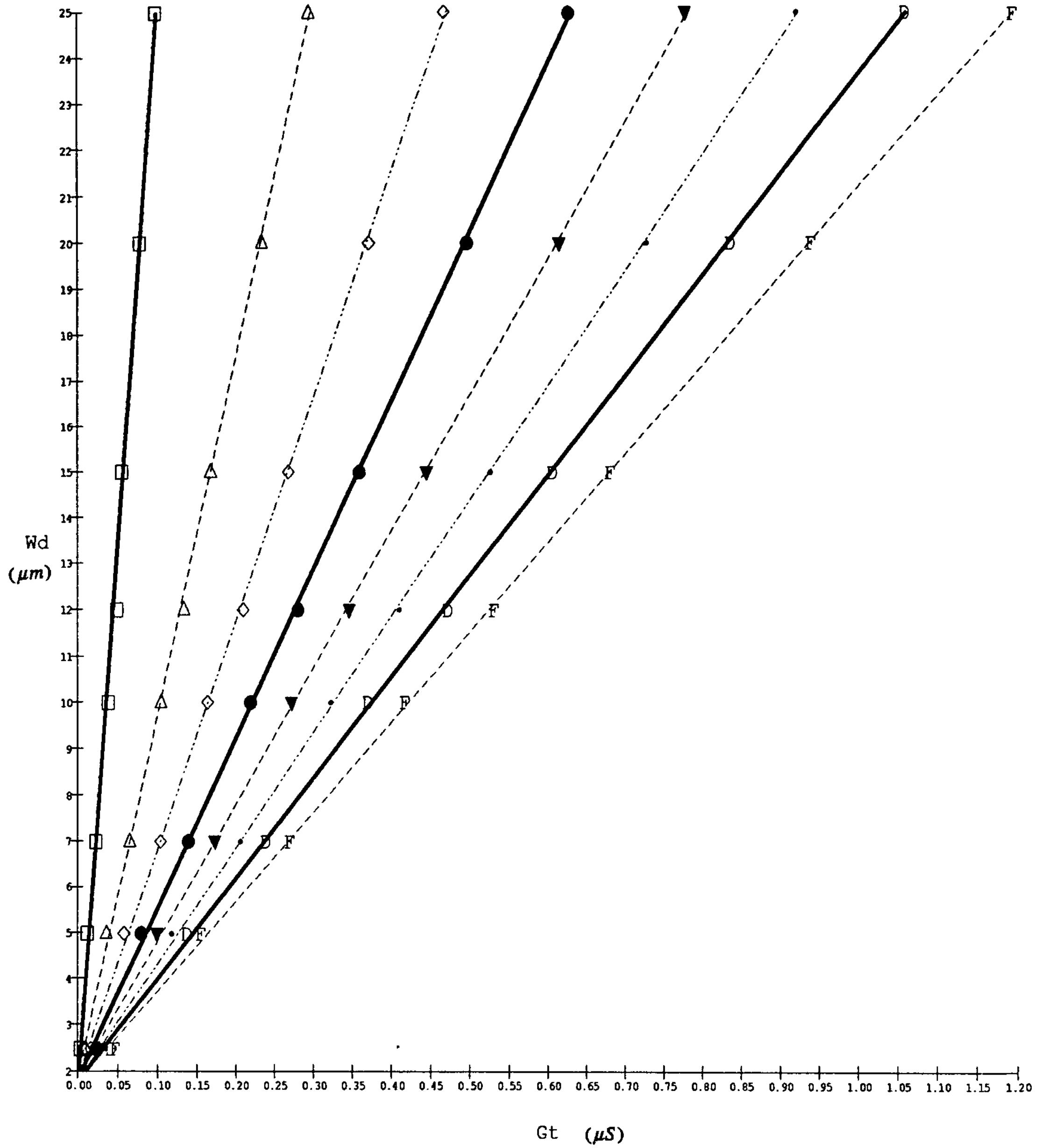
$$R_{chan} = \frac{A}{W_{eff} V_{gt}} + \frac{A\theta}{W_{eff}}$$

and therefore

$$G_t = \frac{W_{eff} V_{gt}}{A(1 + \theta V_{gt})} \tag{6.4}$$

In order to define an edge current, a parallel conductance must be added at zero effective width. Therefore the analogous equation used in [4] is

Wafer 5.



□ Vgt=0.5
—— 234.907854*X +1.541549
Δ Vgt=1.5
------ 78.107431*X +1.824884
◇ Vgt=2.5
·-·-·- 49.218298*X +1.838076
● Vgt=3.5
—— 36.793113*X +1.824934
▼ Vgt=4.5
------ 29.744585*X +1.807472
• Vgt=5.5
·-·-·- 25.169649*X +1.788891
D Vgt=6.5
—— 21.91616*X +1.774874
F Vgt=7.5
------ 19.467328*X +1.766612

Figure 6.5. G_t vs W_D showing intercept ΔW.

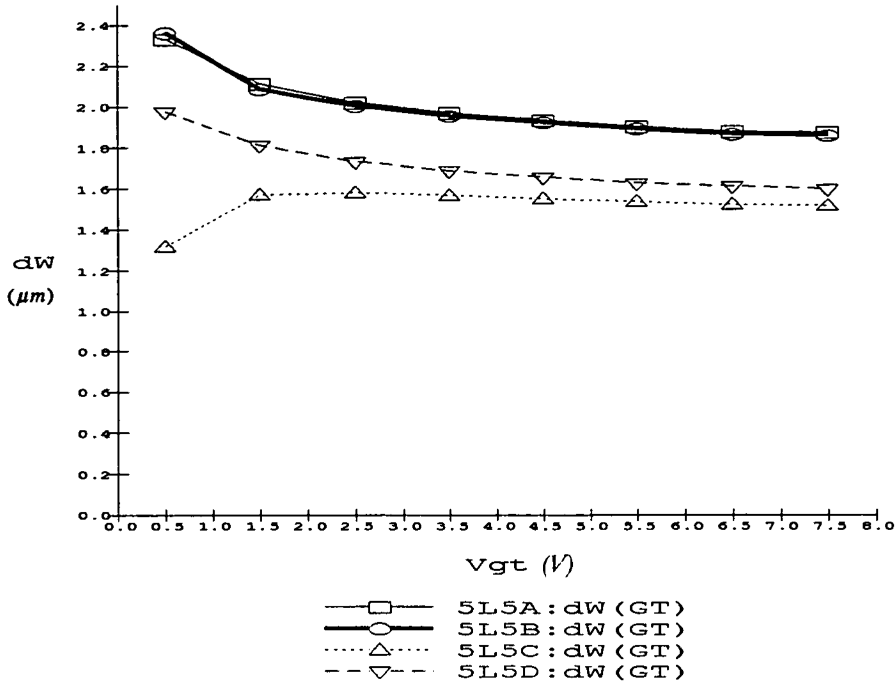


Figure 6.6 ΔW using G_t extrapolation for wafer 5.

$$G_t = \frac{W_{eff} V_{gt}}{A(1+\theta V_{gt})} + 2G_p \tag{6.5}$$

and since $W_D = W_{eff} + \Delta W$

$$G_t = \frac{W_D V_{gt}}{A(1+\theta V_{gt})} - \frac{\Delta W V_{gt}}{A(1+\theta V_{gt})} + 2G_p \tag{6.6}$$

Defining

$$P = \frac{V_{gt}}{A(1+\theta V_{gt})}$$

then

$$G_t = PW_D - (P\Delta W + 2G_p) \tag{6.7}$$

A further extraction is performed by plotting the slope, P, against the intercept $P\Delta W + 2G_p$. The slope of this graph is then given by $-\Delta W$ and the intercept gives the value of the parallel conductance. Deen & Zuo quote a correlation coefficient of better than 0.99 for their data.

An example of the $G_t - V_{gt}$ graph is shown in figure 6.7 where the V_{gt} values are 5.2, 5.1, 5.0, 4.9, and 4.8 volts. Correlation coefficients for the straight line fits are shown as better than 0.999.

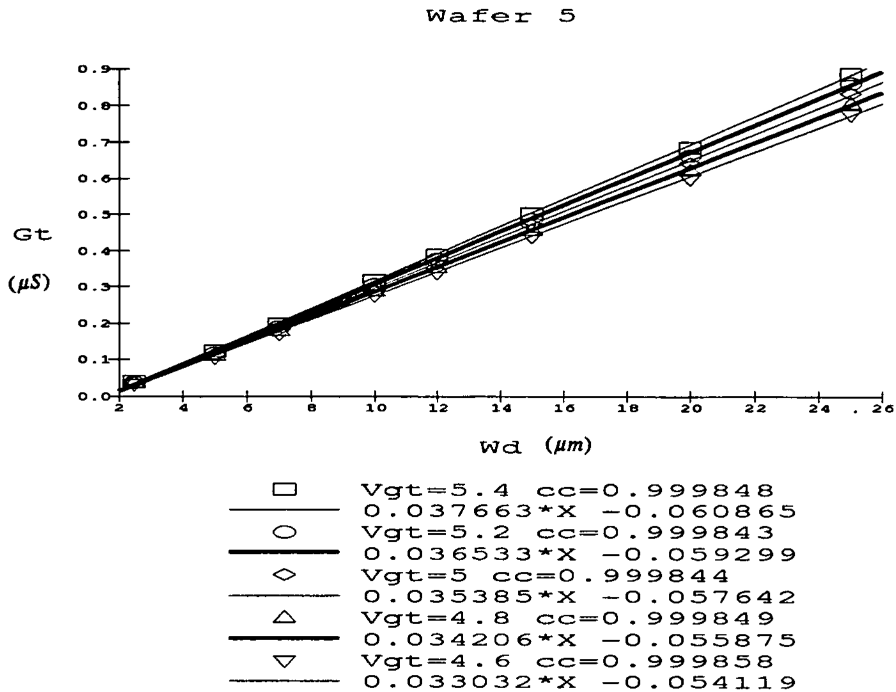


Figure 6.7 $\Delta W - G_t$ extrapolation for wafer 5.

The resultant $\Delta W - V_{gt}$ plots of these extraction routines is shown in figure 6.8.

6.5. Errors And Approximations

The comparison of the techniques show an inconsistency in the results. Whilst the general trends are evident, in each case absolute values show large variations. Two possible sources of input error in the extraction process were identified. The first of these was the spread in R_x, and the second the extrapolated threshold voltage, V_t. The effect of error in these parameters is now discussed.

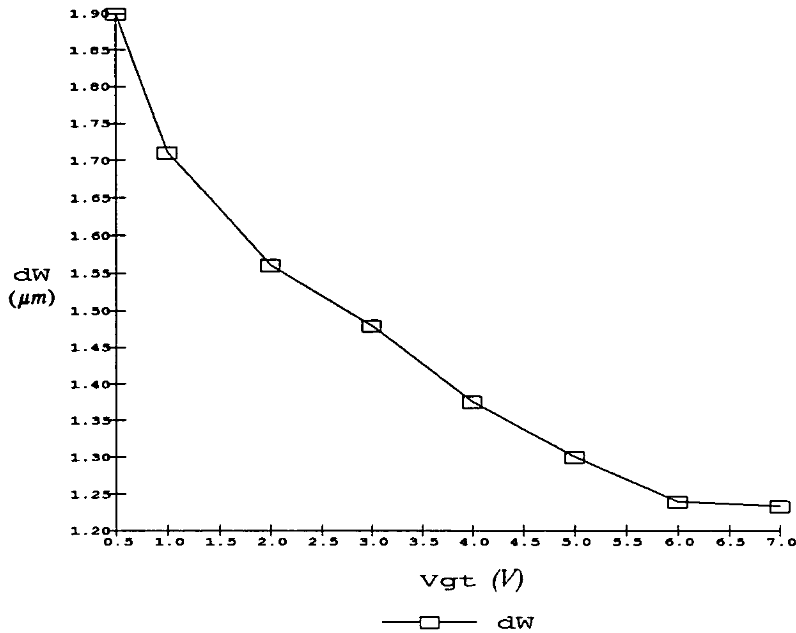


Figure 6.8. ΔW vs V_{gt} for Wafer 5.

6.5.1. Channel Normalisation: Error In Series Resistance

For all the measurements made the average value of $R_x \times W_D$ was 2250 $\Omega\mu m$. Figure 6.9 shows a graph of the V_{gt} against the product $R_m \times W_D$ and figure 6.10 an expanded view at the high values of V_{gt}. It can be seen that even at the highest values of V_{gt}, $R_m \times W_D$ is greater than 20,000 $\Omega\mu m$ and this is an order of magnitude greater than the product $R_x \times W_D$. At lower values of V_{gt} the difference is very much greater. Therefore, an error of 100% in the series resistance will only produce an error of 10% in the channel resistance, and even then only at the high values of V_{gt}. The impact of a change by a factor of two in the extracted series resistance can be seen in figures 6.11 and 6.12 where the channel normalisation technique has been used to extract ΔW. In 6.11 the product of the series resistance and the drawn width has been set to the extracted value of 2250 $\Omega\mu m$. In figure 6.12, R_x has been doubled to 4500 $\Omega\mu m$. No noticeable difference in the $\Delta W - V_{gt}$ graphs can be observed.

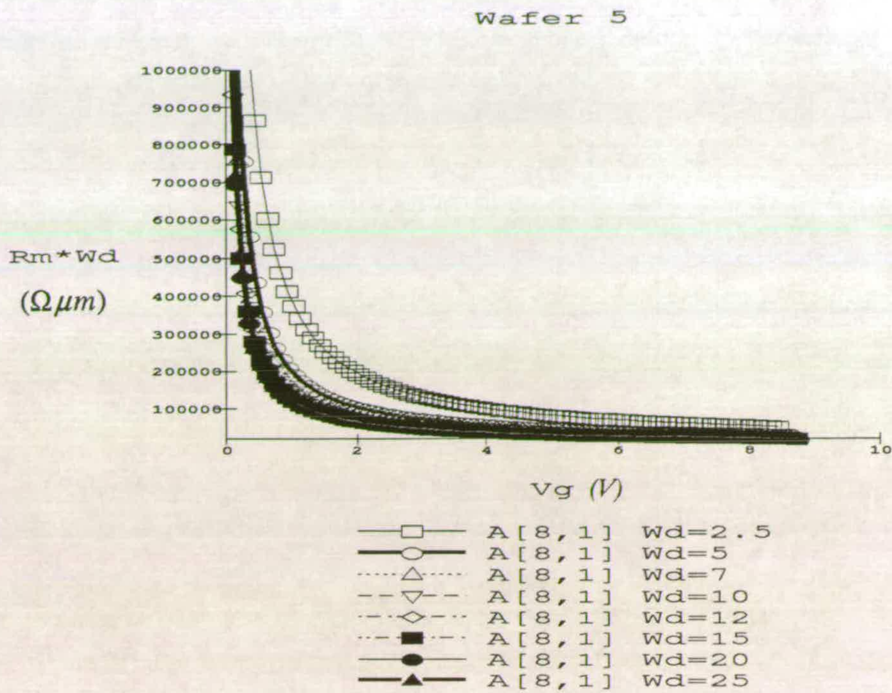


Figure 6.9 $R_m W_D$ product for wafer 5.

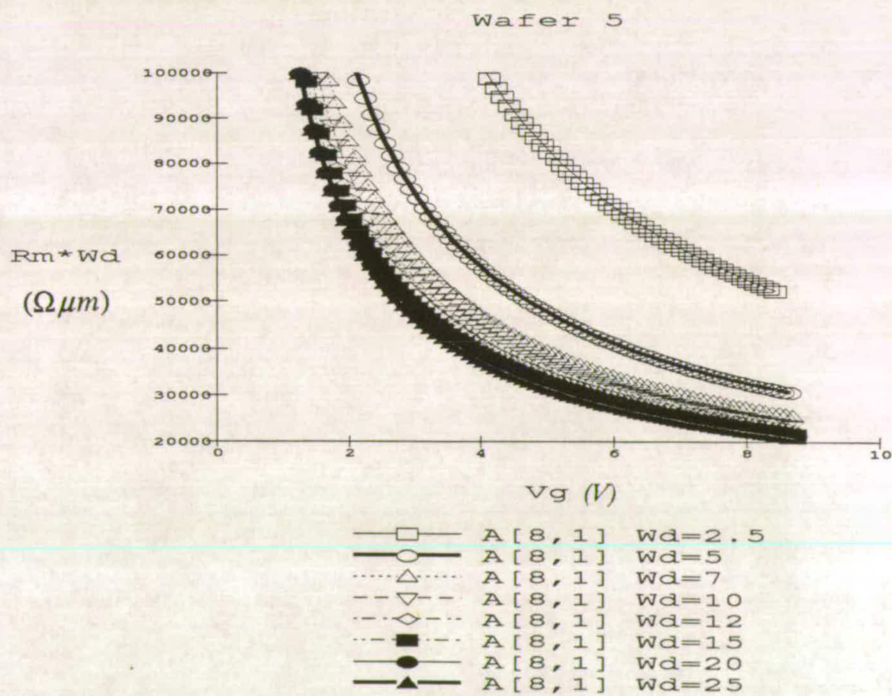


Figure 6.10. Expanded view of $R_m W_D$ product for wafer 5.

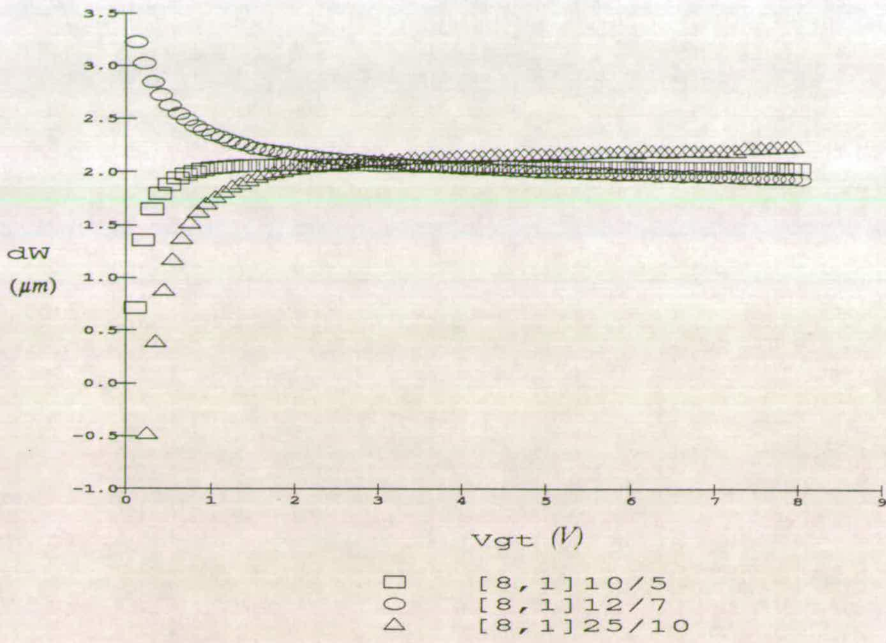


Figure 6.11. Extracted ΔW using for wafer 5 using nominal R_x.

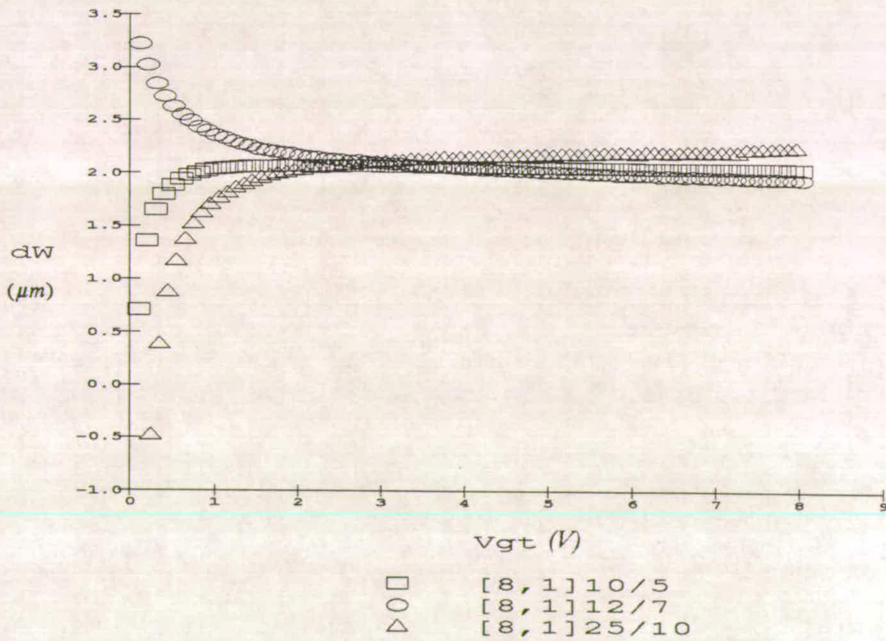


Figure 6.12. Extracted ΔW using for wafer 5 using $2 \times R_x$.

6.5.2. Channel Normalisation: Error In The Threshold Voltage

In all cases, the threshold voltage has been calculated by the linear extrapolation technique. The extraction techniques use V_t as the reference voltage. Therefore any variation in the threshold voltage will impact the resulting $\Delta W - V_{gt}$ relationship. This may arise in the definition used for the threshold voltage or through measurement error.

The $I_D - V_{gs}$ measurement itself was discretised. The V_{gs} voltage ranged from 0 to 10 volts in 0.1 volt steps. To assess the impact of small changes in the V_t, the extraction technique was performed with an offset of 0. 1 volts on the measured data. Figure 6.13 shows the results of a 0.1 volt shift on the V_t for the 10 μm and 7 μm devices used in figure 6.11. The 10&5 μm and 12&7 μm pairings show a large change in the two cases.

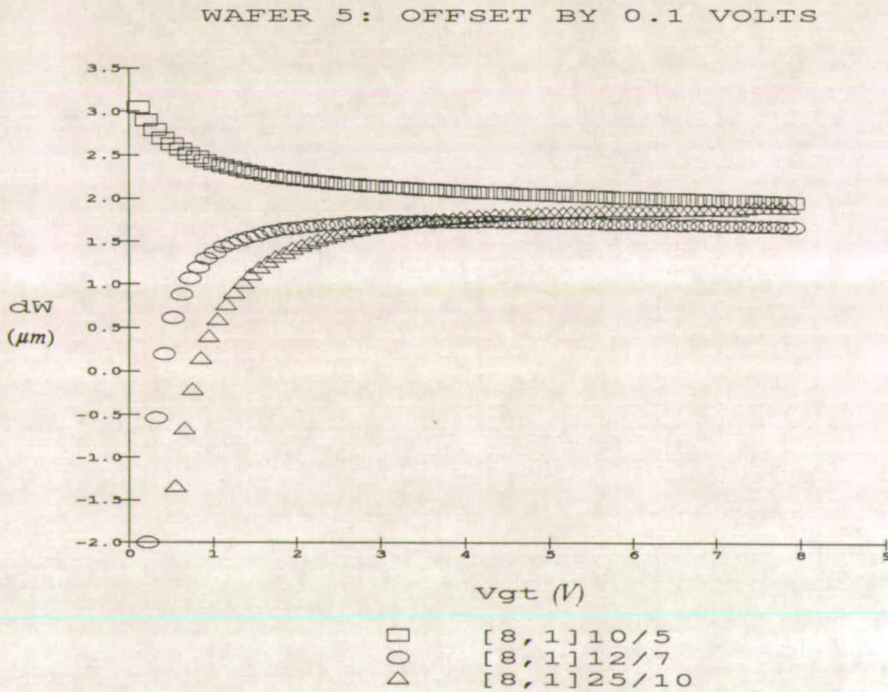


Figure 6.13. Extracted ΔW for wafer 5 using 0.1 volt V_t offset.

It can be seen that a large difference may occur in the output for the 'small' offset in V_t. The graph of the resultant G_t' as a function of V_{gt} is shown in figure 6.14.

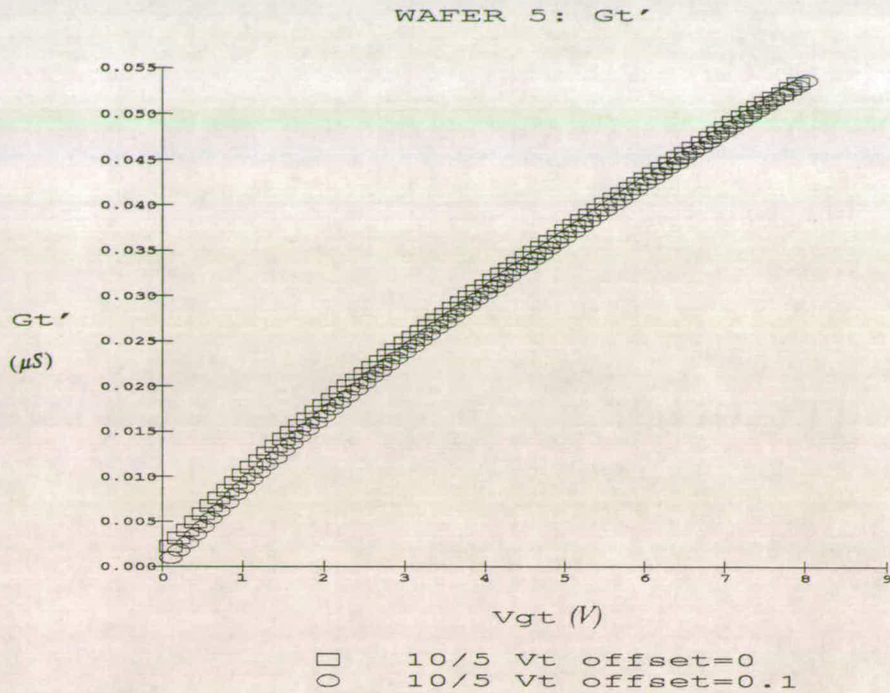


Figure 6.14. G_t' for wafer 5 using 0.1 volt V_t offset.

Here the value of V_t has been offset by 0.1 volts for the W_D =10 μm device and, though small, the difference in G_t' is critical to the final plot of ΔW *vs* V_{gt}.

Backwards linear extrapolation was used to define V_t. Since the measurement itself discretises the V_{gs} values, a linear extrapolation is unlikely to produce an intercept on the V_{gs} axis at a measurement point. Therefore in producing the $V_{gt} = V_{gs} - V_t$ values, errors of ±0.1volts are possible. In an effort to reduce this sensitivity, a linear interpolation between the resultant G_t and V_{gt} data points was used. This produced a common set of V_{gt} values and minimised the errors when subtracting the two sets of G_t values. Figure 6.15 shows the extracted G_t' values for 3 sets of devices within one chip, after the linear interpolation has been performed. The resultant $\Delta W - V_{gt}$ plots are shown in figure 6.16

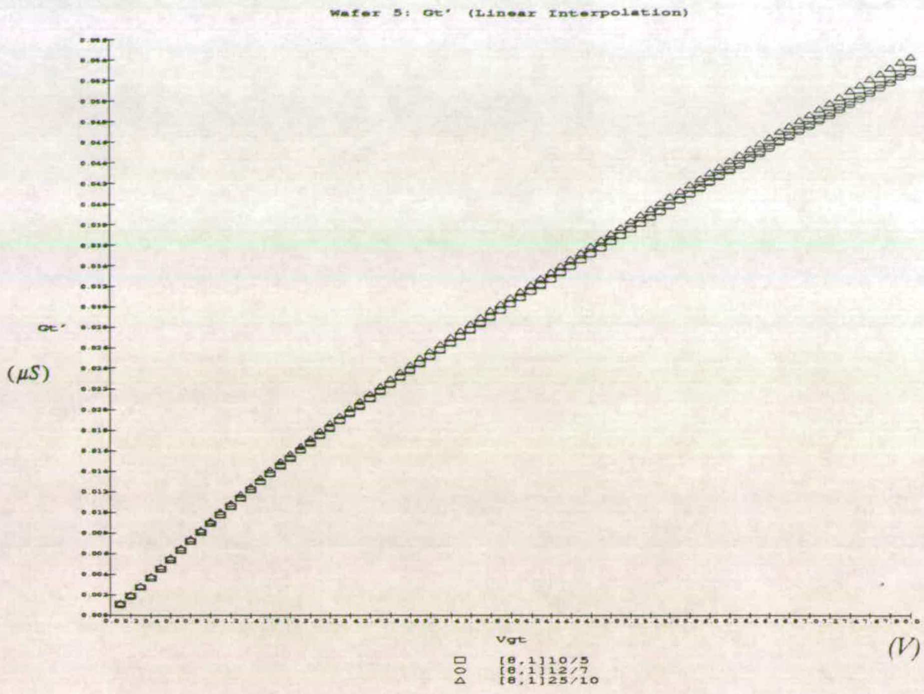


Figure 6.15. G_t' for wafer 5 using linear interpolation.

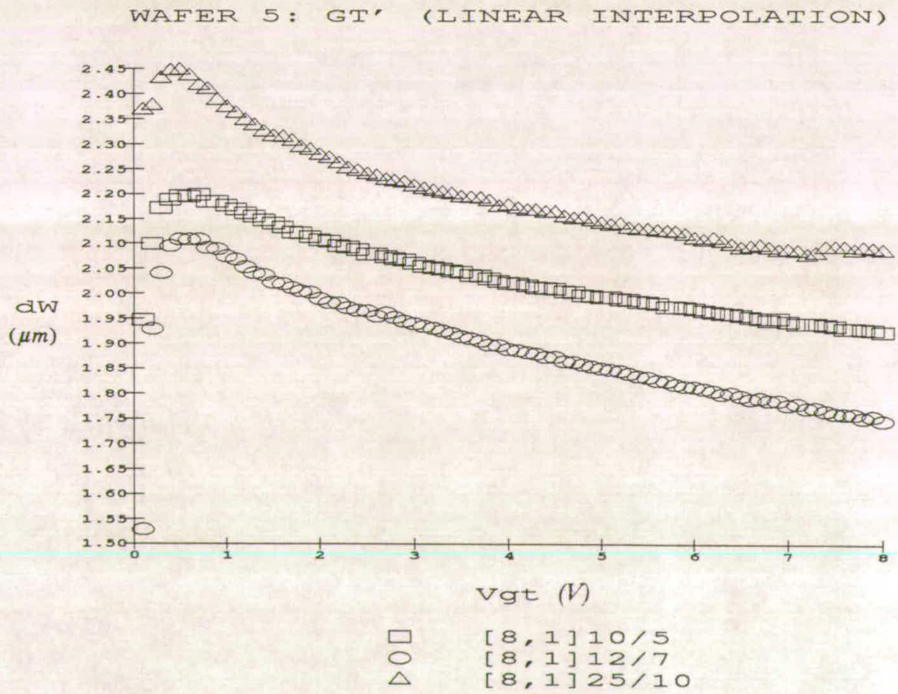


Figure 6.16. ΔW for wafer 5 using linear interpolation for G_t'.

The large variation in the ΔW values suggested errors still in the measurement or extraction process.

The equation used to calculate W_{eff} as a function of V_{gt} is

$$W_{eff} = \frac{G_{t2}}{G_t'}$$

where

$$G_t' = \frac{G_{t2} - G_{t1}}{W_{D2} - W_{D1}}$$

and

$$\Delta W = W_D - W_{eff}$$

Assuming G_t to be fixed, and allowing an error of $\pm\Delta G_t$ on G_t' enables the error in ΔW, ε_w, to be written as

$$\varepsilon_w = \frac{G_t}{G_t' \pm \Delta G_t'} \tag{6.8}$$

which reduces to

$$\varepsilon_w = \frac{2G_t \Delta G_t'}{G_t'^2 - \Delta G_t'^2} \tag{6.9}$$

The observed error in G_t' then scales with G_t and is a function of V_{gt}. The resultant plot is shown in figure 6.17

Using these values, the error in ΔW given by (6.9) has been calculated and is shown in figure 6.18. It can be seen that despite small variations in observed values of G_t' large errors can result in the value of ΔW. The greatest errors appear at low V_{gt}, which is reflected in the high and low extremes of ΔW seen in figure 6.16.

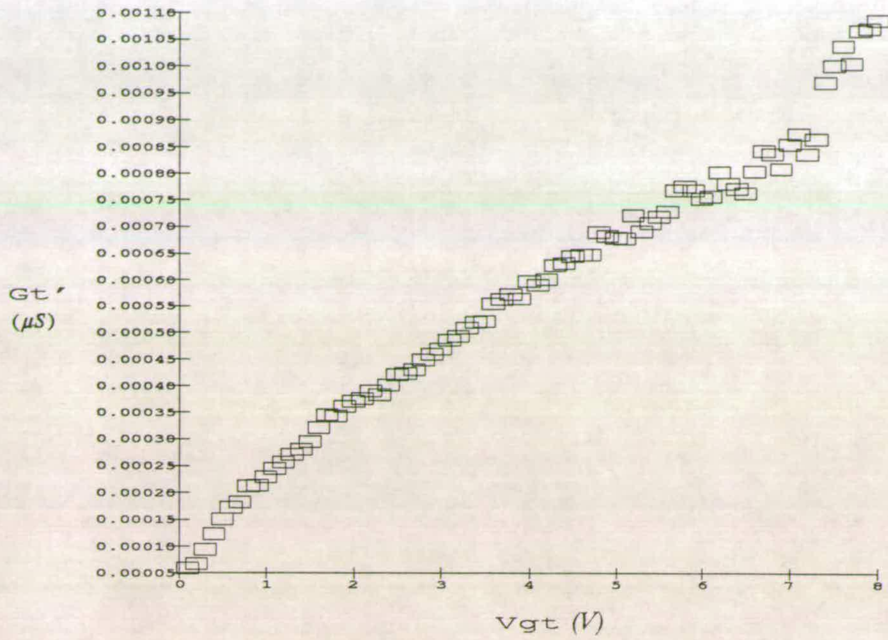


Figure 6.17. Error in G_t'

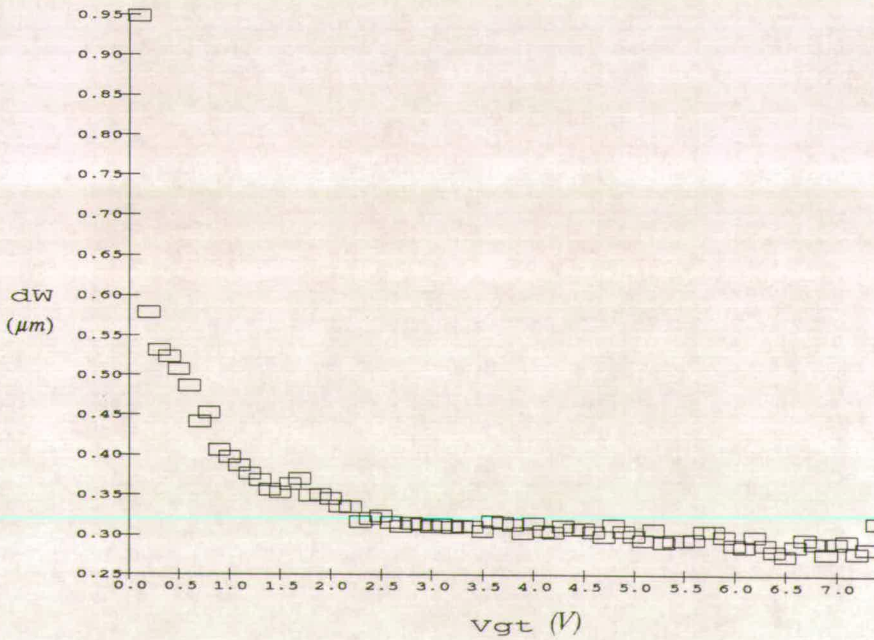


Figure 6.18. Error in ΔW calculated from ε_w values.

6.5.3. G_t' Extrapolation: Errors In Series Resistance

As with the previous extraction technique, errors in the series resistance measurement have little impact on the final $V_{gt} - W_D$ relationship for the same reason that R_x is only around 10% of R_m. Figure 6.19 shows the resultant $\Delta W - V_{gt}$ relationship using the average $R_x \times W_D$ and figure 6.20 the resultant using twice the value of $R_x \times W_D$.

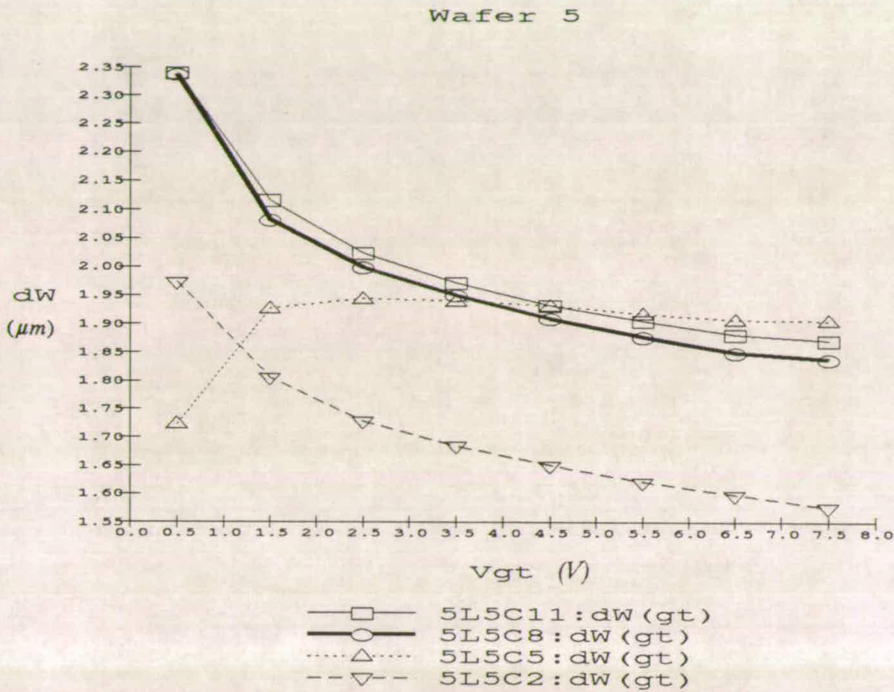


Figure 6.19. ΔW using mean $\{R_x \times W_D\}$.

6.5.4. G_t' Extrapolation: Errors In Threshold Voltage

The linear least squares fit to the measured data showed a consistently high correlation coefficient. Using the raw data of wafer 5 in table 6.3, the ΔW value for each V_{gt} and the corresponding correlation coefficient, (CC) can be observed.

Figure 6.21 shows the resultant $\Delta W - V_{gt}$ graph. Again however, this technique is susceptible to variations in the threshold voltage, since the V_t value defines the reference for the subsequent V_{gt} values. Figure 6.22 shows the effect of a 0.2 volt

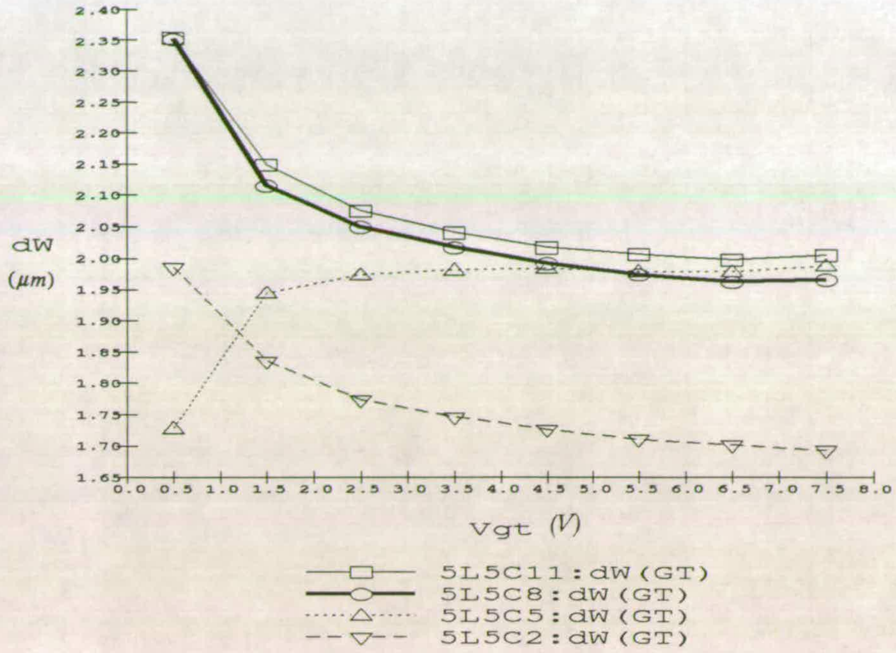


Figure 6.20. ΔW using mean $\{ 2 \times R_x \times W_D \}$.

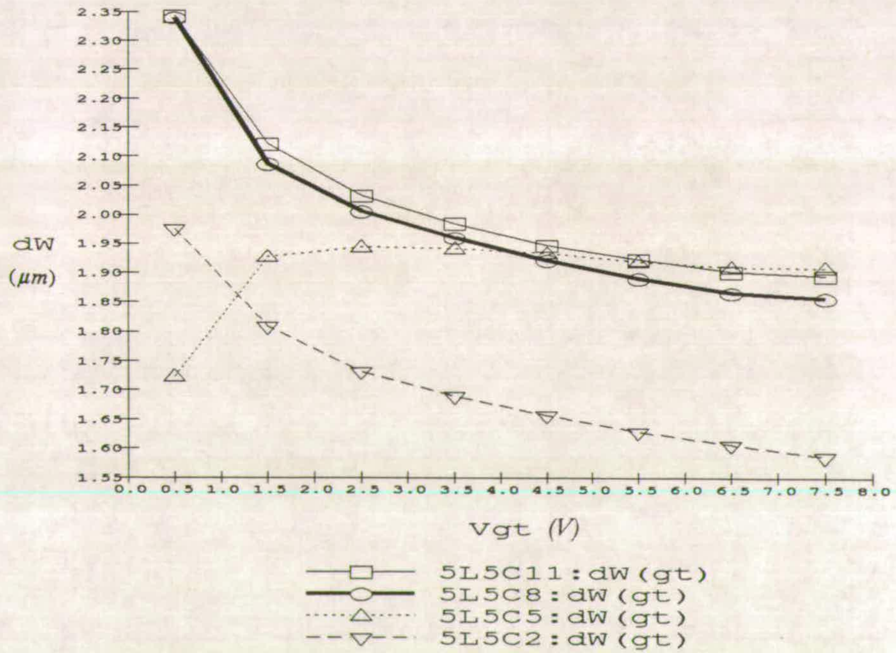


Figure 6.21 ΔW for wafer 5 using G_t'.

V_{gt}	$\Delta W(11)$	$CC(11)$	$\Delta W(8)$	$CC(8)$	$\Delta W(5)$	$CC(5)$	$\Delta W(2)$	$CC(2)$
0.5	2.342	0.99864	2.339	0.99332	1.724	0.99264	1.975	0.99884
1.5	2.121	0.99916	2.086	0.99924	1.927	0.99855	1.808	0.99980
2.5	2.032	0.99916	2.005	0.99962	1.944	0.99902	1.731	0.99986
3.5	1.984	0.99907	1.959	0.99969	1.940	0.99914	1.689	0.99986
4.5	1.946	0.99900	1.920	0.99972	1.933	0.99918	1.656	0.99985
5.5	1.922	0.99888	1.889	0.99972	1.918	0.99916	1.628	0.99984
6.5	1.900	0.99877	1.865	0.99971	1.908	0.99914	1.606	0.99982
7.5	1.893	0.99858	1.855	0.99966	1.907	0.99912	1.585	0.99980

Table 6.3. ΔW and CC for wafer 5.

shift in the threshold voltage on the final $\Delta W - V_{gt}$ graphs.

Large errors in ΔW are possible at low voltages, where the linear $I_D - V_{gt}$ relationship approximation is poor. As with the G_t' normalisation technique, errors can exist arising from the discretisation of the V_{gt} data. To reduce this a linear extrapolation was once more performed. The effect of this is seen in figure 6.23 where the result is to smooth out the discontinuities at low V_{gt} voltages. In the low V_{gt} regime small changes in V_{gt} have a large effect in the channel conductance [5] and measured data in this regime is therefore highly prone to error.

6.6. Measurement Analysis

The measurement of these wafers showed inconsistency in the results. Reasons for this were identified and have been discussed. However, even after accommodating these possible sources of error, unexplained variation in results are

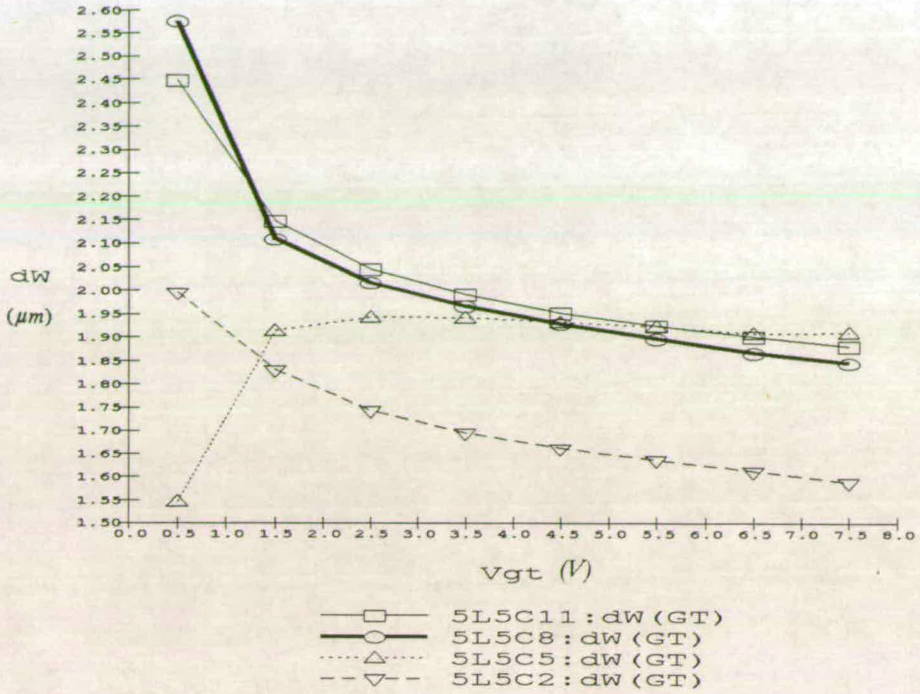


Figure 6.22 ΔW for wafer 5 using G_t' with V_t offset.

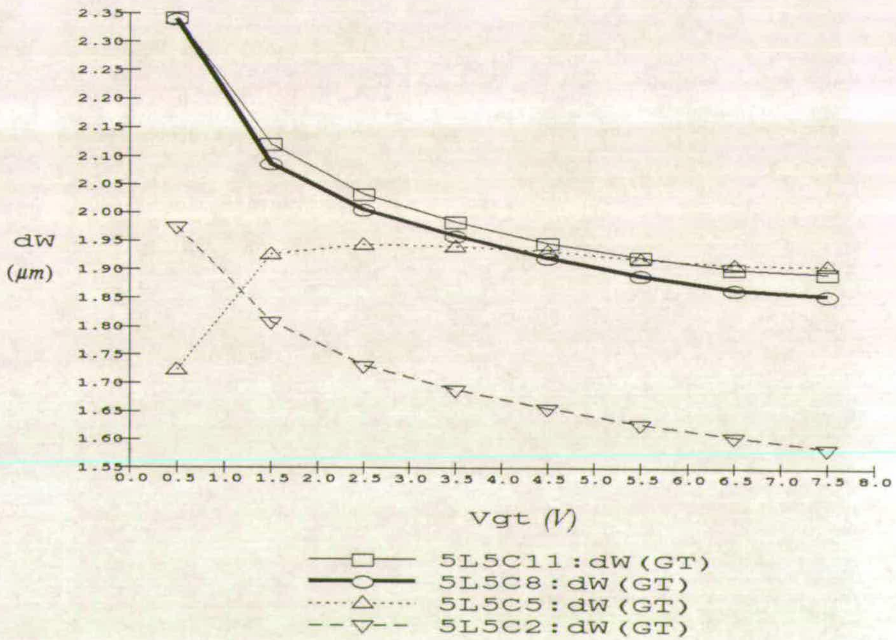


Figure 6.23. ΔW for wafer 5 using G_t' method and linear interpolation.

still observed. To examine this in detail, one wafer was measured 4 times. Four sections on the wafer corresponding to different T_{pad} thicknesses were selected to allow a family of curves to be produced. The purpose of this was to compare the width reduction derived from theoretical considerations with that predicted by simulation and SEM cross-section.

Nine different drawn widths were measured. The first measurement run tested all four sections of the wafer, (measurement A). The measurement was repeated on the same devices, (measurement B). Measurement C used the identical adjacent device as shown previously in figure 5.1. The fourth measurement pass re-tested these devices, (measurement D). Linear interpolation was then performed on the measured data. Figure 6.24 shows the graph of the measured threshold voltage and the maximum transconductance of each device.

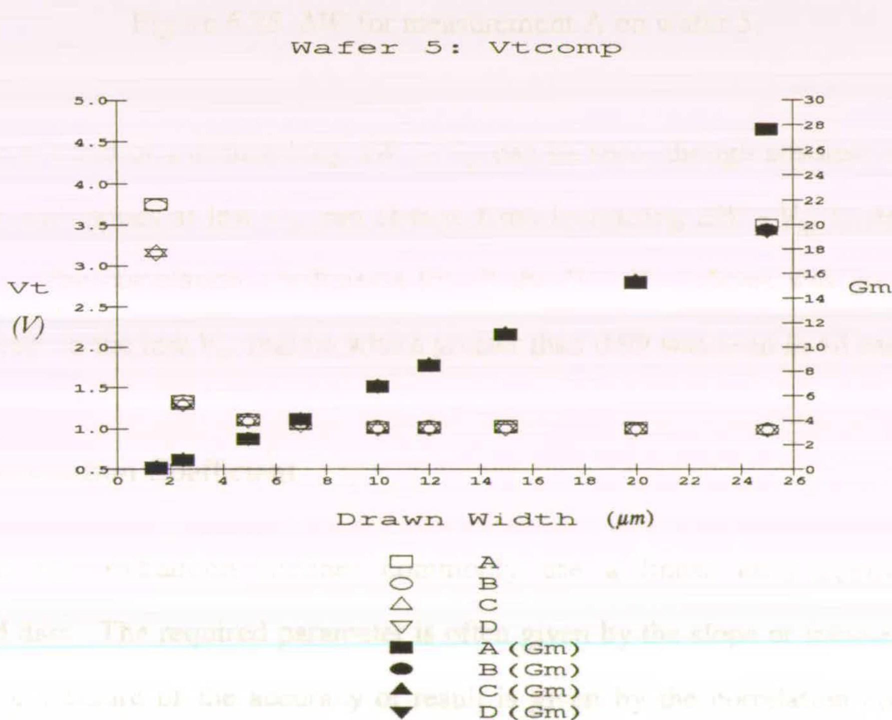


Figure 6.24. V_t and G_m for wafer 5.

The threshold voltages show a consistency except for the 1.5 μm device. The values of G_m also show consistency in measurement. Resultant $\Delta W - V_{gt}$ plots are shown in

figures 6.25, 6.26, 6.27 and 6.28.

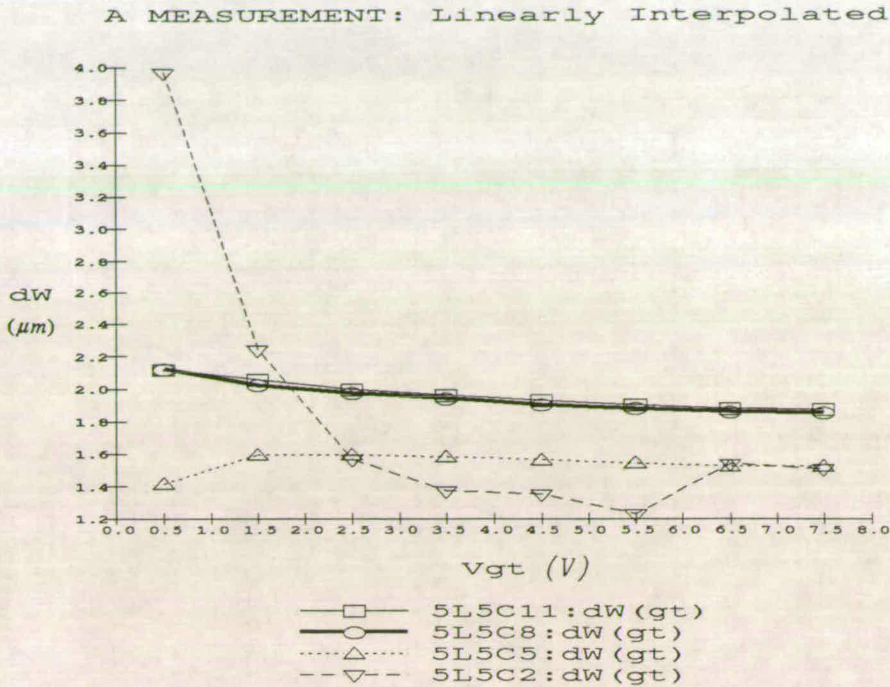


Figure 6.25. ΔW for measurement A on wafer 5.

The general trend of a diminishing ΔW *vs* V_{gt} can be seen, though absolute values do not agree and values at low V_{gt} can change from increasing ΔW - V_{gt} to decreasing ΔW - V_{gt}. The correlation coefficients for all the G_t - W_D curves was greater than 0.999 except in the low V_{gt} regime where greater than 0.99 was seen in all cases.

6.6.1. Correlation Coefficient

Parameter extraction routines commonly use a linear least squares fit to measured data. The required parameter is often given by the slope or intercept on the axis and a measure of the accuracy of result is given by the correlation coefficient. The correlation coefficient parameter is an assessment of the degree of linear relationship between two variables [6] . A figure of 1 denotes perfect correlation and 0 no correlation. A value of -1 means that the variables are negatively correlated. The conclusions which can be drawn about the degree of linear relationship between

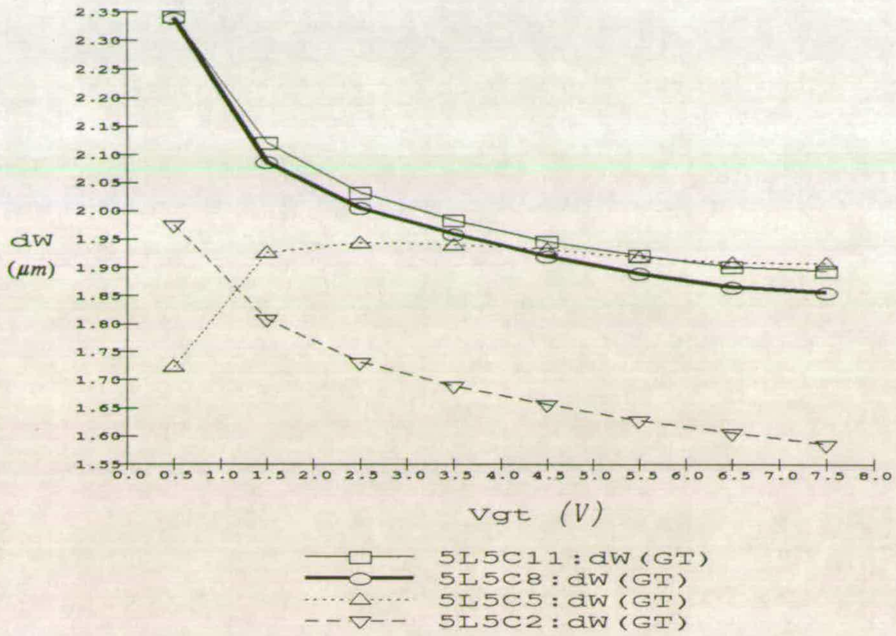


Figure 6.26. ΔW for measurement B on wafer 5.

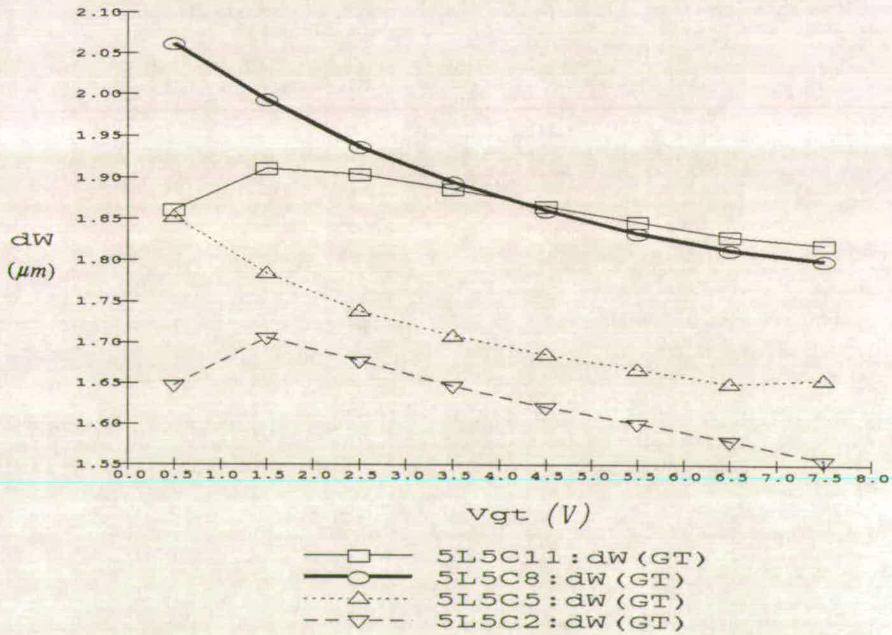


Figure 6.27. ΔW for measurement C on wafer 5.

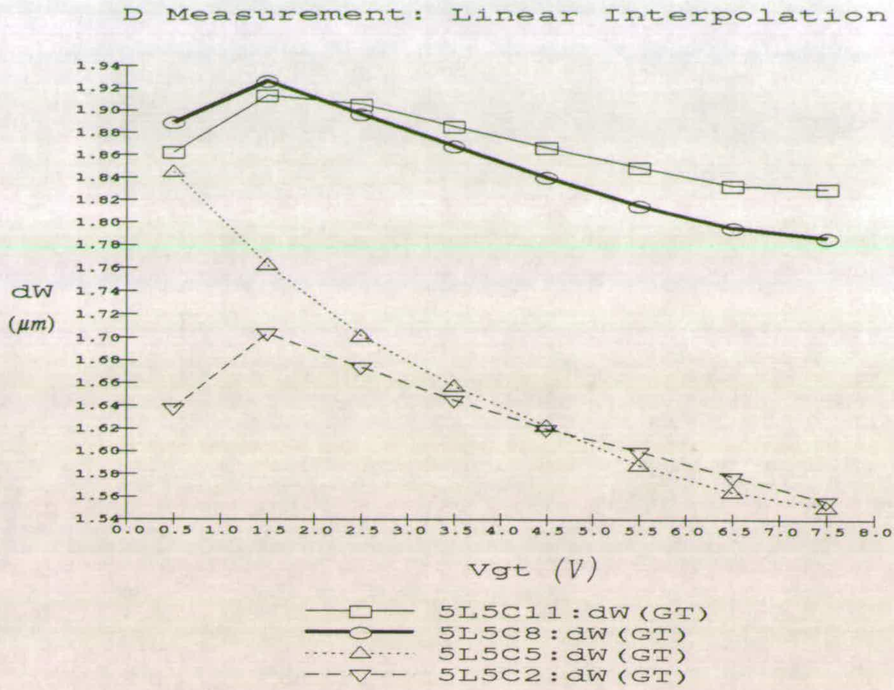


Figure 6.28. ΔW for measurement D on wafer 5.

two variables depend on the number of pairs of observations that have been made. It is obviously easier for 10 points to lie close to a line than 100. In the case of ΔW extraction, a dual linear fit is performed on 4 points and a correlation coefficient is quoted as 0.999 [7] or on 5 points and a correlation coefficient of 0.99 [4] . The data presented in figures 6.25 to 6.28 quote a correlation coefficient of better than 0.999 for $V_{gt} > 0.5$ and 0.99 for $V_{gt} = 0.5$ volts for 8 data points. (Generally 9 data points were used, except where $\Delta W > W_D$). However,

> "the interpretation of the correlation coefficient as a measure of the strength of the linear relationship between two variables does not apply if the values of x are selected as desired because the value of the correlation coefficient will usually depend heavily on the choice of x -values"[8] .

Thus in these processes, care must be taken when quoting a correlation coefficient as an assessment of accuracy.

6.7. Data Set Selection

Other than the Channel Normalisation method, all of the above extraction techniques use a linear least squares fit to a set of data points. Extrapolating beyond the data set reduces the confidence in the value of the extracted parameter. The ΔW extraction techniques all use extrapolation beyond the data set, since negative values of W_D cannot be used.

The choice of data points can critically influence the extrapolated value as can be seen in figures 6.29 to 6.31. Here ΔW has been extracted using the extrapolated $G_m - W_D$ graphs where no allowance has been made for the series resistance value. Figure 6.29 has been derived using the data set W_D = 1.5, 2.5, 5.0, 7.0 μm. and shows an extracted ΔW of 1.4 μm, which is close to the minimum drawn dimension of 1.5 μm. Figure 6.30 shows a $\Delta W = 1.6 \mu$m when the data set is 2.5, 5.0, 7.0 and 10.0 μm. Using a least squares fit to the combined data set of 6.29 and 6.30. viz 1.5, 2.5, 5.0, 7.0 and 10.0 μm produces the figure 6.31 where ΔW =1.49 μm. Therefore it can be seen that data selection in the simplest of cases can influence the extracted parameter value. In all cases the correlation coefficient is greater than 0.999.

Care must be taken when adding more data points. Using more data points away from the origin weights the fit further from the point of interest. This is seen in figure 6.32 where the W_D values are 12, 15, 20 and 25 μm. Whilst an excellent fit can be obtained small variations in the fitted line are magnified when extrapolated to the origin. Tables 6.4, 6.5, 6.6, 6.7, 6.8, and 6.9 show the extracted ΔW for eight chips across one wafer using different data selection sets for the extraction technique. The correlation coefficient is denoted as CC.

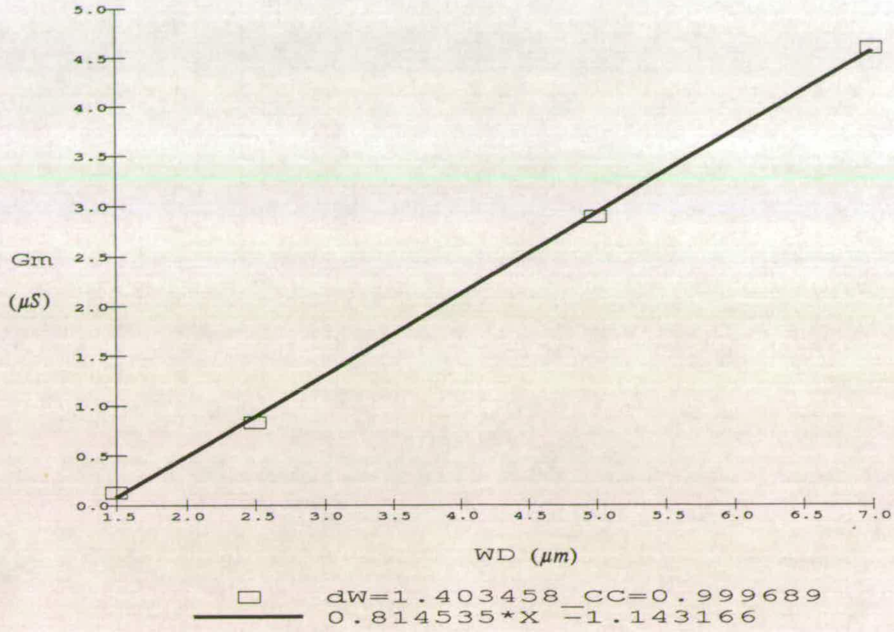


Figure 6.29. ΔW using W_D = 1.5, 2.5, 5 and 7 μm.

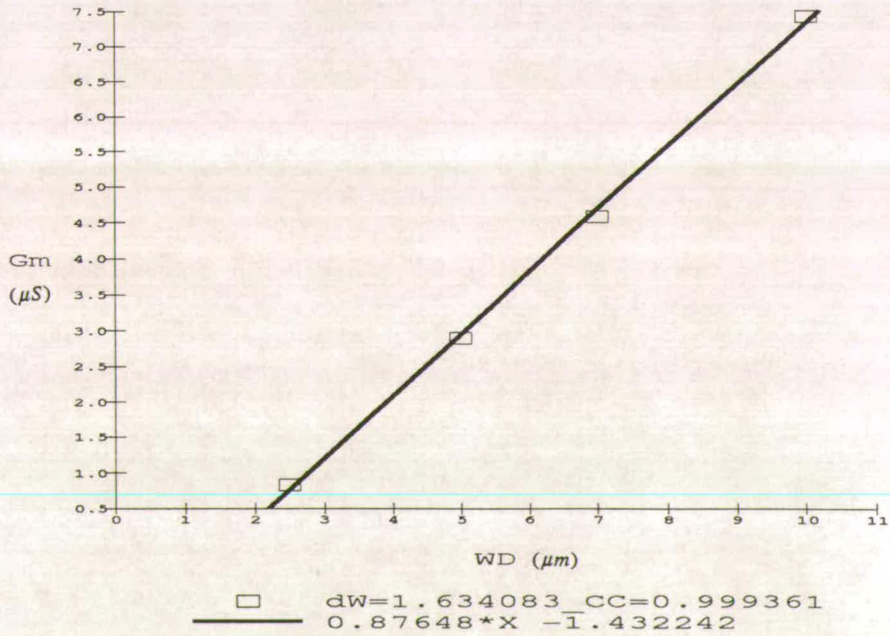


Figure 6.30. ΔW using W_D = 2.5, 5, 7 and 10 μm.

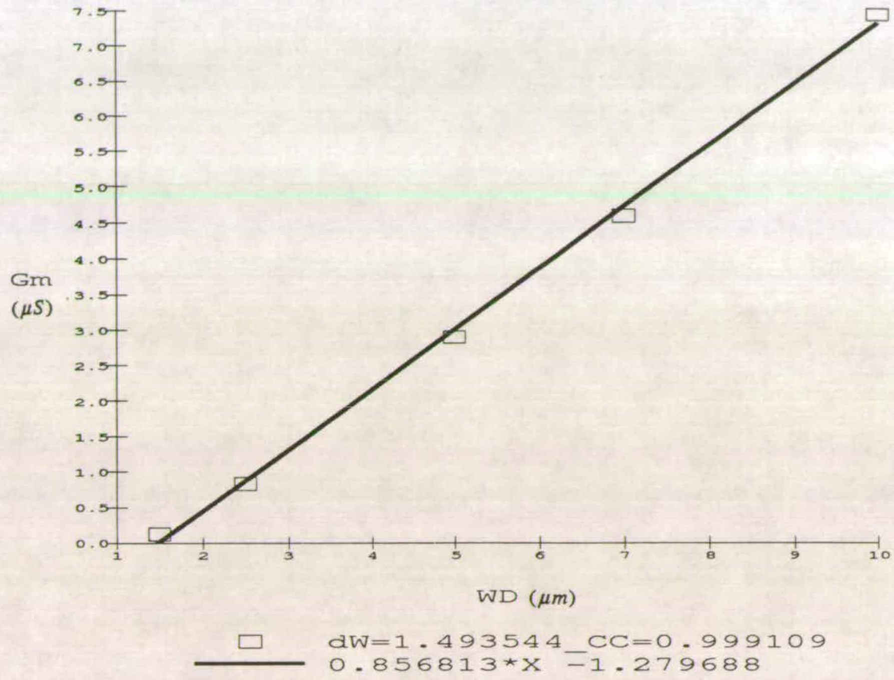


Figure 6.31. ΔW using W_D = 1.5, 2.5, 5, 7 and 10 μm.

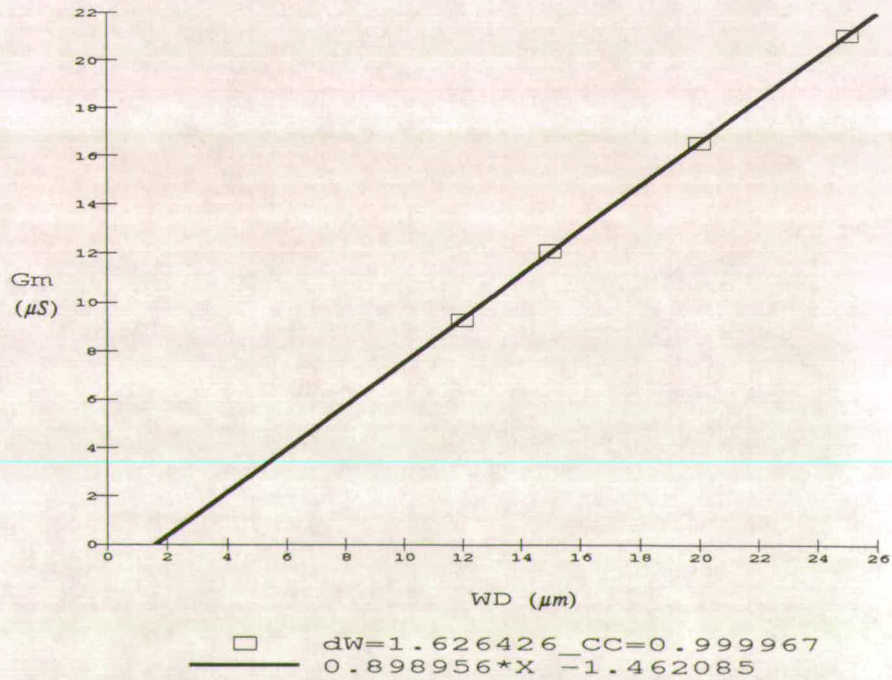


Figure 6.32. ΔW using W_D = 12, 15, 20 and 25 μm.

Data set	X-coord	Y-coord	ΔW	CC
1234	4	1	1.4395	0.996323
1234	4	2	1.5805	0.996731
1234	4	3	1.5476	0.996016
1234	4	4	1.5422	0.996765
1234	5	1	1.4287	0.999586
1234	5	2	1.4833	0.997590
1234	5	3	1.4035	0.999689
1234	5	4	1.4295	0.998687

Table 6.4. ΔW and CC for wafer 5 with data set [1234].

Data set	X-coord	Y-coord	ΔW	CC
2345	4	1	1.6859	0.998913
2345	4	2	1.8989	0.998466
2345	4	3	1.6237	0.997823
2345	4	4	1.8154	0.998864
2345	5	1	1.6794	0.999217
2345	5	2	1.7568	0.998976
2345	5	3	1.6341	0.999361
2345	5	4	1.6758	0.999258

Table 6.5. ΔW and CC for wafer 5 with data set [2345].

Data set	X-coord	Y-coord	ΔW	CC
3456	4	1	1.9760	0.999524
3456	4	2	2.2578	0.999944
3456	4	3	1.9058	0.997727
3456	4	4	2.1679	0.999788
3456	5	1	1.9405	0.999809
3456	5	2	2.1077	0.999895
3456	5	3	1.9074	0.999742
3456	5	4	2.0365	0.999882

Table 6.6. ΔW and CC for wafer 5 with data set [3456].

Data set	X-coord	Y-coord	ΔW	CC
4567	4	1	1.9948	0.999468
4567	4	2	2.2988	0.999868
4567	4	3	2.1396	0.99667
4567	4	4	2.3463	0.999348
4567	5	1	1.9757	0.999853
4567	5	2	2.1829	0.999809
4567	5	3	2.0921	0.999992
4567	5	4	2.1021	0.99992

Table 6.7. ΔW and CC for wafer 5 with data set [4567].

Data set	X-coord	Y-coord	ΔW	CC
5678	4	1	2.24576	0.999933
5678	4	2	2.3648	0.999952
5678	4	3	2.9395	0.999874
5678	4	4	2.6337	0.999921
5678	5	1	1.4809	0.999913
5678	5	2	2.2428	0.999926
5678	5	3	1.7627	0.999886
5678	5	4	2.0779	0.999942

Table 6.8. ΔW and CC for wafer 5 with data set [5678].

Data set	X-coord	Y-coord	ΔW	CC
6789	4	1	1.2455	0.998646
6789	4	2	1.8678	0.999537
6789	4	3	0.9752	0.993834
6789	4	4	2.0971	0.999513
6789	5	1	1.6994	0.99979
6789	5	2	1.9082	0.999889
6789	5	3	1.6264	0.999967
6789	5	4	1.7205	0.999928

Table 6.9. ΔW and CC for wafer 5 with data set [6789].

Figure 6.33 shows the spread of ΔW according to the data set chosen.

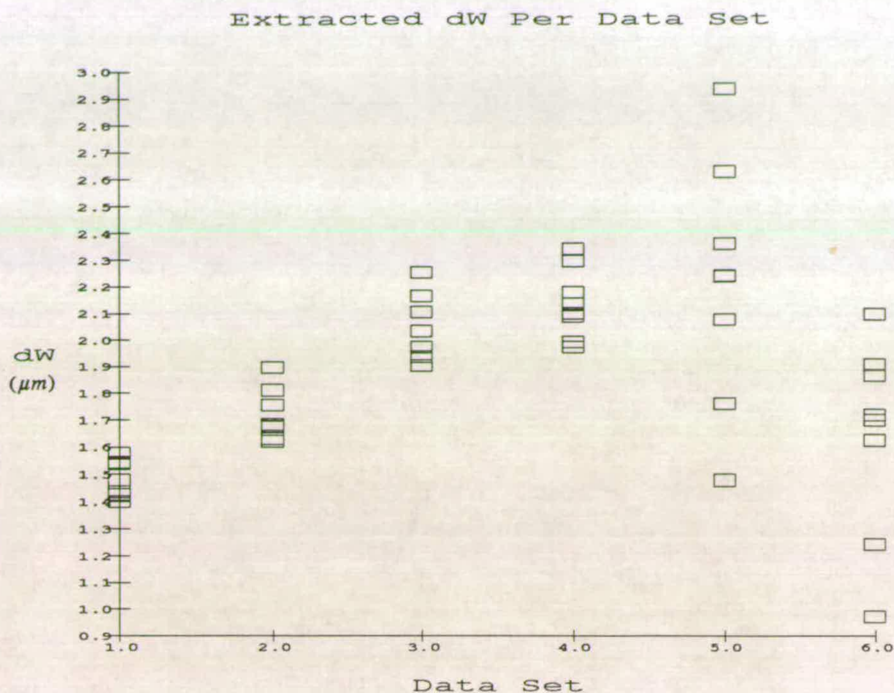


Figure 6.33. ΔW spread for wafer 5.

It can be seen that the spread increases as points further from the axis are chosen. However, as tables 6.4, 6.5 and 6.6 show, the correlation coefficient tends to increase in an inverse relationship to the spread in the extracted ΔW values. According to figure 6.32, the optimum data set, as defined where the spread in ΔW values is a minimum, is given by the set chosen closest to the origin. The extracted ΔW value in this case is $\approx 1.50 \pm 0.1$ μm. However the data set used includes a $W_D = 1.5$ μm device and is therefore unjustifiable to use the minimum data set in this extraction.

6.7.1. Extrapolation Using Two-Stage Processes

The $G_m - W_D$ extraction technique is a one-stage extraction process. The methods of [7] and [4] are both two stage extraction processes whereby the resulting slopes and intercepts from the $G_t - W_D$ graphs are plotted against each other. However the same argument is applicable. This can be seen in the published data of

[7] . Figure 6.34 shows the published graph whereby an extracted width, ΔW, is given by the value of the slope and can be estimated here as $\Delta W = 1.0$ μm.

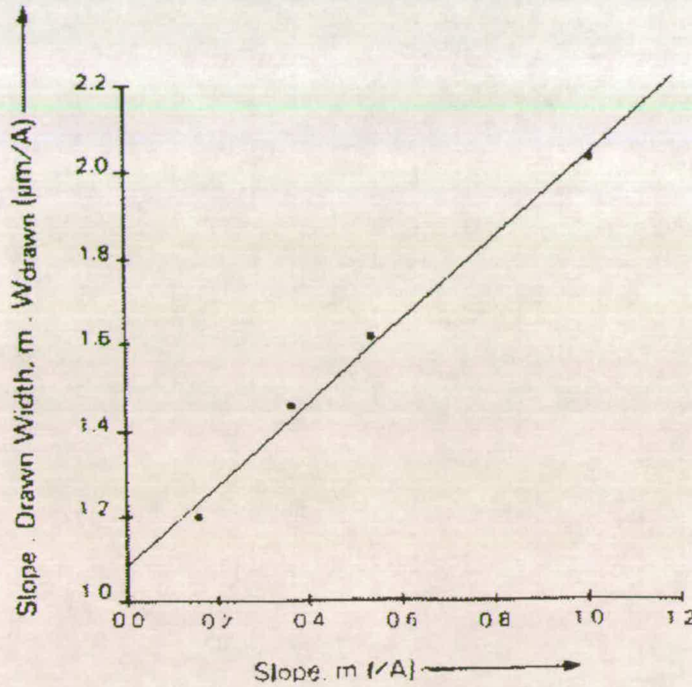


Figure 6.34. Data of [7]

Figure 6.35 shows the same graph, omitting in one instance the point near the origin and in the other the point furthest from the origin. In these cases the value for ΔW can be estimated as 1.14 and 0.92 μm, respectively. (In fact, choosing a worst case fit for these data points presented can result in $\Delta W = 0.7\mu$m.)

This result has been confirmed in the measured data presented here. Figure 6.36 has been obtained using the extraction technique of [3] where the the difference is seen in omitting the highest and lowest value of W_D in the data set. The spread in extracted values is again evident.

Allowing for edge currents and series resistance in the extraction process shows the same trend in the spread of results as shown in figure 6.37 where the variation in ΔW according to the data set and the site location is shown when $V_{gt} = 5$ volts. Again the spread in ΔW increases as the data set moves further from the origin. Figure 6.38

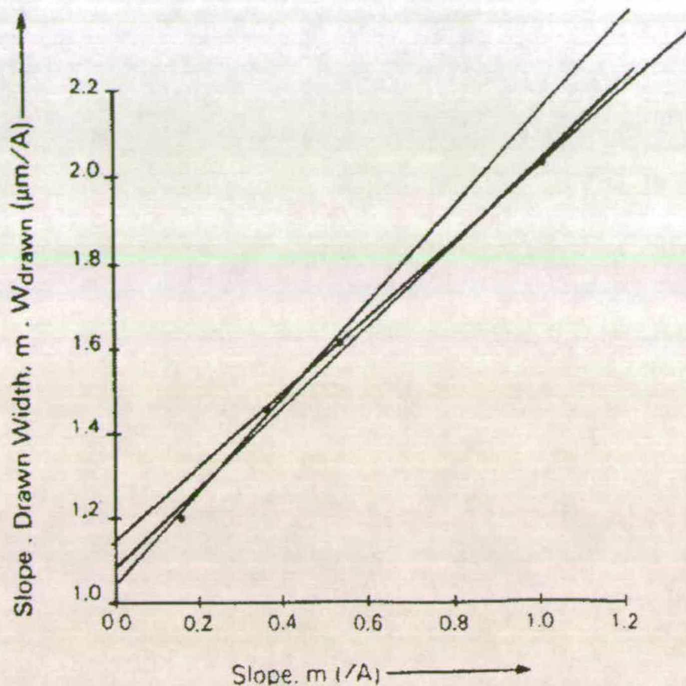


Figure 6.35. Best-worst fit using data of [7]

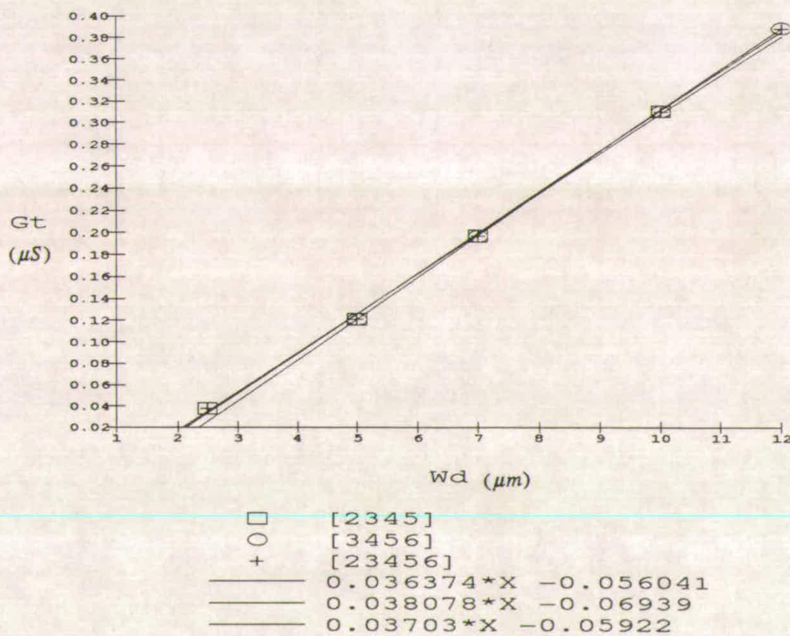


Figure 6.36. Spread in ΔW using technique of [3]

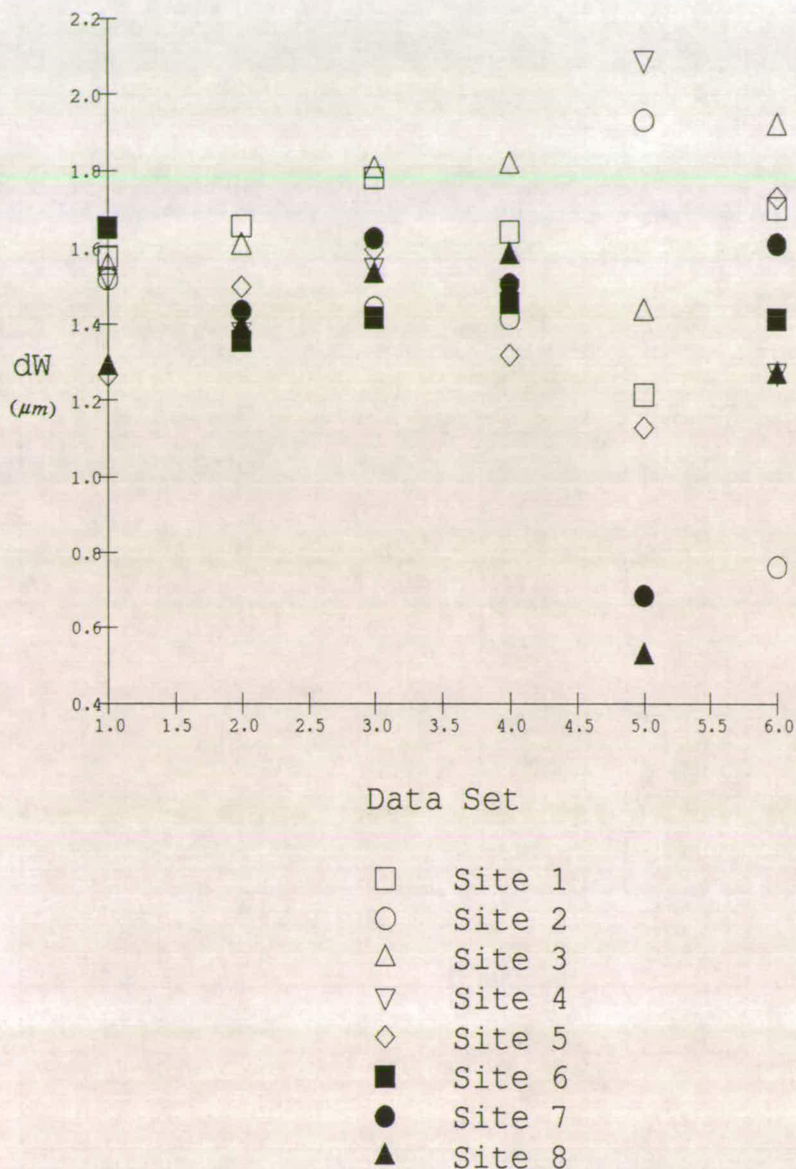


Figure 6.37. Spread in ΔW allowing for edge currents.

shows the spread in V_t. It can be seen that for $W_D = 1.5$ μm, $V_t \approx 6$ volts and consequently this data point has been excluded in calculating the $\Delta W - V_{gt}$ spread, since the maximum V_{gt} value is 10 volts.

To minimise variations in effective length, the same process was run using $W_D = 10$ μm for wafer 15. The spread in threshold voltages is shown in figure 6.39.

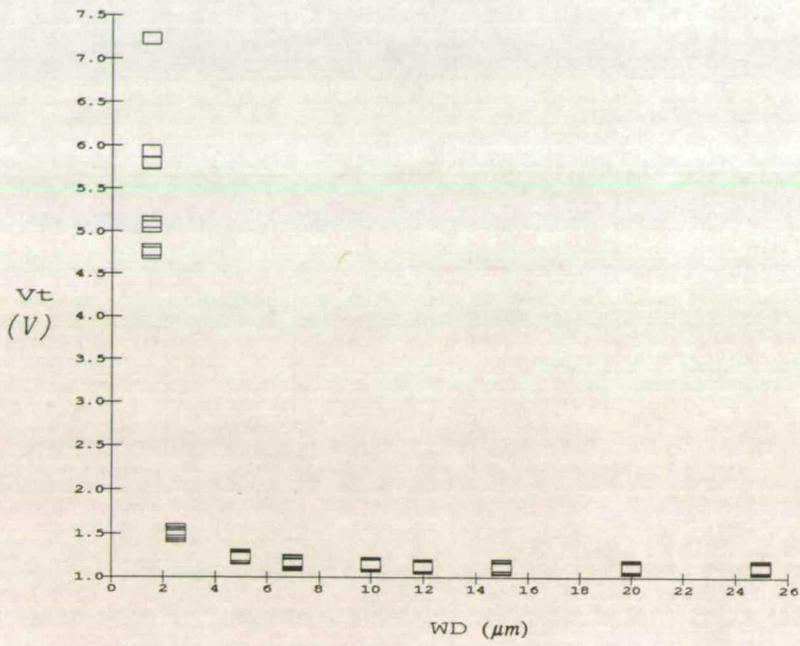


Figure 6.38. Spread in V_t.

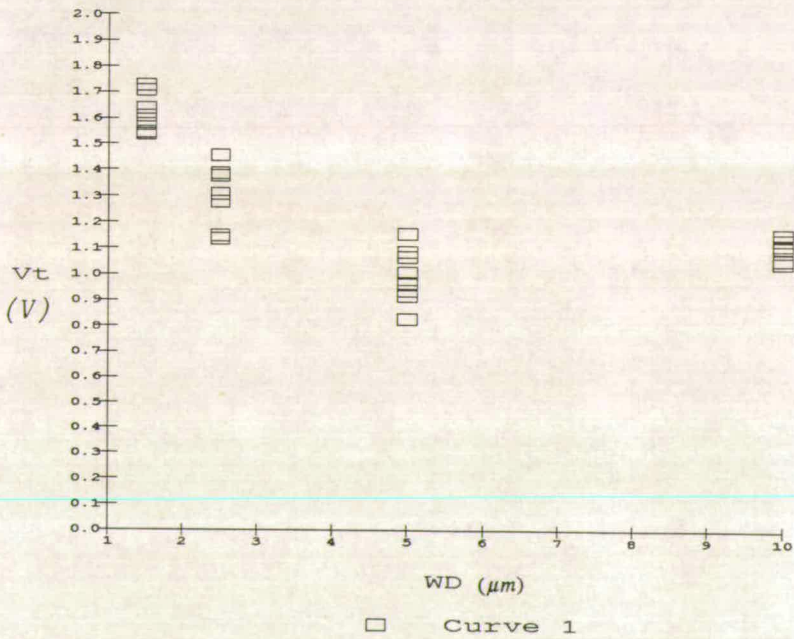


Figure 6.39. V_t spread using $L_D = 10\ \mu m$ devices.

Only 4 different drawn widths were available using this drawn length: 1.5, 2.5, 5.0 and 10.0 μm. The extracted ΔW values are shown in figure 6.40. Data set 1 uses all 4 drawn widths. Data set 2 omitted the 10 μm device and data set 3 omitted the 1.5 μm device.

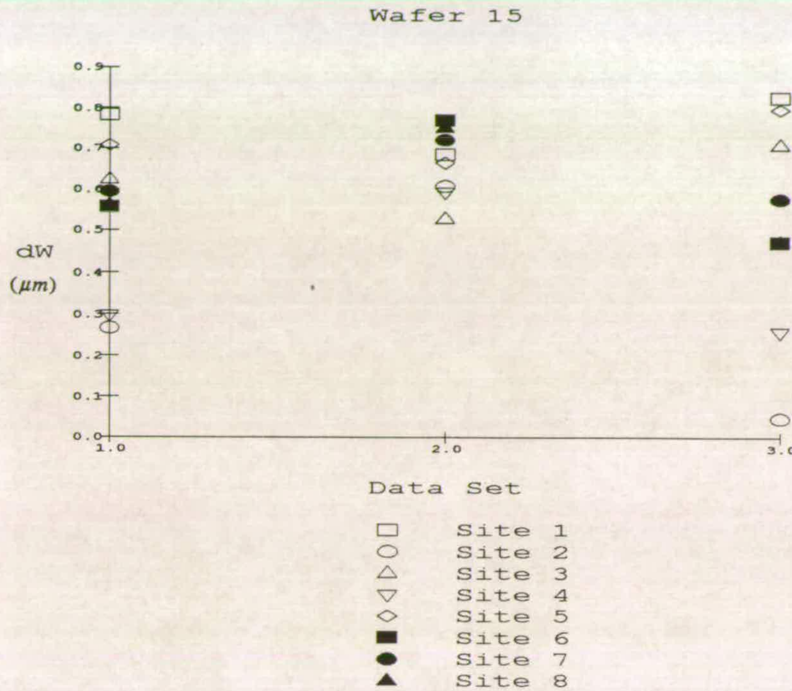


Figure 6.40. Spread in ΔW using $L_D = 10$ μm devices.

The sensitivity to the one 10 μm data point is clearly seen. Slight variations at this point can pivot the straight line to produce large variations in the extracted ΔW value and the effect is more pronounced the fewer the data points. Thus apparent 'rogue' values can be greatly influenced by slight variations in one carefully selected data point.

6.7.2. Data Set Optimisation

Whilst it is desirable to cluster the data set close to the axis, this can cause further problems since edge effects can dominate when W_D approaches ΔW. This can be seen in the $G_t - W_D$ graphs in figures 6.41 and in expanded view in fig 6.42.

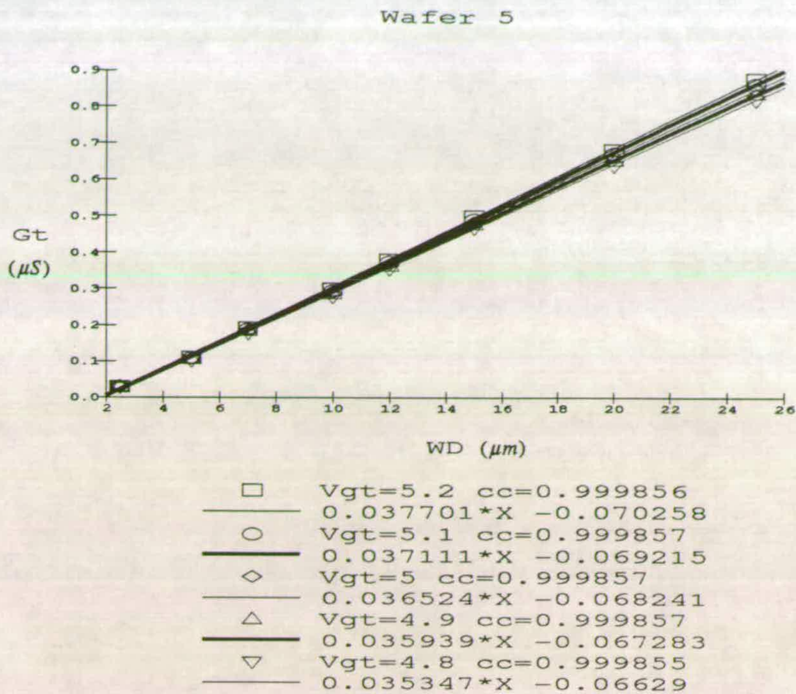


Figure 6.41. Extrapolation showing $\Delta W - G_t$ deviation.

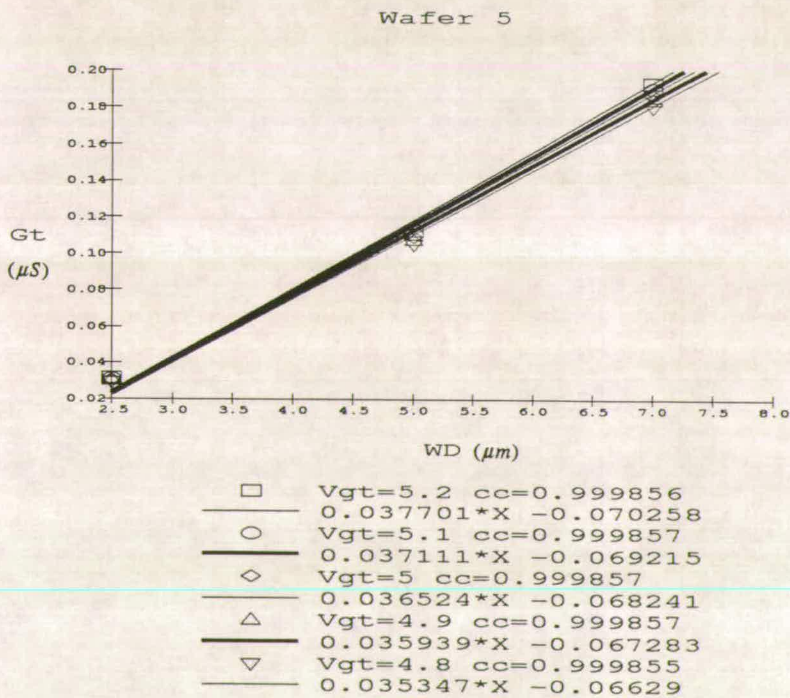


Figure 6.42. Expanded view of G_t extrapolation.

The deviation of G_t for the lowest value of W_D from the straight line tends to reduce the ΔW value. The outcome of this is that the extraction data set should be close to the axis but far enough away to prohibit edge effect domination. This, then, suggests that a dynamic range of data is necessary when comparing different processes to extract a given parameter. If $W_{\min}$ is the minimum value which can be used as a data point and W_s the set of data points then the optimum range for the data set is given by $W_{\min} + W_s$. If two processes give different $W_{\min}$ then the data set for the two will differ by the difference in $W_{\min}$ values. Figure 6.32 suggests the optimum data set is that closest to the axis where the spread in W_D values is a minimum, but here $\Delta W = 1.5\ \mu m \pm 0.1\ \mu m$. However this data set has included the $W_D = 1.5\ \mu m$ device itself.

6.8. Results

Since the data points were defined prior to the extraction process, the only control was in the selection of the data sets themselves. Figure 6.43 shows the plot for wafers 3, 5, 7 and 9 using 4 sets of 4 drawn widths: 2.5, 5, 7 and 10 μm and 5, 7, 10 and 12 μm. The inclusion of the smaller drawn width (the filled symbols) reduces the extracted ΔW in each case. The characteristic shape of the curve at higher V_{gt} is seen in all cases. Low values of V_{gt} can produce reverse trends whereby the extracted width may appear to go up or down as V_{gt} increases. Figure 6.44 shows a similar result for wafers 13, 15, 19 and 21. It should be noted however, that the data set chosen is different in each case, since the wafers have received different processing.

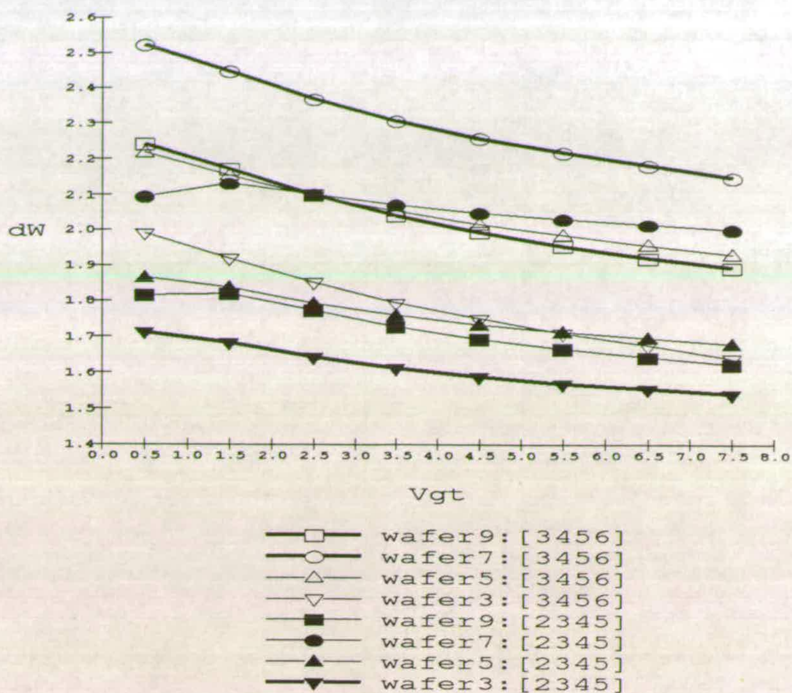


Figure 6.43. ΔW for wafers 3, 5, 7 and 9.

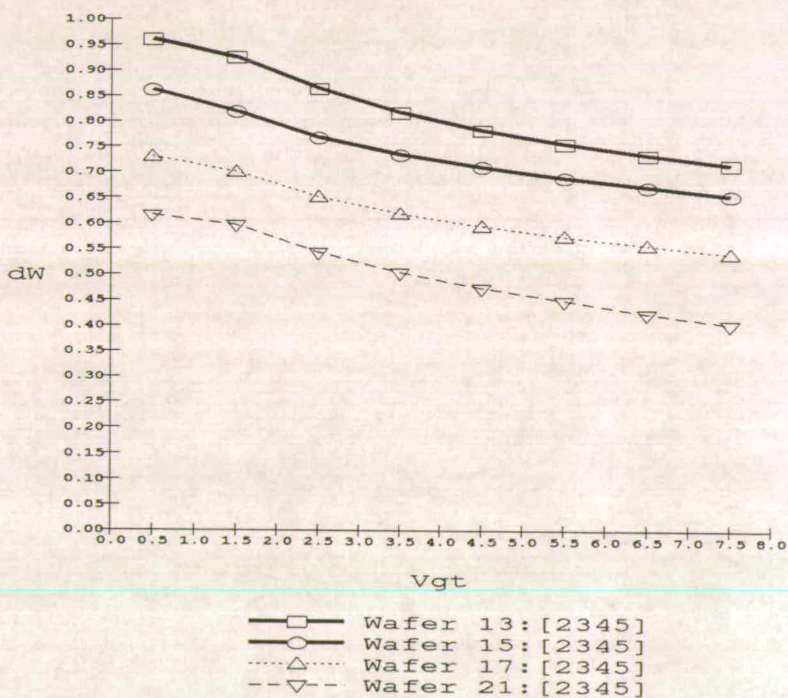


Figure 6.44. ΔW for wafers 13, 15, 19 and 21.

6.9. Summary

Different width extraction techniques have been used in extracting the W_{eff} values. Several points have arisen from the electrical measurements. These are summarised as follows.

(i) The correlation coefficient does not give an assessment of the error in the extraction process.

(ii) The $\Delta W - V_{gt}$ graph shows a diminishing dependence against increasing V_{gt} for each of the measurement techniques used.

(iii) Variations in position of the $\Delta W - V_{gt}$ line are possible according to the threshold voltage extraction technique.

(iv) The ΔW extraction technique is critically dependent on the choice of data points, with the optimum set being close to the extracted value.

References

1. P. I. Suciu and R. L. Johnston, "Experimental Derivation of the Source And Drain Resistance of MOS Transistors", *IEEE Trans. Elec. Dev.*, vol. ED-27, no. 9, pp. 1846-1848, Sep 1980.

2. Y-R. Ma and K. L. Wang, "A New Method To Electrically Determine Effective MOSFET Channel Width", *IEEE Trans. Elec. Dev.*, vol. ED-29, no. 12, pp. 1825-1827, Dec 1982.

3. G. J. Hu, C. Chang, and Y-T Chia, "Gate-Voltage-Dependent Effective Channel Length And Series Resistance of LDD MOSFETs", *IEEE Trans. Elec. Dev.*, vol. ED-34, no. 12, pp. 2469-2475, Dec. 1987.

4. M. J. Deen and Z. P. Zuo, "Edge-Effects In Narrow-Width MOSFETs", *IEEE Trans. Elec. Dev.*, vol. 38, no. 8, pp. 1815-1819, Aug 1991.

5. Y. P. Tsividis, *Operation And Modelling Of The MOS Transistor,* McGraw-Hill, New York, 1987.

6. F. Caswell, *Success In Statistics*, 1982.

7. N. D. Arora, L. A. Bair, and L. M. Richardson, "A New Method To Determine The MOSFET Effective Channel Width", *IEEE Trans. Elec. Dev.*, vol. 37, no. 3, pp. 811-814, Mar 1990.

8. P. G. Hoel, *Introduction To Mathematical Statistics,* Wiley & Sons, 1964.

Chapter 7

Discussion and Conclusions

7.1. Introduction

Three approaches have been taken in the investigation of the width effects in MOS devices: SEM, simulation and electrical extraction. Correlation between the three proved difficult. In the first instance the SEM measurement produces a single value result for the width loss. The simulation and electrical extraction show a width dependency on the applied gate voltage. Correlation between these two is limited for two reasons: the simulation process could only model the 'shape 1' profiles to end of process, and the electrical measurement showed a further width dependency on the data set chosen. However, conclusions can be drawn from the available data.

7.2. Results Comparison

Figure 7.1 shows the $\Delta W/2$ value for wafers 3,5, 9 and 15. The corresponding SEM values for these are shown in table 7.1.

Wafer	3	5	9	15
$\Delta W/2$	0.75	0.65	0.85	0.3

Table 7.1. SEM ΔW values.

Two curves are shown for each of samples 5 (5D and 5DE) and 9 (9A and 9AE). These represent the difference of overetch prior to the gate oxidation, and extend the bird's beak down the channel.

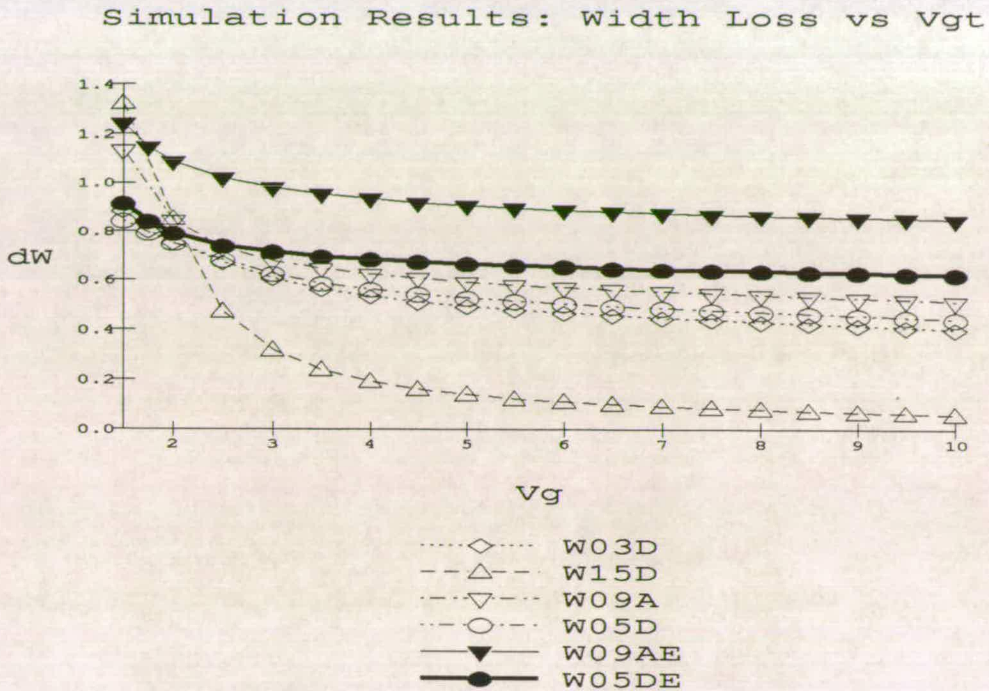


Figure 7.1. Simulation results for $\Delta W/2$.

For wafer 5, the SEM width loss value of 0.65 μm lies close to the y-axis value. i.e. close to the value at threshold. For wafer 9, the SEM value of 0.85 μm again lies close to the y-axis, but a greater difference is seen between the two etchback conditions for this wafer sample. (The appended E on the sample labels denotes those samples which have *not* been overetched). The latter sample shows a larger bird's beak, which extends more due to etchback than the sharper active area to field transition of wafer 5 which presents a thicker oxide at the gate edge. The similar result is seen for wafers 3 and 15. This is borne out in the results presented in chapter 5 where it was shown that the depletion region extension under the field at high applied gate voltages extended the channel width.

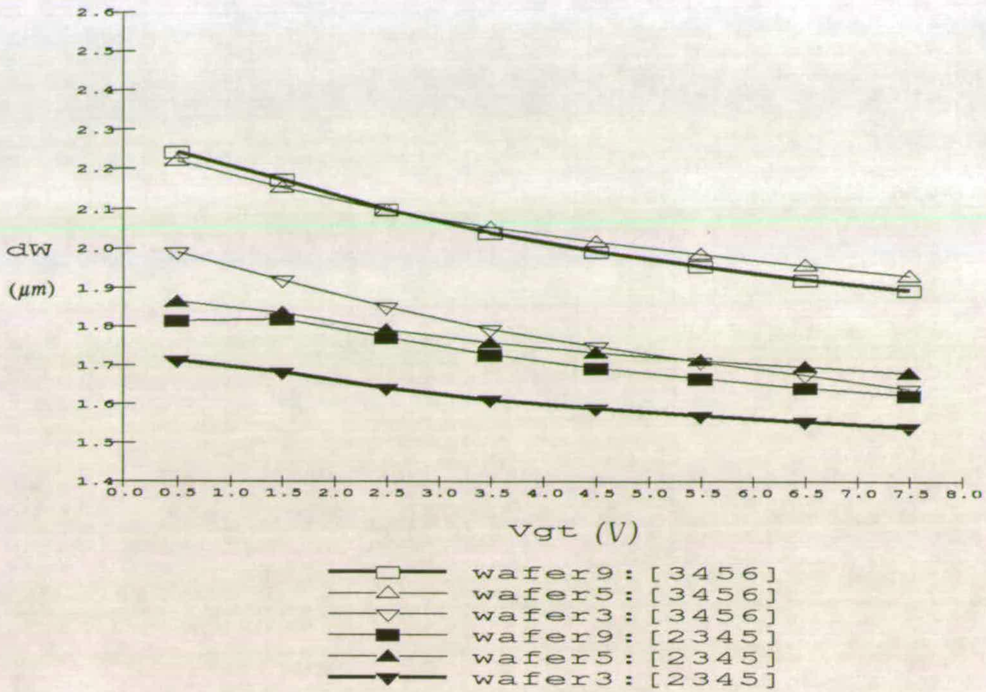


Figure 7.2. Electrical extraction results for ΔW

Comparison with the electrical results is dependent on the data set chosen. Figure 7.2 shows the ΔW values for wafer 3, 5 and 9. In almost all cases the equivalent width loss is greater when measured electrically. i.e. even at high V_{gt} when the depletion region extension widens the gate dimension, the extracted width loss is greater than that observed by SEM.

This discrepancy diminishes as the data set is chosen closer to the minimum data point, where the deviation in the $G_t - W_D$ graph is seen.(Figure 6.42).

Several different extraction techniques have been used in the verification of these results. A new method has been proposed, but shown to be no more accurate because of the inherent inaccuracies in the extraction of the primary width parameter. Hence satisfactory results for the edge currents could not be quoted.

7.3. Review

The thesis covered 3 main topics:

(a) SEM measurement

(b) simulation

(c) parameter extraction.

The impact on each is discussed.

7.3.1. SEM

The SEM sample cross-sections were obtained by sample polishing and etching to delineate the features. The process was greatly accelerated by the use of a dicing saw to provide "coarse" positioning for the initial cleave. This also ensured a clean break along the line of scribe with a 100% consistency. The final preparation before etching was done by polishing on a glass wheel, to smooth the the edge profile and largely eliminated the need for grind-back. The etched removal of 1000 Angstroms of silicon dioxide from the cleaved profile ensured a sharp edge to delineate the topographic profile.

The large number of samples meant that a great deal of time was spent on this technique, thus honing the process down to produce an efficient quick-turnaround time. Identical samples could then be prepared as a calibration check of the SEM process.

The results can be seen in the typical photographs shown in chapter 5. The start and stop points of the transistor width are still open to interpretation to some extent, though better than $0.1\mu m$ accuracy can be obtained.

These results have augmented the published data on bird's beak profiles over the given time and temperature range and have provided comprehensive information for the end-of-process results.

7.3.2. Simulation

The increased model sophistication currently available demands a higher level of knowledge on the part of the user to extract the maximum benefit from the software. This is particularly true when defining the initial grid even with the state-of-the-art software which was used during the course of this work. Problems have been highlighted in the *TSUPREM4* etch model ,whereby the lack of an isotropic etch facility restricts full modelling of the commonly used etch-back process. Apart from identifiable oxide etches, this process also frequently occurs as part of an RCA clean and is thus a major obstacle.

The latest solution to this problem is to interface the software to another package, *DEPICT*, which models lithography, and layer deposition and etching. In this way the process modelling is compartmentalised into discrete packages, allowing similar processes to be modelled, but necessitating more than one package to model a complete process. Unfortunately, this option was not available at the time. Within these constraints, and interfacing the device simulator *PISCES*, identifiable trends in the results were obtained and matched to those from the measured parameters.

7.3.3. Parameter Extraction

Different algorithms have been compared. A new technique has been proposed. All of the extraction routines have been shown to be prone to error.

The new extraction method proposes a normalisation of the transistor conductance by the conductance per unit micron in the channel centre. Since different transistors are necessary to obtain this value, uniformity of processing is critical. However it is possible that fluctuations as small as 3% in the normalised channel conductance technique are sufficient to invalidate the extraction routine.

Analysis of the inherent errors in the extraction process has exposed the problem of a dynamic data set. The linear extrapolation technique is prone to error depending upon the number and position of the points chosen. The use of the correlation coefficient has been shown to be inappropriate as a measure of accuracy and wrongly applied in the case of parameter extraction where one set of data points are pre-determined.

Parameter extraction results show a consistent trend of diminishing dependency on applied gate voltage. Large errors have been shown to occur at gate voltages close to threshold.

The thesis has for the first time compared all three techniques for width extraction and a large number of samples have been used. For all samples and all techniques, similar trends can be seen. The problem of inherent errors has been highlighted to show that considerable care must be taken when quoting extracted values.

7.4. Recommendations and Future Work

Each of the main topics could benefit from further work in these areas. The SEM processing was developed through trial and error and a useful technique produced. Current sample preparation is often performed by ion milling using a focused ion beam. Whilst this provides great positional accuracy, it is a very expensive process and the alternative used here provides a similar degree of accuracy,

at a much reduced cost.

As a result of this work, another project has been proposed and attracted industrial interest [1] . A commercial system is available to to digitise the SEM image for data storage. Because of the large amount of data, this is stored on an optical disk. Much of the SEM analysis is focused on edge profiles. This being so, the proposal is to set grey levels to identify the edges, so reducing the data. Once stored as a line, the information can be coded as $x - y$ coordinates. The next stage is to utilise a suitable curve fitting package to fit a curve to the SEM profile. The application of this could mean etch, resist, deposition, etc profiles being monitored and accurately recorded on a week-to-week, day-to-day or run-to-run basis. Deviations can be quantified mathematically, rather than 'eyeballing' the differences. Interest has also been shown from the modelling software manufacturers [2] who can then adjust the model parameters to an analytical equation and check for goodness of fit.

The impact of this thesis on the parameter extraction process is perhaps the largest of the three. The existence of a possible dynamic data set means that process development vehicles, (p.d.v.s), need themselves be characterised before being implemented in a production run. Algorithms which use extrapolation beyond the data set are most prone to error. Preferable are routines which confine themselves within the known data set, such as the slope value of the fitted line. Since the parameter extraction work is often used as a measure of process control, recognition of the flaws in the routines is imperative to prevent misinterpretation and consequent incorrect action.

References

1. B. Reece [Reece Scientific]. Private Communication

2. M. Kump [TMA]. Private Communication

APPENDIX

EDINBURGH MICROFABRICATION FACILITY

N-CHANNEL SILICON GATE PROCESS 2

BATCH NUMBER: START DATE:

DEVICE IDENTIFICATION: Eu

MASK SET:

MASKING SEQUENCE: 1 2 3 4 6 7 8
MASK REV.LETTERS: A A A A A A A

STARTING MATERIAL: 14-20 ohm.cm.(100) P-type,3in.Dia.

No. OF WAFER STARTS:

INITIAL CLEAN

Start date: Start time: Initials:

10 min. boil in 2:1, Sulphuric acid:Hydrogen peroxide in Teflon jig
D.I water wash
Dip 3 min. in 10% HF in Polypropylene jig
Wash and spin dry

Finish date: Finish time: Initials:

INITIAL OXIDE

Start date: Start time: Initials:

Furnace #1, 950oC, idling on oxygen
Preset gas flows as follows:
Oxygen 20% (1.5 l/min.)
HCl 15% (0.15 l/min.)
Hydrogen 10% (1.7 l/min.)
Load wafers into furnace with Oxygen only flowing.
5 min.Oxygen + HCl
10min.Oxygen + HCl + Hydrogen
5 min.Oxygen

Measure oxide thickness:

Finish date: Finish time: Initials:

SILICON NITRIDE DEPOSITION
Furnace #3, 800oC
Preset flows: Dichlorosilane,30 cc/min. Ammonia,90 cc/min.
Deposition time: 10 min.
Pressure during deposition: mTorr

Measure Si3N4 thickness:

Finish date: Finish time: Initials:

1st PHOTO (POSITIVE RESIST) LAYER# 1

Start date: Start time: Initials:

HMDS vapour box prime for 30 min.

Spin HPR 204 at 6000 rpm for 30 secs
Soft bake at 105oC for 30 min in static oven

Align and expose for secs. (2.5 mW/sq.cm.)

```
          |
          |
     x x x x x x x
    x     |     x
   x  Mask |      x
  x       |      x
   off left |
 x        |       x
    half of |
  x       |      x
   x  T/W  |      x
    x     |    x
     x    |   x
          x
          |
          |
```

Develop in 1 vol LSI 2 Developer, 3 vols DI water at 30oC for 60 secs
Inspect for proper development.

Measure resist image: microns

Hard bake at 130,oC for 30 min. in static oven
Inspect for proper baking

Finish date: Finish time: Initials:

BORON IMPLANT

Start date: Start time: Initials:

B11+ 2 e13 atoms/sq.cm. 130 keV

Finish date: Finish time: Initials:

4:1 ETCH

Start date: Start time: Initials:

Immerse in 4:1, Ammonium Fluoride soln.(40%w/v) : HF
Etch time: 30 secs
Wash and spin dry

Finish date: Finish time: Initials:

SILICON NITRIDE PLASMA ETCH

Start date: Start time: Initials:

Turn function switch to MANUAL.
Press STANDBY.
Press VENT to open system.
Press STANDBY when door opens.
Load wafers, one per pedistal, face upwards
Close door and press START.
Pump down to 30 mTorr with throttle fully open.
Press on PROCESS 3 (CF_4 + 5%O_2) and adjust flow to give 100 mTorr.
Add O_2 to increase pressure to 145 mTorr.
Press on POWER and set at 2 amps.
Etch until patterns clear.(approx. 1 min.)
Press off POWER and PROCESS 3.
Pump to 30 mTorr.

Press STANDBY and then VENT to open system.
Press STANDBY when door opens.
Remove wafers.
Close door and press START.

Repeat above with wafers face downwards to etch backs.

Finish date: Finish time: Initials:

RESIST STRIP

Start date: Start time: Initials:

Immerse in Fuming Nitric Acid 10 min
Wash and spin dry
Inspect for removal of resist.

Measure etched image: microns

Finish date: Finish time: Initials:

PRE-DIFFUSION CLEAN

Start date: Start time: Initials:

Dip in 10% HF for 15 secs.
Wash and spin dry.

Finish date: Finish time: Initials:

FIELD OXIDE

Start date: Start time: Initials:

Furnace #7, 950oC, idling on oxygen
Preset gas flows as follows:
Oxygen 24.5% (1.22 l/min.)
Hydrogen 20% (2.0 l/min.)
HCl 10% (0.1 l/min.)
Load wafers into furnace with Oxygen only flowing.
5 min.Oxygen + HCl
0.5 hours Oxygen + Hydrogen + HCl
15.5 hours Oxygen + Hydrogen
5 min.Oxygen

Measure oxide thickness:

Finish date: Finish time: Initials:

RESIST COAT

Start date: Start time: Initials:

Spin HPR 204 at 6000 rpm for 30 secs.
Hard bake for 30 min. at 130oC in static oven.
Inspect for proper baking

Finish date: Finish time: Initials:

4:1 ETCH

Start date: Start time: Initials:

Immerse in 4:1, Ammonium Fluoride soln.(40%w/v) : HF
Etch time: (Backs dewet)
Wash and spin dry

Finish date: Finish time: Initials:

RESIST STRIP

Start date: Start time: Initials:

Immerse in Fuming Nitric Acid 10 min
Wash and spin dry
Inspect for removal of resist.

Finish date: Finish time: Initials:

OXIDE ETCH

Start date: Start time: Initials:

Immerse in 4:1, Ammonium Fluoride soln.(40%w/v) : HF for 15 secs.
Wash and spin dry

Finish date: Finish time: Initials:

SILICON NITRIDE PLASMA ETCH

Start date: Start time: Initials:

Turn function switch to MANUAL.
Press STANDBY.
Press VENT to open system.
Press STANDBY when door opens.
Load wafers, one per pedistal, face upwards
Close door and press START.
Pump down to 30 mTorr with throttle fully open.
Press on PROCESS 3 (CF4 + 5%O2) and adjust flow to give 100 mTorr.
Add O2 to increase pressure to 145 mTorr.
Press on POWER and set at 2 amps.
Etch until patterns clear.(approx. 1 min.)
Press off POWER and PROCESS 3.
Pump to 30 mTorr.

Press STANDBY and then VENT to open system.
Press STANDBY when door opens.
Remove wafers.
Close door and press START.

Finish date: Finish time: Initials:

OXIDE ETCH

Start date: Start time: Initials:

Immerse in 4:1, Ammonium Fluoride soln.(40%w/v) : HF to dewet scribelines
Wash and spin dry

Measure etched image: microns

Finish date: Finish time: Initials:

2nd PHOTO (POSITIVE RESIST) LAYER# 2

Start date: Start time: Initials:

Pre-spin bake, furnace#1, 950oC, Nitrogen 20% (2 l/min), 5 min.

Spin HPR 204 at 6000 rpm for 30 secs
Soft bake at 105oC for 30 min in static oven
Align and expose for secs. (2.5 mW/sq.cm.)

Develop in 1 vol LSI 2 Developer, 3 vols DI water at 30oC for 60 secs
Inspect for proper development.

Measure resist image: microns

Hard bake at 130oC for 30 min. in static oven.
Inspect for proper baking

Finish date: Finish time: Initials:

ARSENIC IMPLANT

Start date: Start time: Initials:

As75++ 1.5e12 atoms/sq.cm. 90 keV

Finish date: Finish time: Initials:

RESIST STRIP

Start date: Start time: Initials:

Immerse in Fuming Nitric Acid 10 min
Wash and spin dry
Inspect for removal of resist.

Finish date: Finish time: Initials:

PRE-DIFFUSION CLEAN

Start date: Start time: Initials:

Dip in 10% HF for 15 secs.
Wash and spin dry.

Finish date: Finish time: Initials:

GATE OXIDE

Start date: Start time: Initials:

Furnace #1, 950oC, idling on oxygen
Preset gas flows as follows:
Oxygen 20% (1.5 l/min.)
HCl 15% (0.15 l/min.)
Hydrogen 10% (1.7 l/min.)
Load wafers into furnace with Oxygen only flowing.
5 min.Oxygen + HCl
17min.Oxygen + HCl + Hydrogen
5 min.Oxygen

Measure oxide thickness:

Finish date: Finish time: Initials:

BORON IMPLANT

Start date: Start time: Initials:

B11+ 4.0 e11 atoms/sq.cm. 40 keV

Finish date: Finish time: Initials:

RESIST STRIP

Start date: Start time: Initials:

Immerse in Fuming Nitric Acid 10 min
Wash and spin dry

Finish date: Finish time: Initials:

ANNEAL

Start date: Start time: Initials:

Furnace #1, 950oC, idling on nitrogen
Preset gas flows as follows:
Nitrogen 20% (2.0 l/min.)
Load wafers into furnace for 30 min.

3rd PHOTO (POSITIVE RESIST) LAYER# 3

Start date: Start time: Initials:

Spin HPR 204 at 6000 rpm for 30 secs
Soft bake at 105oC for 30 min in static oven
Align and expose for secs. (2.5 mW/sq.cm.)

```
   x x x x x x x
  x           x
 x             x
x               x
    Do NOT
x               x
   expose T/W
x               x
 x             x
  x           x
   x         x
       x
```

Develop in 1 vol LSI 2 Developer, 3 vols DI water at 30oC for 60 secs
Inspect for proper development.

Measure resist image: microns

Hard bake at 130oC for 30 min. in static oven.
Inspect for proper baking

Finish date: Finish time: Initials:

OXIDE ETCH

Start date: Start time: Initials:

Immerse in 4:1, Ammonium Fluoride soln.(40%w/v) : HF to dewet scribelines
Wash and spin dry

Finish date: Finish time: Initials:

RESIST STRIP

Start date: Start time: Initials:

Immerse in Fuming Nitric Acid 10 min
Wash and spin dry
Inspect for removal of resist.

Measure etched image: microns

Finish date: Finish time: Initials:

PRE-DIFFUSION CLEAN

Start date: Start time: Initials:

Dip in 10% HF for 5 secs.
Wash and spin dry.

Finish date: Finish time: Initials:

POLY-SILICON DEPOSITION

Start date: Start time: Initials:

Furnace #4, 600oC
Preset flows: Silane,60 cc/min.
Deposition time: 45 min.
Pressure during deposition: mTorr

Measure poly-silicon thickness:

Finish date: Finish time: Initials:

POLY OXIDE

Start date: Start time: Initials:

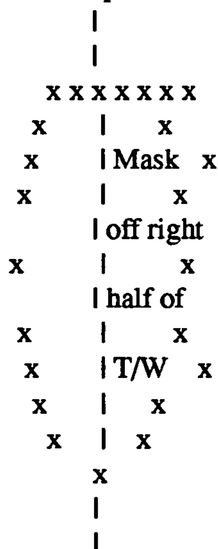
Furnace #1, 950oC, idling on oxygen
Preset gas flows as follows:
Oxygen 20% (1.5 l/min.)
HCl 15% (0.15 l/min.)
Hydrogen 10% (1.7 l/min.)
Load wafers into furnace with Oxygen only flowing.
5 min.Oxygen + HCl
18min.Oxygen + HCl + Hydrogen
5 min.Oxygen

Finish date: Finish time: Initials:

4th PHOTO (POSITIVE RESIST) LAYER# 4

Start date: Start time: Initials:

Spin HPR 204 at 6000 rpm for 30 secs
Soft bake at 105oC for 30 min in static oven
Align and expose for secs. (2.5 mW/sq.cm.)



Develop in 1 vol LSI 2 Developer, 3 vols DI water at 30oC for 60 secs
Inspect for proper development.

Measure resist image: microns

Hard bake at 130oC for 30 min. in static oven.
Inspect for proper baking

Finish date: Finish time: Initials:

OXIDE ETCH

Start date: Start time: Initials:

Immerse in 4:1, Ammonium Fluoride soln.(40%w/v) : HF to dewet t/w

Wash and spin dry

Finish date: Finish time: Initials:

POLY SILICON PLASMA ETCH

Start date: Start time: Initials:

Turn function switch to MANUAL.
Press STANDBY.
Press VENT to open system.
Press STANDBY when door opens.
Load wafers, one per pedistal, face upwards
Close door and press START.
Pump down to 30 mTorr with throttle fully open.
Press on PROCESS 3 (CF_4 + 5%O_2) and adjust flow to give 150 mTorr.
Press on POWER and set at 2 amps.
Re-adjust gas flow to keep pressure in the range 150-175 mTorr.
Etch until patterns clear.(approx. 3 min.)
Press off POWER and PROCESS 3.
Pump to 30 mTorr.

Repeat above with wafers face downwards to etch backs.

Press STANDBY and then VENT to open system.
Press STANDBY when door opens.
Remove wafers.
Close door and press START.

Finish date: Finish time: Initials:

RESIST STRIP

Start date: Start time: Initials:

Immerse in Fuming Nitric Acid 10 min
Wash and spin dry
Inspect for removal of resist.

Measure etched image: microns

Finish date: Finish time: Initials:

OXIDE ETCH

Start date: Start time: Initials:

Immerse in 4:1, Ammonium Fluoride soln.(40%w/v) : HF to dewet t/w
Wash and spin dry

Measure etched image: microns

Finish date: Finish time: Initials:

PRE-DIFFUSION CLEAN

Start date: Start time: Initials:

Dip in 10% HF for 15 secs.
Wash and spin dry.

Finish date: Finish time: Initials:

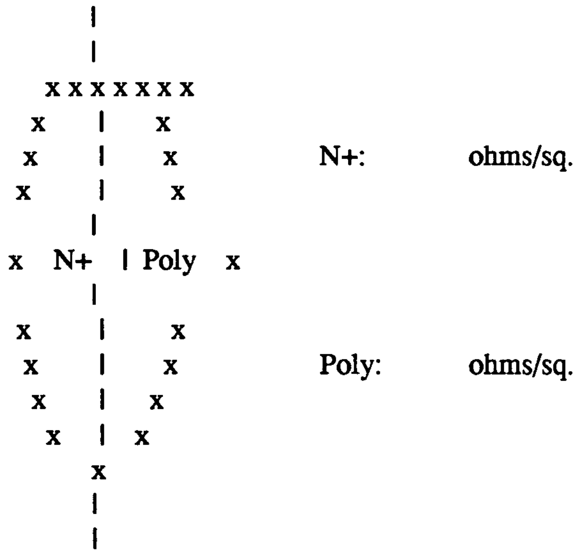
PHOSPHORUS DEPOSITION (SOLID SOURCE)

Start date: Start time: Initials:

Furnace #5, 850oC, idling on Nitrogen
Preset gas flows as follows:
Nitrogen 50 (2.0 l/min.)

Insert wafers at 4 ins./min.
20 min. ramp at 8.1oC/min. to up 1000oC
15 min. soak
20 min. ramp at 8.1oC/min. down to 850oC
Remove wafers at 4ins./min.

Measure sheet resistances



N+: ohms/sq.

Poly: ohms/sq.

Finish date: Finish time: Initials:

PHOSPHORUS DEGLAZE

Start date: Start time: Initials:

Immerse for 30 secs. in 10% HF (1 vol HF, 9 vols DI)
Wash and spin dry

Finish date: Finish time: Initials:

POLY OXIDE

Start date: Start time: Initials:

Furnace #7, 950oC, idling on oxygen
Preset gas flows as follows:
Oxygen 20% (1.5 l/min.)
HCl 15% (0.15 l/min.)
Hydrogen 10% (1.7 l/min.)
Load wafers into furnace with Oxygen only flowing.
5 min.Oxygen + HCl
18min.Oxygen + HCl + Hydrogen
5 min.Oxygen

Finish date: Finish time: Initials:

5th PHOTO (POSITIVE RESIST) LAYER# 6

Start date: Start time: Initials:

Spin HPR 204 at 6000 rpm for 30 secs
Soft bake at 105oC for 30 min in static oven
Align and expose for secs. (2.5 mW/sq.cm.)

```
   x x x x x x x
  x           x
 x             x
x               x
     Do NOT
x               x
   expose T/W
x               x
 x             x
  x           x
   x        x
        x
```

Develop in 1 vol LSI 2 Developer, 3 vols DI water at 30oC for 60 secs
Inspect for proper development.

Measure resist image: microns

Hard bake at 130oC for 30 min. in static oven.
Inspect for proper baking

Finish date: Finish time: Initials:

OXIDE ETCH

Start date: Start time: Initials:

Immerse in 4:1, Ammonium Fluoride soln.(40%w/v) : HF to dewet t/w
Wash and spin dry

Finish date: Finish time: Initials:

RESIST STRIP

Start date: Start time: Initials:

Immerse in Fuming Nitric Acid 10 min
Wash and spin dry
Inspect for removal of resist.

Measure etched image: microns

Finish date: Finish time: Initials:

REFLOW PYRO DEPOSITION

Start date: Start time: Initials:

PWS 2000 Hotplate at 430oC, in right hand rest position.
Preset SPEED: 200
Preset gas flows as follows:
Nitrogen(02),80 (4.0 l/min):Nitrogen(SiH4),60 (2.9 l/min)
5%Silane,110 (1.3l/min): 1%Phosphene,80 (0.96l/min): Oxygen,60 (0.65l/min)
Water flow, 100

Include one fresh 14-20 ohm.cm. P-type test wafer per batch.

Place up to 6 wafers onto hotplate as close as possible to centre.
Press button <-- to deposit
When hotplate stops, remove wafers to steel worktop with Bernoulli tweezer
Reload with fresh wafers and press button -->
Continue with whole batch.

Measured thickness:

NOTE: Proceed immediately to First Reflow

Finish date: Finish time: Initials:

FIRST REFLOW
FURNACE# 2 (O2)

Temperature: 1050oC
Idling ambient: Oxygen

Preset gas flows as follows:
Oxygen 50 (2.0 l/min.)

Start date: Start time: Initials:

20 min. furnace time

Etch reflow pyro off extra T/W in 4:1, NH4F:HF.

Measure sheet resistance at 5 points: ohms/sq.
Discard T/W after use.

Finish date: Finish time: Initials:

DENSIFICATION OF REFLOW PYRO

Start date: Start time: Initials:

Furnace #7, 950oC, idling on oxygen
Preset gas flows as follows:
Oxygen 28% (1.4 l/min.)
Hydrogen 20% (2.0 l/min.)
Load wafers into furnace with Oxygen only flowing.
5 min.Oxygen
15 min.Oxygen + Hydrogen
5 min.Oxygen

NOTE: Proceed immediately to resist spin for contact photo.

Finish date: Finish time: Initials:

6th PHOTO (POSITIVE RESIST) LAYER# 6

Start date: Start time: Initials:

Spin HPR 204 at 6000 rpm for 30 secs
Soft bake at 105oC for 30 min in static oven
Align and expose for secs. (2.5 mW/sq.cm.)

```
     x x x x x x x
   x           x
  x             x
 x               x
     Expose  T/W
x                 x
     completely
 x               x
  x             x
   x           x
     x       x
         x
```

Develop in 1 vol LSI 2 Developer, 3 vols DI water at 30oC for 60 secs
Inspect for proper development.

Measure resist image: microns

Hard bake at 130oC for 30 min. in static oven.
Inspect for proper baking

Finish date: Finish time: Initials:

OXIDE ETCH

Start date: Start time: Initials:

Immerse in 4:1, Ammonium Fluoride soln.(40%w/v) : HF to dewet t/w
Wash and spin dry

Finish date: Finish time: Initials:

RESIST STRIP

Start date: Start time: Initials:

Immerse in Fuming Nitric Acid 10 min
Wash and spin dry
Inspect for removal of resist.

Measure etched image: microns

Finish date: Finish time: Initials:

SECOND REFLOW

Start date: Start time: Initials:

Furnace #2, 1050oC, idling on Nitrogen
20 min. Nitrogen ,25,(1 l/min.)

Finish date: Finish time: Initials:

PRE-ALUMINIUM EVAPORATION CLEAN (N-CH.SI GATE)

Start date: Start time: Initials:

Dip 20 secs. in reflow etch: 25 vols Ammonium Fluoride soln.(40%w/v),1 vol HF
Wash and spin dry

Re-measure sheet resistances

```
         |
         |
     x x x x x x x
    x    |    x
   x     |     x            N+:        ohms/sq.
  x      |      x
         |
 x   N+  | Poly   x
         |
```

```
 x      |     x
  x     |    x          Poly:        ohms/sq.
   x    |   x
    x   |  x
        x
        |
        |
```

Discard test wafer

Finish date: Finish time: Initials:

ALUMINIUM EVAPORATION (Si GATE)

Start date: Start time: Initials:

Load wafers on palettes and load into Balzers BAS 450 coater.
Pump system to 5e-6 or better with Meissner trap.
Close shutter
Throttle pump and admit argon at 2e-3.
Set integrator at 11500, Range 3
Run up Aluminium/Silicon target to 6 kW.
Open shutter until integrator times out.
Warm up Meissner and chamber.
Vent system and remove wafers.

Finish date: Finish time: Initials:

7th PHOTO (POSITIVE RESIST) LAYER# 7

Start date: Start time: Initials:

Spin HPR 204 at 4000 rpm for 30 secs
Soft bake at 105oC for 30 min in static oven
Align and expose for secs. (2.5 mW/sq.cm.)
Develop in 1 vol LSI 2 Developer, 3 vols DI water at 30oC for 60 secs
Inspect for proper development.

Measure resist image: microns

Hard bake at 130oC for >60 min. in static oven.
Inspect for proper baking

Finish date: Finish time: Initials:

ALUMINIUM ETCH

Start date: Start time: Initials:

R.I.E. etch to clear patterns.

Finish date: Finish time: Initials:

RESIST STRIP

Start date: Start time: Initials:

Immerse in Fuming Nitric Acid 10 min
Wash and spin dry
Inspect for removal of resist.

Measure etched image: microns

Measure AlSi thickness: Microns

Finish date: Finish time: Initials:

SINTER

Start date: Start time: Initials:

Furnace #8, 435oC, idling on Nitrogen
5 min. Nitrogen,25,(1 l/min.)
10 min. 40% Hydrogen/Nitrogen,25,(1 l/min.)
5 min. Nitrogen

Finish date: Finish time: Initials:

ELECTRICAL TEST

Start date: Start time: Initials:

Finish date: Finish time: Initials:

OVERLAY PYRO DEPOSITION

Start date: Start time: Initials:

PWS 2000 Hotplate at 430oC, in right hand rest position.
Preset speeds: 330 <-- and 600 -->
Preset gas flows as follows:
Nitrogen(02),80 (4.0 l/min):Nitrogen(SiH4),60 (2.9 l/min)
5%Silane,110 (1.3 l/min): 1%Phosphene,20 (0.15 l/min): Oxygen,60 (0.65l/min)
Water flow,100

Place up to 6 wafers onto hotplate as close as possible to centre.
Press PROGRAM button and START button
When hotplate stops, remove wafers to steel worktop with Bernoulli tweezer
Continue with whole batch.

Measure thickness: Angstroms

Finish date: Finish time: Initials:

8th PHOTO (POSITIVE RESIST) LAYER# 8

Start date: Start time: Initials:

Spin HPR 204 at 4000 rpm for 30 secs
Soft bake at 105oC for 30 min in static oven
Align and expose for secs. (2.5 mW/sq.cm.)
Develop in 1 vol LSI 2 Developer, 3 vols DI water at 30oC for 60 secs
Inspect for proper development.
Hard bake at 130oC for 30 min. in static oven.
Inspect for proper baking

Finish date: Finish time: Initials:

OVERLAY PYRO ETCH

Start date: Start time: Initials:

Immerse in 4:1, Ammonium Fluoride soln.(40%w/v) : HF to dewet scribelines
Wash and spin dry

Finish date: Finish time: Initials:

RESIST STRIP

Start date: Start time: Initials:

Immerse in Fuming Nitric Acid 10 min
Wash and spin dry
Inspect for removal of resist.

Finish date: Finish time: Initials:

Papers associated with this thesis.

M. Fallon, J. M. Robertson, A. J. Walton, R. J. Holwill "Examination Of LOCOS Process Parameters And The Measurement Of Effective Width" Proc. IEEE 1991 Int. Conf. for Microelectronics Test Structures, Japan March 1991

R.S Ferguson, A.J. Walton and M. Fallon "An Interactive Process Modelling Simulator; its Capabilities and Performance" Swansea, Proc Process and Device Modelling Alvey Meeting, 19-21 Sept 1988, PP54-71

R.S. Ferguson, M. Fallon and A.J. Walton "The Capability and Design of an Interactive Process Simulator" UK IT 88 Swansea 4-7 July 1988

R.S. Ferguson, A.J. Walton and M. Fallon "Developments with EQUIPS" The Third Process and Device Modelling Alvey Conference, PP42-46 Oct 1987

EXAMINATION OF LOCOS PROCESS PARAMETERS AND THE MEASUREMENT OF EFFECTIVE WIDTH

M. Fallon, J.M. Robertson, A.J. Walton, R.J. Holwill

Edinburgh Microfabrication Facility
Department of Electrical Engineering
King's Buildings
University of Edinburgh,
Edinburgh, EH9 3JL, UK.

\bstract: Device isolation by means of LOCOS and field mplantation are commonly incorporated in current MOS rocesses. These two process steps interact to impact the ffective MOS transistor width. This paper examines the opographical features determined by pad oxide and nitride hicknesses and compares the physical with the effective lectrical width.

. INTRODUCTION

Device isolation by means of LOCOS [1], is a widely used echnology and has been the subject of much research, articularly as device dimensions have reduced towards the ubmicron regime. Inherent in the LOCOS process is the loss f silicon caused by the lateral diffusion of the oxidant and the o-called bird's beak [2]. As the bird's beak may be as great s 1μm, it can play a significant part in determining device haracteristics [3], as well as consuming relatively large areas or sub-micron technologies.

Whilst the topography of the bird's beak impacts the urface width of the MOS transistor, it is normally omplicated by the presence of a field inversion implant [4]. Thus the quoted bird's beak length, L_{bb}, is not sufficient in tself to predict the effective loss of the drawn dimension.

This work investigates variation of topographical structure vith pad oxide-nitride thicknesses and compares the lectrically extracted width against these topographical eatures.

. FORMATION OF THE BIRD'S BEAK

The bird's beak features are detailed in figure 1 where he encroachment of the field oxide into the active area can be bserved. It has been recognised [5], that thicker pad oxide nd/or thinner silicon nitride can increase the length of the ird's beak, L_{bb}: a thicker pad oxide exposes more area to xidant diffusion and a thinner nitride lessens the rigidity of he mask allowing greater deformation which effectively ncreases the size of the oxidation window. Methods of educing this feature have included the introduction of an ntermediate oxynitride layer. Other technologies involve ilicon etching [6] or sacrificial polysilicon [7]. All these ntroduce further process steps with the aim of improving levice isolation and/or increasing packing density. The conventional LOCOS process is limited by the presence of the ird's beak and while its effect may be minimised it can never e eradicated. For good isolation an appreciable field oxide hickness is required and so the option used here to reduce the

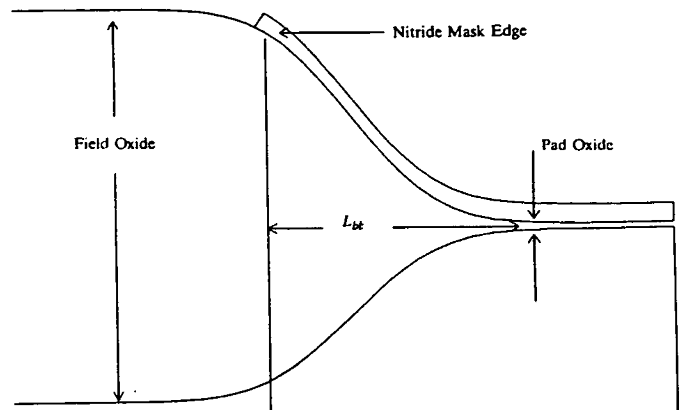


Figure 1. LOCOS Profile

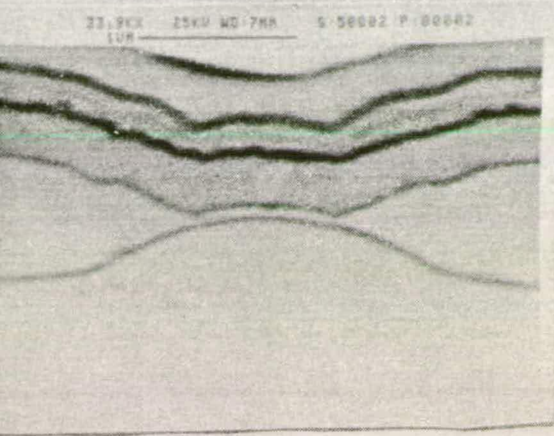
length of the bird's beak is to select an appropriate combination of pad oxide and nitride thickness. This paper examines this relationship and its effect upon the effective width of MOS transistors.

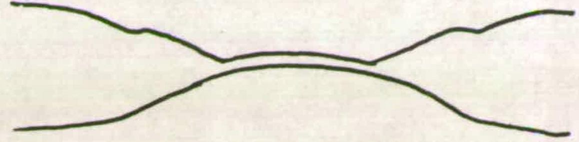
3. EXPERIMENT

A standard NMOS process using <100> 14-20 Ω-cm p-type material was employed. The boron field and threshold adjust implants were 2×10^{13} at 100keV and 4×10^{11} at 40keV respectively. The field and gate oxides were 1.2 μm and 800A respectively. The gate oxidation cycle was repeated to provide a sacrificial oxidation to eliminate the white ribbon effect [8]. To observe the effect of different of pad oxide and nitride thicknesses the following combinations were processed: the pad oxides were incremented in 8 steps from 50A to 930A and the nitride in 5 steps from 370A to 5500A. A wet oxidation at 950°C was used to grow the pad oxide and the LPCVD nitride was deposited at 800°C. The required patterns were defined using RIE. This produced a 40 element matrix after which all wafers received identical processing.

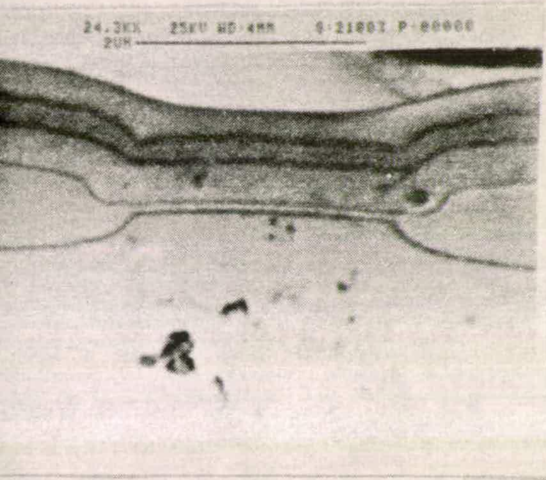
From the matrix of oxide/nitride thicknesses, 3 classes of cross-sectional profiles can result as illustrated in figure 2. For the thin nitride, corresponding to low mask stress, a smooth transition from gate to field oxides can be observed. As the nitride thickens, a step in the silicon substrate appears which is associated with increasing stress in the nitride. In the extreme the familiar bird's beak crest disappears entirely, with virtually complete suppression of oxidation beyond the original silicon surface covered by the nitride mask.

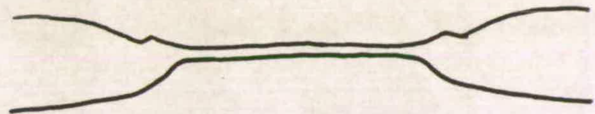
CH2907-4/91/0000-0157$01.00 ©1991 IEEE



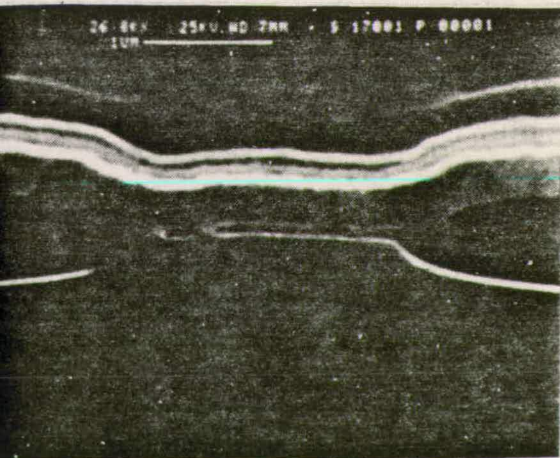


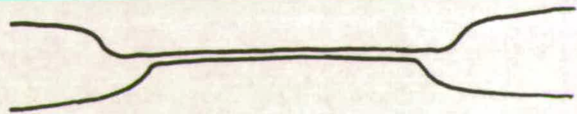
2(a) T_n = 820A





2(b) T_n = 4400A





2(c) T_n = 5500A

Figure 2. The three classes of LOCOS profiles.

4. MEASUREMENTS AND RESULTS

A 2×n pad array layout was used to connect to the discrete enhancement MOS transistors. Six different design widths of 1μm, 1.5μm, 2.5μm, 5μm, 7μm and 10μm were matched to a length of 5μm with each transistor individually connected to its own set of pads.

The measurement consisted of biasing the transistors into inversion and extracting the maximum transconductance (g_m). On average 16 measurements were made for each combination of pad oxide (T_p) and nitride (T_n) thickness. The effective loss in electrical width was obtained from the intercept on the width axis and figure 3 shows one such measurement.

Figure 4 shows the expected relationship between ΔW_{eff} and T_p and T_n: as the pad oxide increases ΔW_{eff} increases, particularly so for thinner oxides. Furthermore, as T_n increases, ΔW_{eff} is reduced. For thin nitrides ΔW_{eff} is between 1.4μm and 2.5μm. This means that either the bird's beak length or the distance that field implant has diffused is between 0.7μm and 1.25μm. The increase of ΔW_{eff} with T_p implies it is the former, thus fixing the maximum length of field implant diffusion at 0.7μm, which is the minimum value of L_{bb}. However, as T_n increases, ΔW_{eff} becomes largely independent of the thickness of the pad oxide and has a value of about 0.7μm. Consequently, we can deduce that maximum distance the field implant diffuses is about 0.35μm.

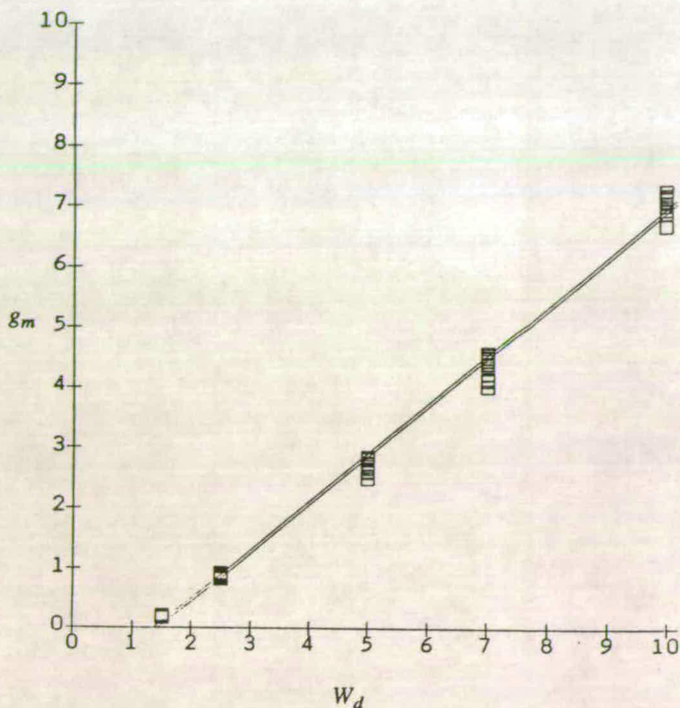


Figure 3. Transconductance (g_m) vs drawn width (W_d)

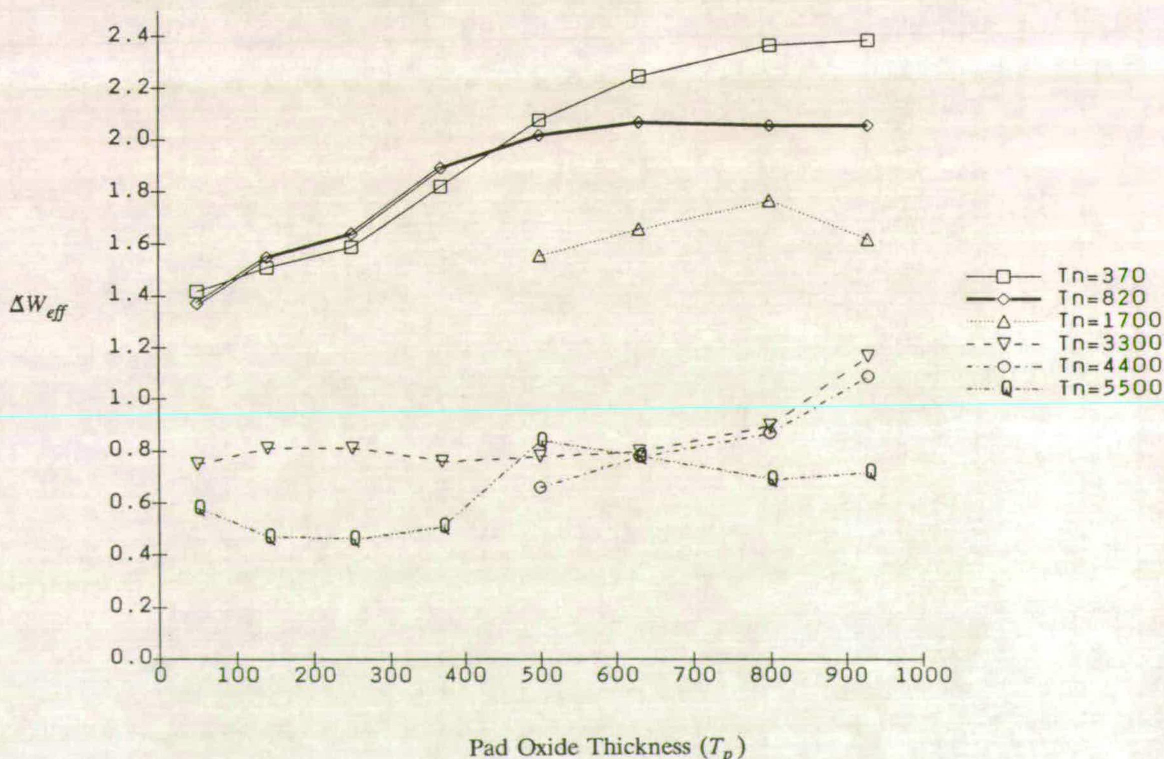


Figure 4. ΔW_{eff} vs pad oxide thickness (T_p) and nitride thickness (T_n)

After measurement of ΔW_{eff}, devices with 2.5μm drawn ns were cross-sectioned and examined under the SEM. topographical width of the conducting channel (W_s) as ed by the onset of the bird's beak was measured. Figure dicates this dimension together with a number of others n will be used later. Although no nitride edge is visible at end of the process, the length of the bird's beak can be red from the drawn device width and the width of the tant channel as given below.

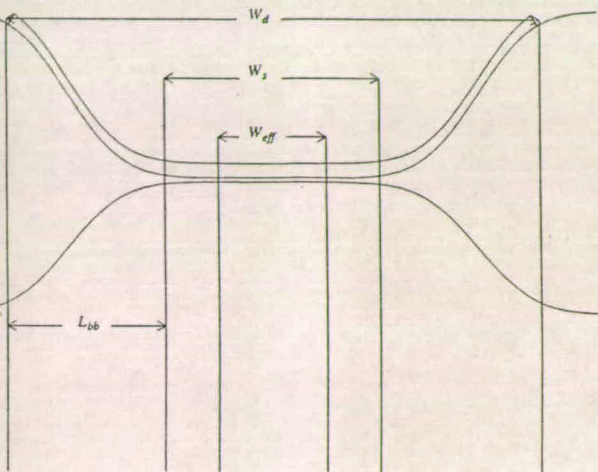


Figure 5. Dimensions used to define LOCOS structure

$$L_{bb} = \frac{W_d - W_s}{2} \qquad (1)$$

Figure 6 uses a transistor with W_d = 2.5μm as a reference to compare W_{eff} and W_s and extract the topographical effect of the bird's beak.

Whilst W_{eff} is never greater than W_s, it can be seen that at low values of nitride thickness, W_{eff} and W_s follow each other quite closely, to within 0.25μm. At higher values of T_n, which produce the extended planar channel of figure 2(c), (shown in close-up in figure 7), the difference between W_s and W_{eff} can be high as 0.5μm. This difference then, is simply the loss in effective width due to the field implant diffusion extending beyond the bird's beak. From figure 6 the minimum value of L_{bb} is obtained when W_s is a maximum. Using equation (1) the minimum value of L_{bb} is 0.11μm, and the distance implant has diffused past this point is 0.24μm, giving a total loss of effective width on one side of 0.35μm which is similar to the measurements made in reference [9], where a pad oxide of 20A was used in in combination with 2000A of silicon nitride.

5. CONCLUSIONS

Since W_s is always greater than W_{eff}, no extension of conducting channel beyond the gate oxide occurs, despite the projection of the planar gate dielectric above the angled silicon substrate [10]. Thus no corner, or Inverse Narrow Width Effect is seen which agrees with previously reported results [3].

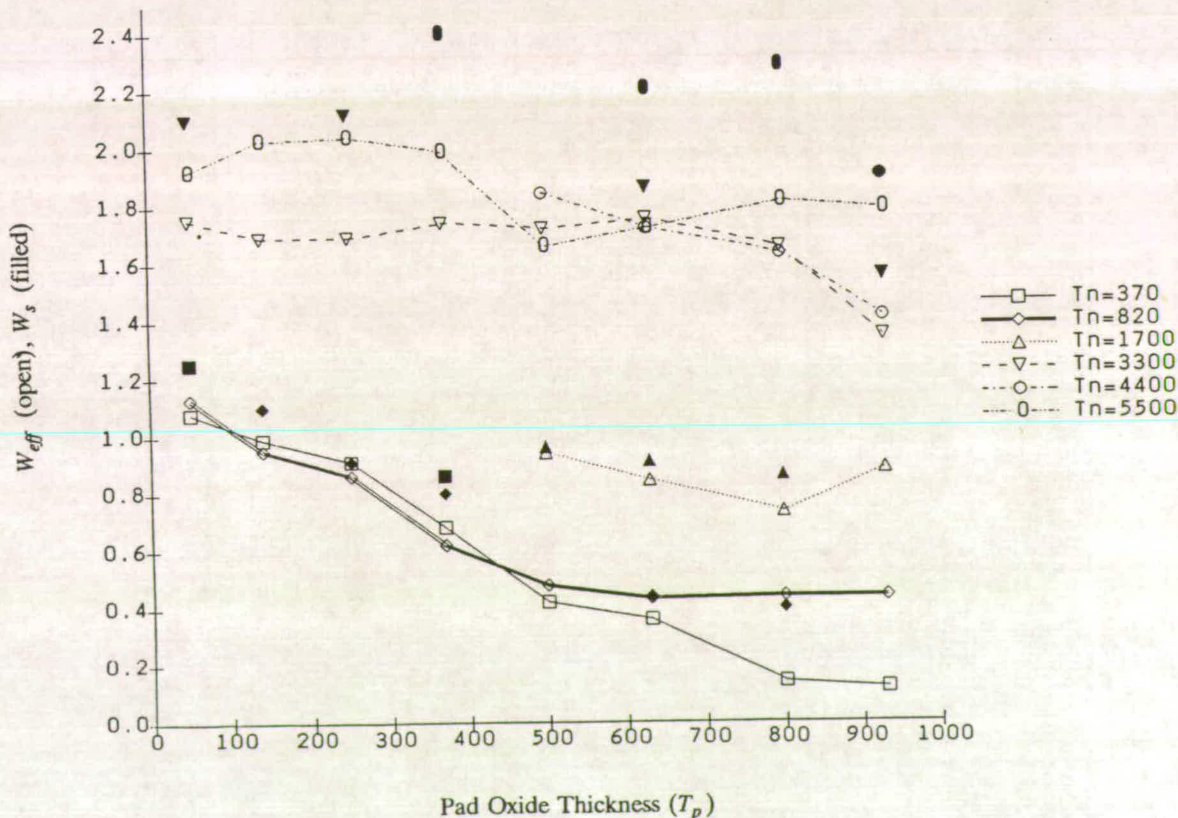


Figure 6. W_{eff} (open) and W_s (filled) vs T_p and T_n

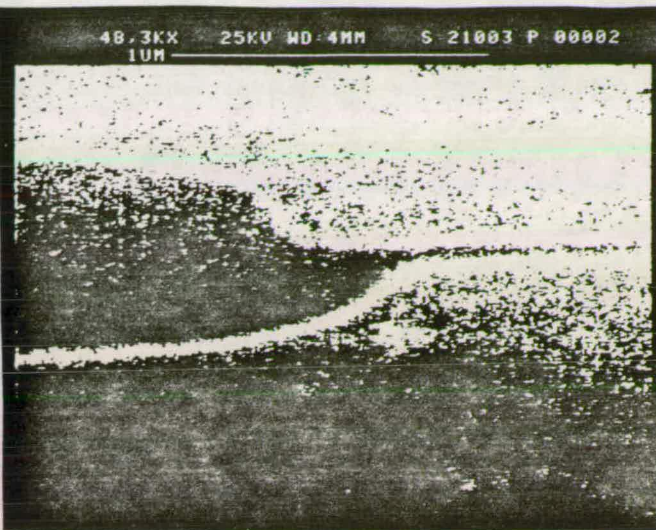


Figure 7: Extension of gate electrode beyond conducting channel.

The limit then in topographical packing density may be achieved by physical reduction of the bird's beak, but for varying pad oxide/nitride mask combinations the effective device width is limited by the presence of the field implant.

6. ACKNOWLEDGEMENTS.

This work has been supported by the Science and Engineering Research Council under grants GR/F 30499 and GR/F 38884.

7. REFERENCES.

[1] J. A. Appels, E. Kool, M. M. Paffen, J. J. H. Schatorje, W. H. C. G. Verkuylen, "Local Oxidation of Silicon and its Application in Semiconductor-Device Technology", Phillips Research Reports, Vol 25, pp 118-132, 1970

[2] T. C. Wu, W. T. Stacy, K. N. Ritz, "The Influence of the LOCOS Processing Parameters on the Shape of the Birds's Beak Structure", J. Electrochem Soc, Vol 130, no 7, pp 1563-1566, July 1983

[3] F. M. Klassen, P. A. van der Plas, R. J. W. Debets, N. A. H. Wils, M. G. Pitt, "Narrow-width Effects in Submicron MOS ICs", Proc ESSDERC'89, Berlin, pp 105-108, 1989

[4] A. M. Asenov, E. N. Stefanov, B. Z. Antov, "Numerical Study of the Effect of the Doping Profile on the Threshold Voltage of Narrow-Channel MOS Transistor", Solid State Electronics, Vol 30, no 12 pp 1305-1315, 1987

[5] N. Guillemot, G. Pananakakis, P. Chenevier, " A New Analytical Model of the Bird's Beak", IEEE Trans ED, Vol ED-34, no 5, pp 1033-1038, May 1987

[6] K. J. Chiu, J. L. Moll, J. Manoliu, "A Bird's Beak Free Local Oxidation Technology Feasible for VLSI Circuits Fabrication", IEEE J Solid-State Circuits, Vol SC-17, no 2, pp 166-170, April 1982

[7] N. Matsukawa, H. Nozawa, J. Matsunaga, S. Kohyama, "Selective Polysilicon Oxidation Technology for VLSI Isolation", IEEE Trans ED, Vol ED-29, no 4, pp 561-567, April 1982

[8] T. A. Shankoff, T. T. Sheng, S. E. Haszko, R. B. Marcus, T. E. Smith, "Bird's Beak Configuration and Elimination of Gate Oxide Thinning Produced during Selective Oxidation", J. Electrochem Soc, Vol 127, no 1, pp 216-222, Jan 1980

[9] J. G. Dil, J. W. Barsten, R. D. J. Verhaar, A. E. T. Kuiper, "LOCOS with Thick and Thin Nitride Masking", Phillips Journal of Research, Vol 40, no 2, pp 72-87, 1985.

[10] K.K. Hsueh, J.J Sanchez, T.A. Demassan, L.A. Akers, "Inverse-Narrow-Width Effect and Small-Geometry MOSFET Threshold Voltage Model", IEEE Trans Dlectron Devices, vol ED-35, no 3, pp 325-338 ,March 1988

[11] J. C. Marchetaux, B. S. Doyle, A. Boudou, "Interface States Under LOCOS Beak Region", Solid State Electronics, Vol 30, no 7 pp 745-753, 1987

AN INTERACTIVE PROCESS MODELLING SIMULATOR: ITS CAPABILITY AND PERFORMANCE

R.S. Ferguson[†] A.J. Walton, M. Fallon

[†]Department of Electrical Engineering
Ashby Building
Queens University
Stranmillis Road
BELFAST

Edinburgh Microfabrication Facility
Department of Electrical Engineering
King's Buildings
Mayfield Road
EDINBURGH EH9 3JL

1. INTRODUCTION

The paper reviews the work leading to the production of the interactive process modelling package EQUIPS.

EQUIPS was developed primarily because there was a requirement for a modelling package that would perform 2-dimensional process simulation with input and output in an easily comprehensible format while not requiring significant computational overheads. The code has been designed to be portable and is written in FORTRAN77. It executes on the latest generation of personal workstation systems, such as the VAX Station 2000.

The constraints that such a specification imposed on the numerical model led us to select a moving/modifiable finite elment procedure for the solution of the equation set that is obtained from the physical models of the fabrication process. The package can be broken down into a set of subunits. Each one was developed separately and although a set of

common nodal points are used the models use different algorithms. It is inappropriate to detail these models in this paper and those interested are referred to [2] [3].

2. EQUIPS: THE MODELS

Although EQUIPS is based on finite element algorithms, the coefficients and the physical models of processes follow closely the well characterised ones such as those in SUPREM II and III. This approach has the advantage that the results given by EQUIPS very closely match those that users have found acceptable in the past. A further advantage here is that the process specific parameters for SUPREM models have probably been calibrated and can be readily used in EQUIPS. EQUIPS also offers the user of choice of model so that other physical models can be benchmarked against the existing ones.

2.1 IMPLANTATION

Implantation is simulated in EQUIPS using a multiplicative function of vertical and horizontal distributions. The horizontal model can be selected from a Gaussian, joint half-Gaussian or Pearson type IV. The distribution in the horizontal dimension follows a model proposed by Selberherr et al [1] which in essence is a curve fit to data obtained using Monte Carlo simulation. The results from a typical implant step are presented in Figures 1, 2 and 3. Here a simulation region 0.5μm wide by 0.4μm deep is depicted, a vertical edge Si_3N_4 mask covers the right hand half and a dose of 10^{16} As was implanted at 100 keV.

EQUIPS automatically edits the solution mesh to optimise the positioning of the solution points to pick out important features by iteratively implanting and using the nodal values of concentration to decide if additional nodes are required.

2.2 DIFFUSION

The diffusion model uses an implicit backward difference solution for the 2-D diffusion equations for a single species. Like other packages the diffusion coefficient is concentration dependent and if desired the electric field at each point can be calculated to give a field enhanced term by invoking empirical models. An example of the use of the diffusion model is discussed in section 5.

2.3 OXIDATION

Oxidation is modelled by a solution for the steady state diffusion of oxygen in $Si0_2$ which then reacts at the oxide/silicon interface with the subsequent addition of oxide at the interface. The extra volume of oxide created causes the solution domain to deform as does the solution mesh. To maintain compatibility with the Deal-Grove model that is well calibrated the two-dimensional numerical model oxide thickness is scaled to conform with Deal-Grove predictions in unmasked regions well away from mask edges. EQUIPS consequently predicts growth of the same magnitude as SUPREM but with a good approximation to the shape of the oxide in the vicinity of the mask edges. Figures 4 and 5 show plots of the finite element mesh before and after the simulation of an oxidation step.

3. USER INTERFACE

A significant portion of the code within EQUIPS is devoted to the user interface and the graphical presentation of the result of simulations. The package can cope with a large range of devices. It interfaces particularily well in an environment when two screens(or a device that can simulate two screens) are available. This separates the character input/output from the graphical input/output and gives a very user friendly appearance, providing a continously updated display of the status of the package operation. Figure 6 illustrates the appearance of some of the EQUIPS command menus.

The package EQUGS was created specifically to complement EQUIPS and the implementation allows users to add device drivers for any graphics terminal with very little programming effort.

4. MESH EDITOR AND REFINEMENT

Since EQUIPS uses the FE method optimisation of the mesh is a critical factor in achieving good accuracy and performance. In order to allow the mesh to be readily modified EQUIPS uses a rectangular grid that can be subdivided by allowing a nodal point to be located at the centre of each side as well as on the quadrilateral vertices. A triangular mesh is then superimposed automatically for the calculation stage. The retention of the "square" mesh permits an editing operation. Nodal points and hence elements, can be added or removed without destroying the continuity of so that the triangular mesh. Refining the mesh can consequently be accomplished, the greatest concentration of nodal points can be made to effectively follow the dopant during processing. There is some automation of the refinement process which reduces the need for user intervention but to obtain the most efficient mesh, say for a large problem of typically 1200 nodes some user interaction is desirable. The interactive mesh editor is easy to use and EQUGS provides drivers for the "Graphical Input" devices on most terminals. A device that has mouse, joystick or tablet gives very rapid editing.

5. BENCHMARKING

EQUIPS has been benchmarked against the programs SUPREM II, SUPREM III and ICECREM as well as on the Edinburgh Microfabrication Facility 6 μm NMOS process. Tables 1, 2 and 3 summarises some observations.

Benchmarking has also been carried out against a standard problem proposed by GEC. The diffusion of As at 950°C for 5 hr. after an initial implant. of 10^{16} with $\sigma_{x,y}$ = 0.05 and r_p = 0.25. A comparison is presented in figures 7 to 15. It can be seen that EQUIPS is in good agreement with the benchmark problem. From Table 1 it can be seen

at there is reasonable agreement between the models and experiment. The low value of e source/drain oxide thickness produced by EQUIPS can be easily corrected by adjusting e linear rate-constant as this oxidation takes place over a heavily phosphorus doped sub-rate. This was in fact implemented in the case of SUPREM II but had not yet been cluded in the present version of EQUIPS.

. CONCLUDING REMARKS

EQUIPS appears to fill a gap in the packages available to the device and process ngineer. It operates interactively with very user friendly menu type commands that equire little user training. In tests the accuracy has proved to be as good as the SUPREM ackages while also offering simulation in two-dimensions. If necessary, EQUIPS can be sed in a batch mode whenever a larger number of solution points are required for greater ccuracy.

EFERENCES

1] Hobler, G., Lauger, E. and Selberherr, S. "Two-dimensional modelling of ion implantation", Proc. SISDEPSG, Swansea, July 1986.

2] Ferguson, R.S. and Doherty, J.G. "An interactive two-dimensional finite element process modelling package for a single user mini-computer", Solid State Electronics, Vol 27, No. 12, pp 1043-1054, 1984.

3] Armstrong, G.A, Ferguson, R.S. and Flynn, J. "A two-dimensional finite element algorithm for the simultaneous solution of the semiconductor device equations with automatic convergence", IJNME, Vol 23, pp 635-649, 1986.

Oxide	EQUIPS	SUPREM II	SUPREM 3	ICECREM	EXPERIMENT
Initial Oxide	470A	565A	594A	544A	550A
Gate Oxide	730A	842A	914A	836A	850A
Source-Drain	1150A	1934A			
Field Oxide	14,510A	14,500A	16,300A	15,400A	14,500A

Table 1 : Comparison of oxide thickness

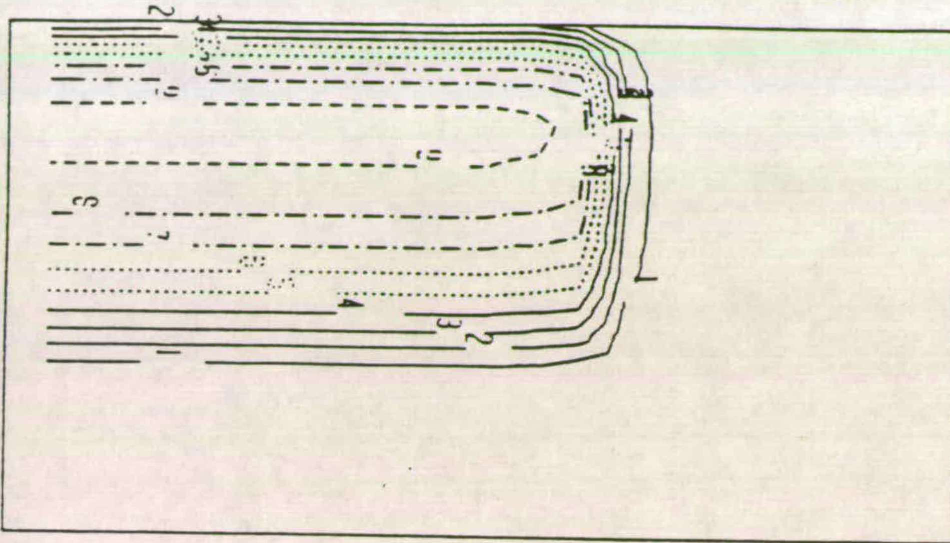
X_j	EQUIPS	SUPREM II	SUPREM 3	ICECREM	EXPERIMENT
Depletion region	0.28μm	0.26μm	0.23μm	0.23μm	
Source-Drain	1.39μm	1.7μm	1.37μm	1.39μm	1.5μm

Table 2 : Comparison of junction depth

	Resistivity (Ω/□)				
Program	EQUIPS	SUPREM II	SUPREM 3	ICECREM	EXPERIMENT
Enhancement Region	2.15×10^4	3.52×10^4	3.97×10^4	2.39×10^4	
Depletion Region	8,810	5,490	4,4140	6,440	
Source-Region	9.0	17.4	11.1	15.0	11.0

Table 3 : Comparison of sheet resistivity

CONTOUR PLOT OF IMPURITY CONCENTRATION



CONTOUR VALUES

1 1.00E 13
2 1.00E 14
3 1.00E 15
4 0.10E 17
5 0.10E 18
6 0.10E 19
7 0.10E 20
8 0.10E 21
9 0.10E 22

FIGURE 1: Contour plot of concentration after implant of 1e16 As at 100 keV into half masked Si substrate 0.5 microns wide and 0.4 microns deep.

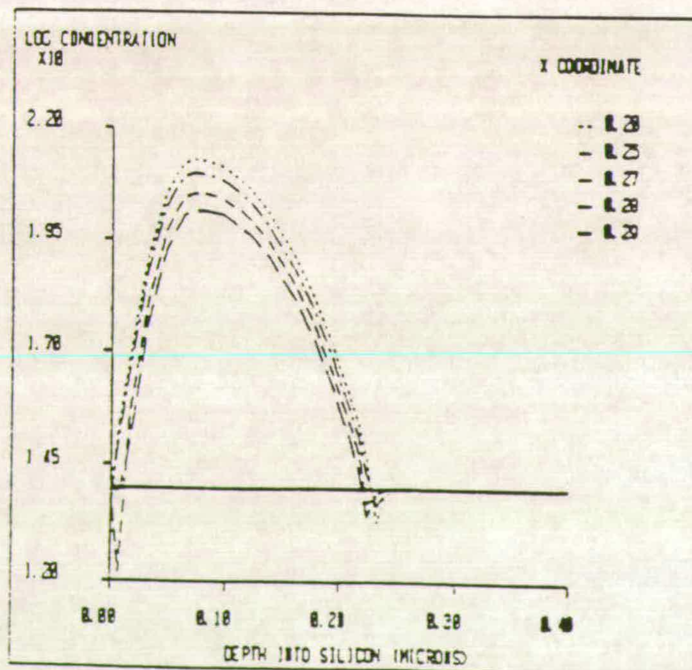


FIGURE 2: Concentration in vertical sections through substrate, after implant described in Fig. 6.1.

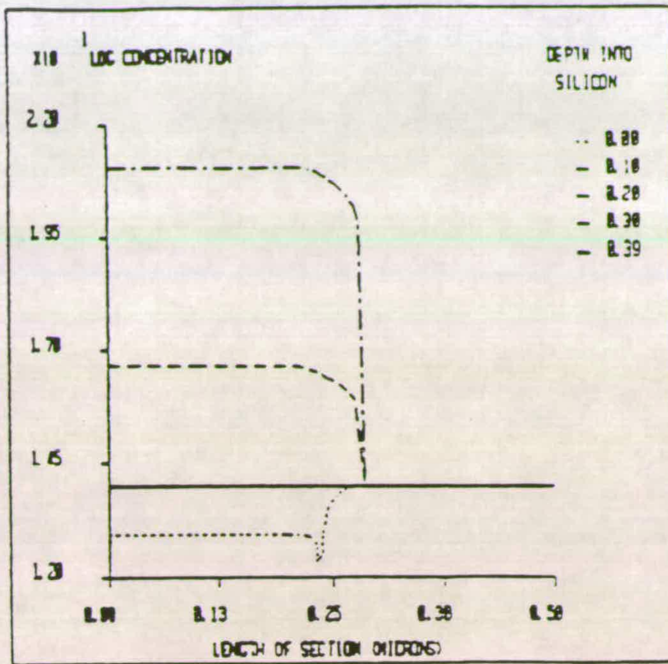


FIGURE 3: Concentration in horizontal sections through substrate, after implant desribed in Fig. 6.

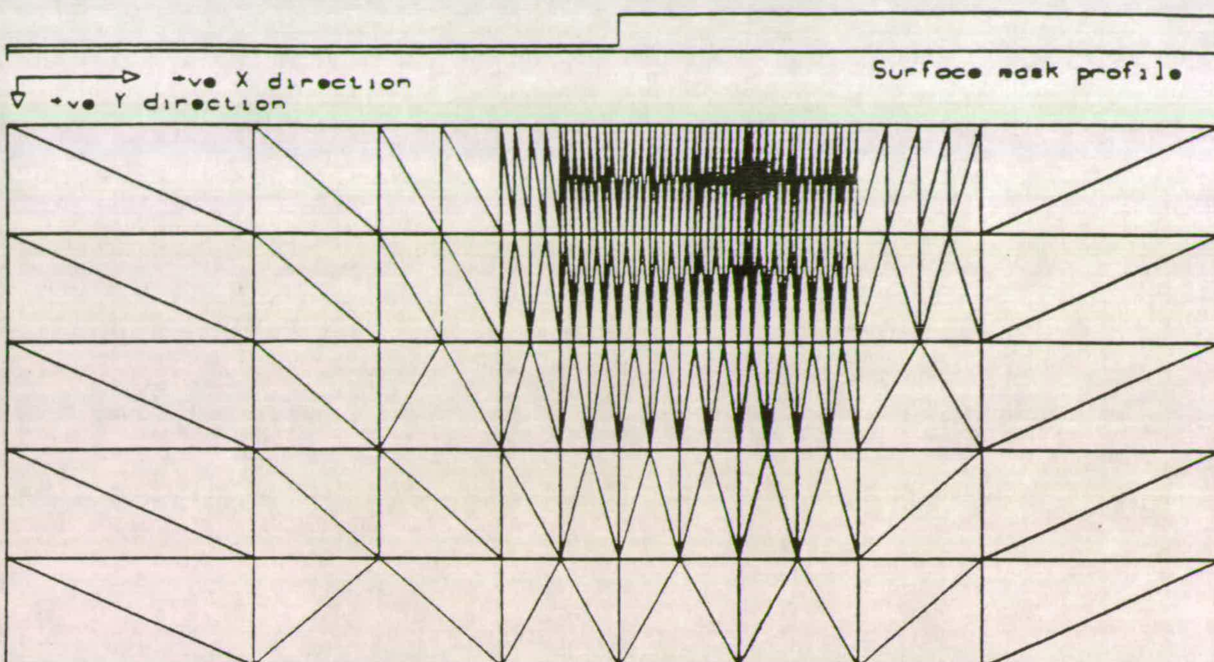


FIGURE 4: E.Q.U.I.P.S. Mesh before simulation of oxidation step.

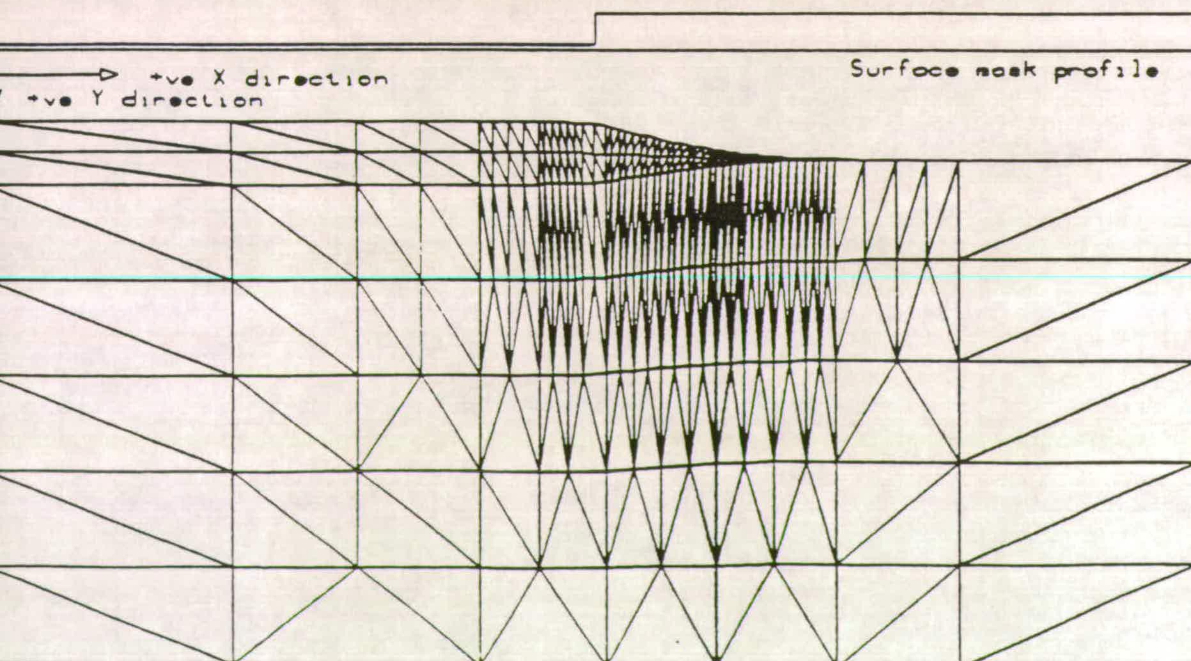


FIGURE 5: Mesh after completion of oxidation step.
Wet oxide 60 mins 1050 C.

```
HH:MM:SS              E.Q.U.I.P.S.  1.5          DD-MMM-YY

                    [ PROGRAM COMMAND MENU ]

IMPlantation       PREdeposition      P-Square mesh     P-Tri. mesh
REDuce             PLOt               SAVe              RE-start
STOp               OXIdation          DRIve-in          PATtern-mask
EDIt-mesh          ETCh               PRInt-con.        METalize
SCAle mesh         HIStory            REFine            P-Level
HELp               EXTract            DEPosition        NEW-plotter
STArt simulation

                           COMMAND >

CPU 0.0 s | NODES  0 | ELEMENTS  0 | WIDTH 0 um | DEPTH 0 um
```

FIGURE 6(a): Main menu.

```
                                                  VT220 Terminal

22-SEP-87              E.Q.U.I.P.S.  1.8             10:00:18

              IMPLANTATION COMMAND OPTIONS

LANT IMPURITY (B) (P) (As) (Sb) (Return - to main menu)
   Dose (cm-2)
   !       Energy (keV)
   !       !       Model (G)auss (P)earson (S)UPREM 3
   !       !       !       Rp Si
   !       !       !       !       Rp SiO2
   !       !       !       !       !       DeltaRp Si
   !       !       !       !       !       !       DeltaRp  SiO2
   !       !       !       !       !       !       !
   1E13    100     P

PU   5.0 s | NODES   176 | ELEMENTS   300 | WIDTH  5.00 um | DEPTH  2.00 um
```

FIGURE 6(b): Implantation data

```
VT220 Terminal

 22-SEP-87          E.Q.U.I.P.S.   1.8                 10:03:58

mperature
    Step time (mins)
    !         Diffusion model (0-3)
    !         !          Ramp up time
    !         !          !         Ramp down time
    !         !          !         !         Oxidation model (W)et (D)ry (S)team
    !         !          !         !         !         Pressure (atm)
    !         !          !         !         !         !
050 1         3                              W         0.5

CPU   10.7 s | NODES   176 | ELEMENTS   300 | WIDTH  5.00 um | DEPTH  2.00 um
```

FIGURE 6(e): Oxidation data

```
VT220 Terminal

 22-SEP-87          E.Q.U.I.P.S.   1.8                 10:06:15
```

Impurity	Step time	Time	Temperature
B	1.00 m	1.00 m	1050.00 C°

```
CPU   17.4 s | NODES   198 | ELEMENTS   340 | WIDTH  5.00 um | DEPTH  2.00 um
```

FIGURE 6(f): Pre-dep processing information

VT220 Terminal

22-SEP-87 E.Q.U.I.P.S. 1.8 10:27:34

[PLOT OPTION MENU]

VER. SECT.	HOR. SECT.	ISO. PROJ.	CONTOUR
PROJECTION	NODes and bdys.	SQUare mesh	TRIangular mesh
NEW graf device	QUIT(end plot)		

OPTION >

CPU 21.9 s	NODES 198	ELEMENTS 340	WIDTH 5.00 um	DEPTH 2.00 um

FIGURE 6(c): Plot menu

VT220 Terminal

22-SEP-87 E.Q.U.I.P.S. 1.8 10:33:24

IMPURITY CHARGE IN EACH LAYER (pC um-1)

LAYER	B	As
OXIDE	1.916E-06	6.360E-10
SILICON	5.165E-04	1.714E-07
NITRIDE	0.000E+00	0.000E+00
POLY SILICON	0.000E+00	0.000E+00
CHARGE SUM	5.184E-04	1.721E-07

PRESS RETURN

CPU 44.3 s	NODES 198	ELEMENTS 340	WIDTH 5.00 um	DEPTH 2.00 um

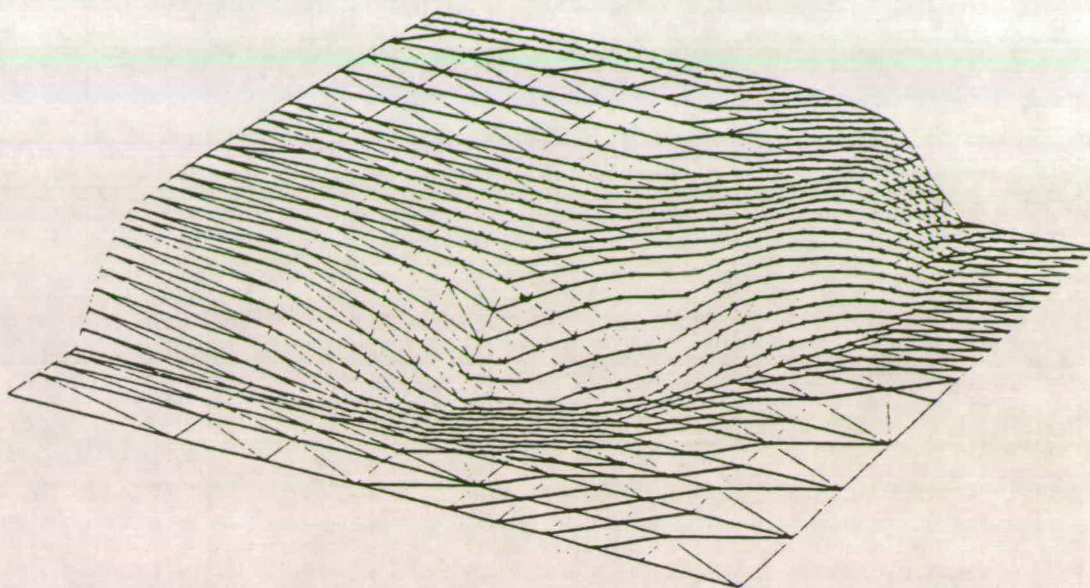


FIGURE 7: G.E.C. benchmark. Implant on an E.Q.U.I.P.S. mesh.

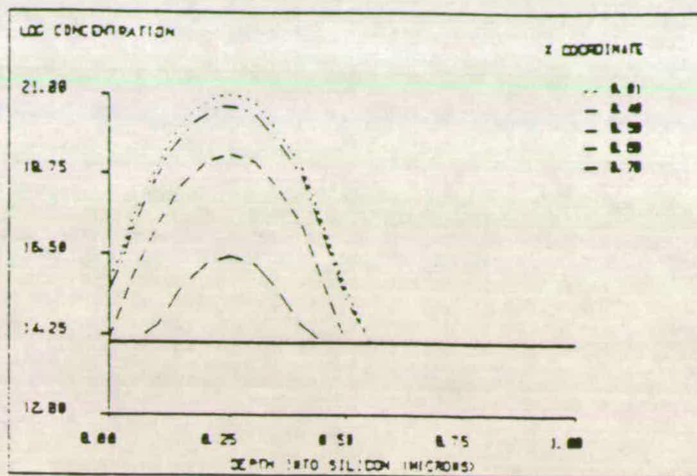


FIGURE 8: G.E.C. benchmark.
Vertical sections through implant of Figure 7.

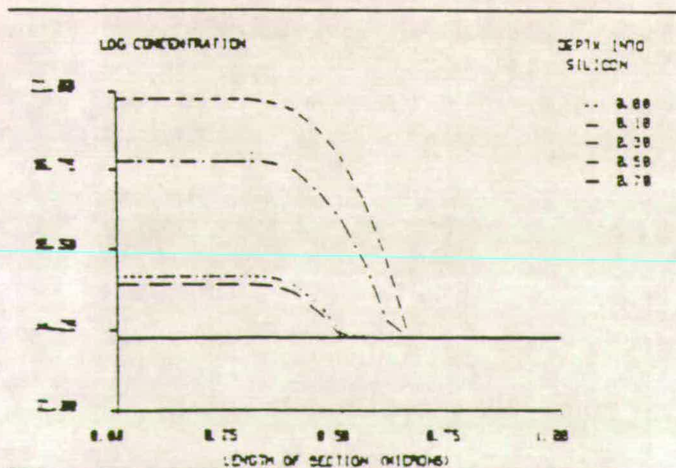


FIGURE 9: G.E.C. benchmark. Horizontal sections through implant of Figure 7.

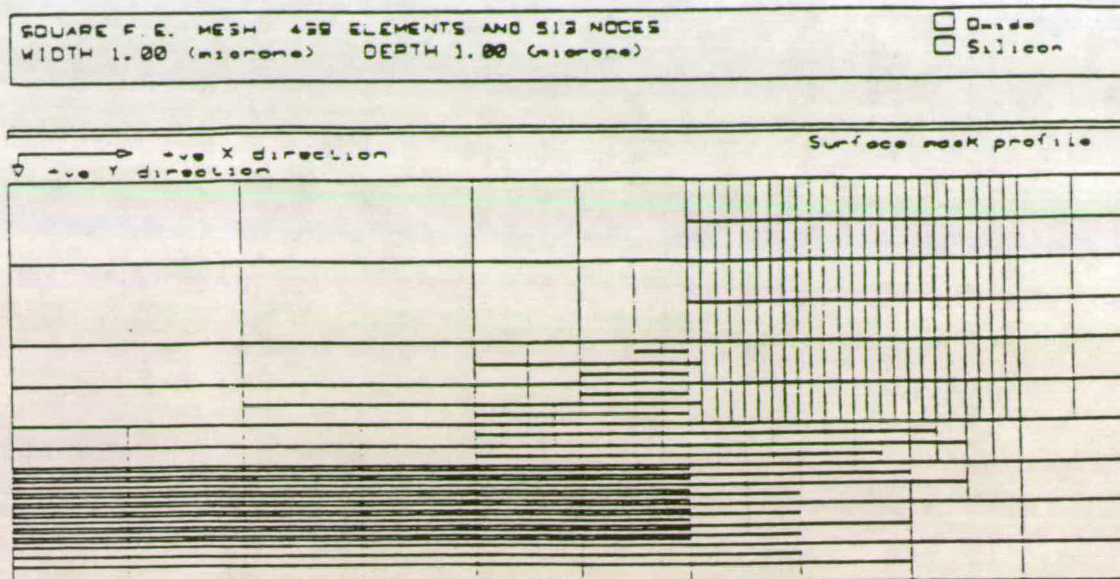


FIGURE 10: G.E.C. benchmark. Mesh used to produce the diffusion solution.

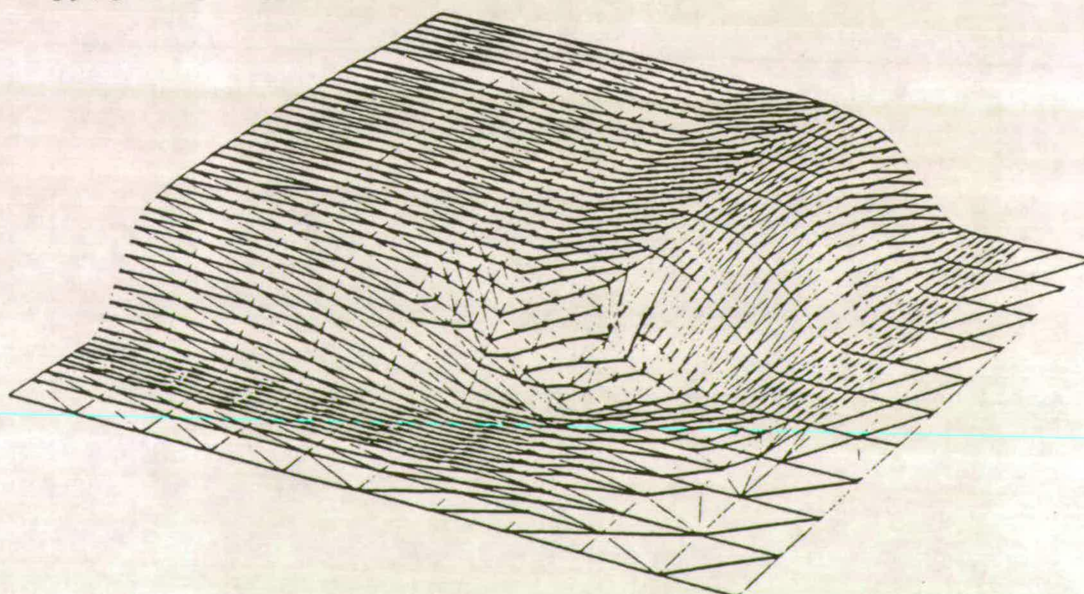


FIGURE 11: G.E.C. benchmark. Diffusion of implant. Projection plot.

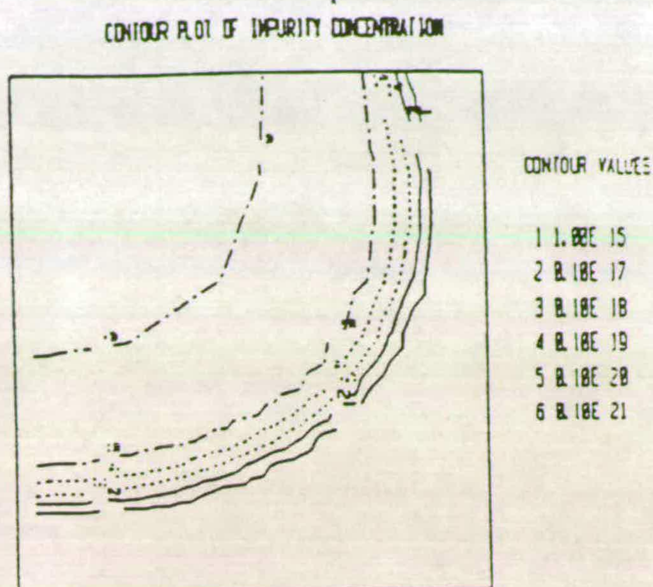


FIGURE 12: G.E.C. benchmark. Diffusion of implant. Contour plot from E.Q.U.I.P.S.

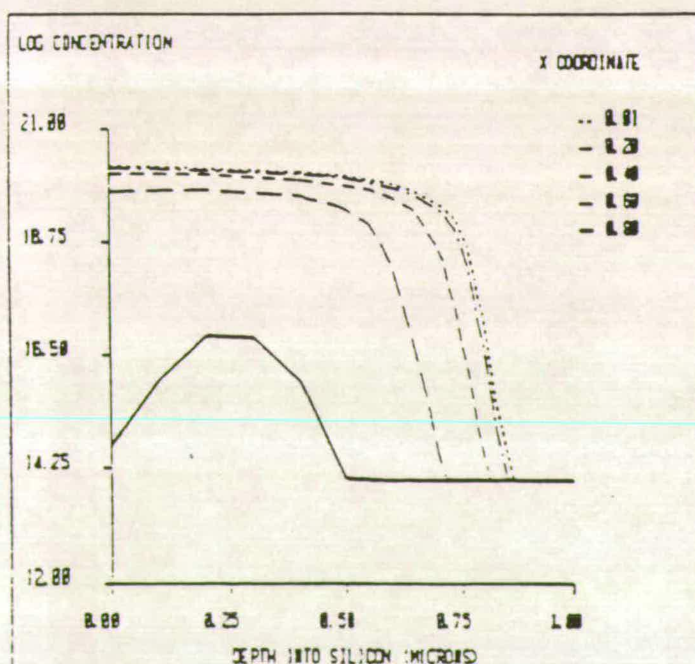


FIGURE 13: G.E.C. benchmark.
VERTICAL sections through diffused profile of Figures 11 and 12.

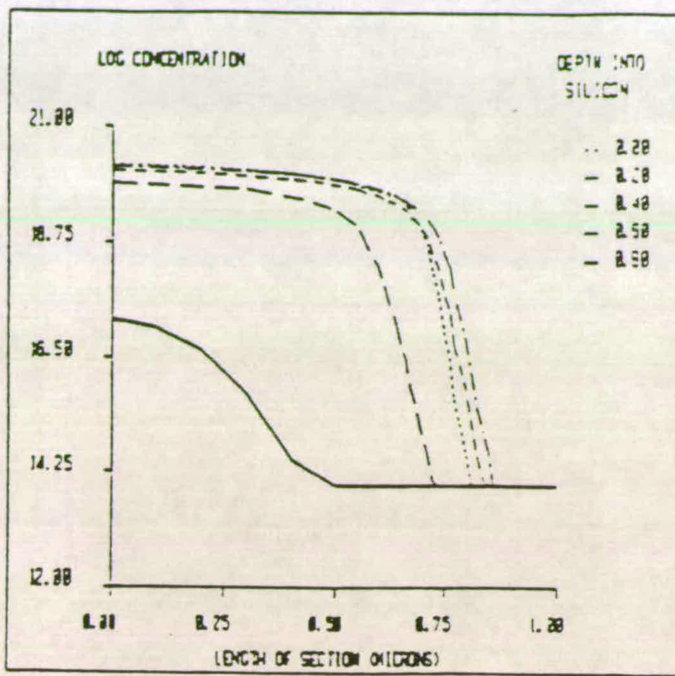
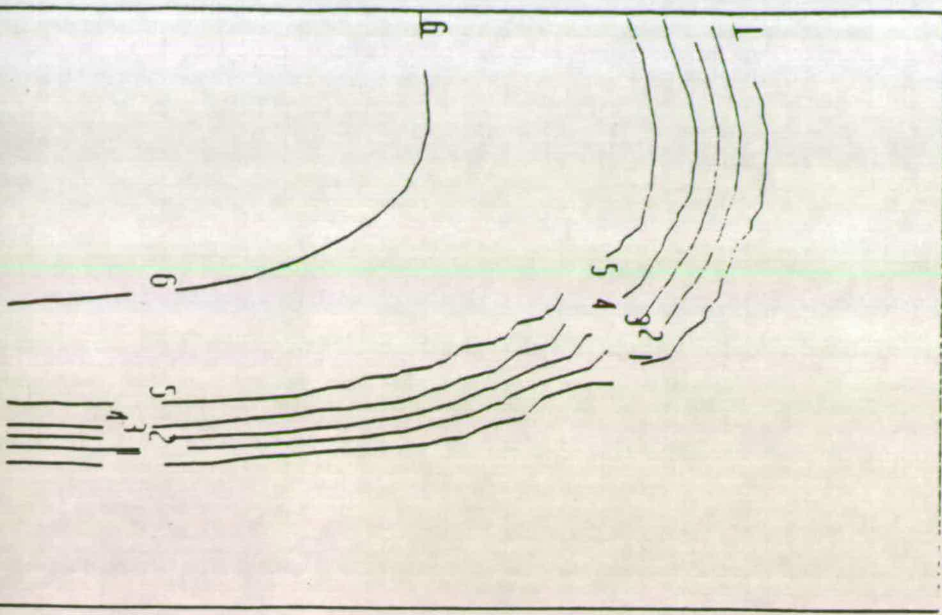


FIGURE 14: G.E.C. benchmark.
HORIZONTAL sections through diffused profile of Figures 11 and 12.

CONTOUR PLOT OF IMPURITY CONCENTRATION

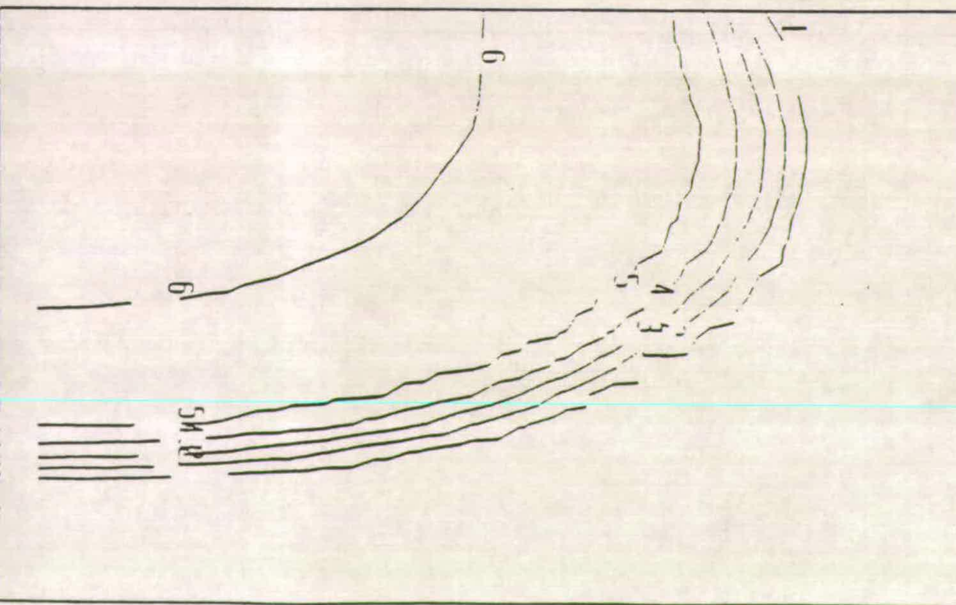


CONTOUR VALUES

1 1.00E 15
2 0.10E 17
3 0.10E 18
4 0.10E 19
5 0.10E 20
6 0.10E 21

FIGURE 15: Contour plot of GEC benchmark

CONTOUR PLOT OF IMPURITY CONCENTRATION



CONTOUR VALUES

1 1.00E 15
2 0.10E 17
3 0.10E 18
4 0.10E 19
5 0.10E 20
6 0.10E 21

FIGURE 16: Contour plot of GEC benchmark problem from E.Q.U.I.P.S simulation.

INTERACTIVE PROCESS SIMULATOR (EQUIPS)

by

R.S. Ferguson

Queens University Belfast

and

A.J. Walton and M. Fallon

Edinburgh Microfabrication Facility
University of Edinburgh

SUMMARY

EQUIPS is a two dimensional interactive process simulator that is designed to execute rapidly on any computer system with 1 Mbyte of memory and 5 Mbytes of disk storage. The source is standard FORTRAN 77 and includes graphics software for 15 types of terminals and plotters. The process steps that can be modelled are IMPLANTATION, PREDEPOSITION, DIFFUSION, OXIDATION, ETCHING and LAYER DEPOSITION. A non-system dependent batching procedure is also included. The solution of the modelling equations is performed by a self consistent algorithm based on the finite element method with a choice of implicit or explicit transient solution techniques. Oxidation is modelled by allowing the mesh to automatically adapt so as to accommodate the volumetric expansion which takes place as the silicon is oxidised.

ABSTRACT

EQUIPS is a menu driven interactive process model for simulating the fabrication of silicon integrated circuits. The finite element method is employed to solve the partial differential equations in two dimensions.

Experience in the field of device modelling has shown that the finite element method requires a very carefully designed mesh if full advantage to be taken of the technique. EQUIPS consequently uses a sophisticated interactive graphical mesh editor that allows the user to quickly prepare a suitable mesh and adjust it appropriately for different phases of the solution. The mesh editor may also be used for automatic mesh refinement as the simulation proceeds. This can be applied during each iterative cycle in the solution but it is time consuming and is not recommended when the program is operated in the interactive mode.

Although EQUIPS is primarily intended for interactive use, it is possible to execute particular process steps in a batch mode if, for example, a higher accuracy is required. This is accomplished without having to rely on system dependent features of the computer. The user requests that the step be executed off line, and the program then sets a flag file and finishes. The user then simply enters a background process and later, when EQUIPS is restarted the initial state is that at the end of the batch step.

Any program which is interactive must present results in such a manner which enables them to be easily interpreted and EQUIPS gives the option of contour and surface plots as well as 1-D sections. To enhance its portability the program interfaces directly with EQUGS which is a graphics library specifically written for EQUIPS. This is capable of driving 15 types of terminals and plotters and the software has been designed to allow the addition of extra devices with minimal effort.

The algorithm upon which EQUIPS is based is straightforward. The non-linear equations are linearised by using short time steps and solving for the diffusion of each impurity separately. The effect of coupling is taken into account by calculating the electric field due to carrier impurities and using this to enhance or retard diffusion. A maximum number of four dopants can be present during the simulation.

Oxidation is modelled by including in the algorithm time step loop a model (steady state Poisson problem with Neumann and mixed boundaries) that predicts the amount of oxide grown at the Si/SiO_2 interface. The finite element mesh is then deformed to allow for

the volumetric expansion and the consequent redistribution of impurities at the interface calculated.

EQUIPS is primarily being used on a MICROVAX II system and give very reasonable response times for all stages with the possible exception of the diffusion and oxidation steps if a large mesh is used (1000 nodes 1500 elements). However, with careful mesh design 500 nodes and 750 elements will give a very good covering domain and result in processing times commensurate with interactive use.

The user can create a default file that allows model parameters to be modified so that the simulation can be tailored to virtually any specific process. With EQUIPS being interactive, its elegance can only be fully demonstrated through use but the following examples give an illustration of some of its capabilities. Figure 1 shows the deformed mesh which results from an oxidation step. It should be noted that in the original mesh (designed using the interactive mesh editor) the oxide was modelled by only four elements in its width. Figure 2 illustrates a simple mesh for which is used for the implant of figure 3. The resulting concentration is shown as a contour plot in figure 3. Figure 4 shows the mesh created by using the automatic mesh refinement procedure after the implant of figure 3.

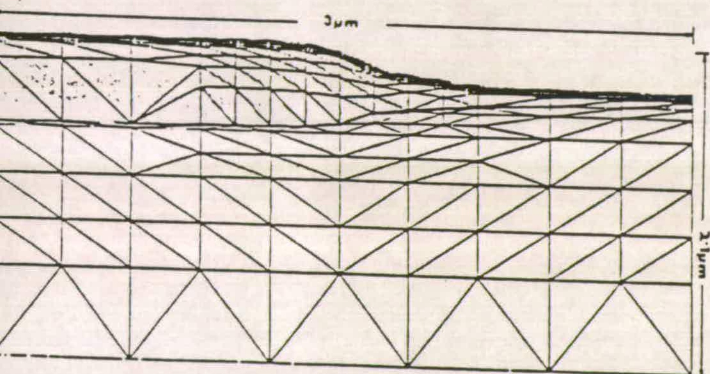


Figure 1. Mesh deformation after oxidation.

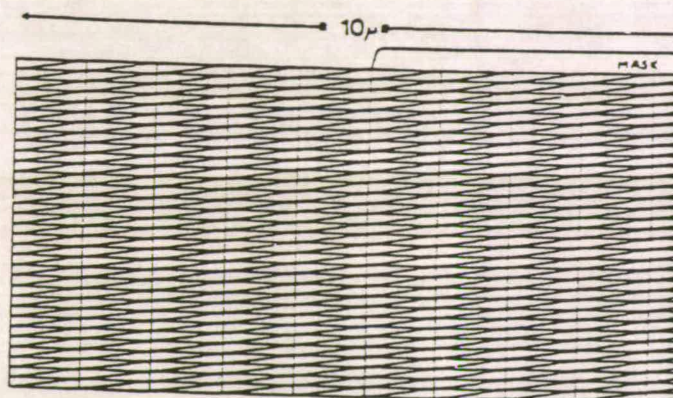


Figure 2. Initial mesh and mask used for the implant of figure 3.

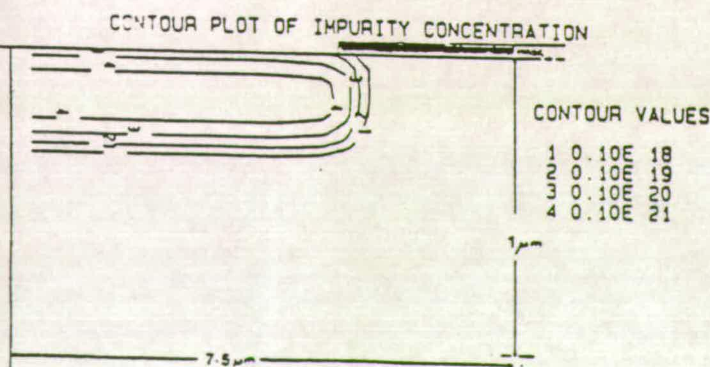


Figure 3. Implant profile for boron implant (10^{16} at 50 KeV).

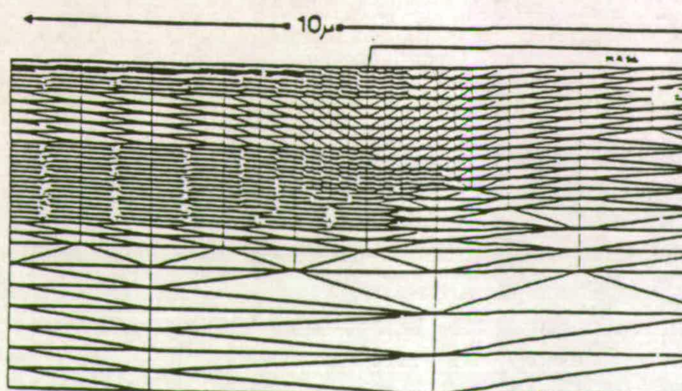


Figure 4. Resulting mesh after the implant of figure 3 using automatic refinement.

DEVELOPMENTS IN THE E.Q.U.I.P.S. PACKAGE 1986-1987

by

R. S. Ferguson, A. J. Walton and M. Fallon.

1.0 Introduction

We will present a paper detailling developments an changes introduced into the E.Q.U.I.P.S. package during the last 12 months, ie. since the last club meeting. We will stress changes in the models and changes in the user interface and will report some results on the verification of the simulations from E.Q.U.I.P.S. by comparison with one dimensional code and with the G.E.C. benchmark problem.

2.0 Changes to the E.Q.U.I.P.S. code.

2.1 The user interface.

2.1.1 E.Q.U.G.S. Graphics library

The graphics package that is required to run along with EQUIPS has also been updated to version 3.5, this again corrects some "bugs" and in particular has corrected the problem found when creating **long strings** of character output with routine CHAHOL and using the **hardware chartacter** facility of the display terminal. This version is also available on request. Other character handling routines CHAFIX etc. have been rewritten to make use of FORTRAN 77 character facilities.

2.1.2 User manual

A new version of the user manual has been prepared, this gives some example simulations of the use of EQUIPS as well as detail about executing the package.

2.1.3 Defaults for diffusion models.

Two default files can now be accessed by the user of EQUIPS SU2PAR.DAT and EQUPAR.DAT in these all the coefficients for the SUPREM and ICECREM like diffusion models can be changed by the user simply by editing the appropiate text.

2.2 Sheet resistivity calculation.

The calculations for sheet resistivity in EQUIPS versions 1.0 to 1.5 are known to be inaccurate, (ie they disagree with SUPREM). The reason for this is that the mobility used is not field dependent, this has been changed in EQUIPS and the corrected code are included in version, 1.6, along with the the correction of a few other minor "bugs" that do not affect the VAX/VMS version but may cause a problem on other systems.

2.3 The implantation model.

This has been completely rewritten. The first point to make here is that it is now designed to give exactly the same output as SUPREM II and III when those models are chosen. This means that for As P and Sb diffusion a joined half gaussian is used and for Pearson IV models an empirical exponential tail is added. The model, of course only looks like SUPREM if comparison is done in one dimension. For example simulating the implantation of As 1E16 100 keV into a Si substrate doped with Boron. The peaks are reported at depths of 0.110 μm and 0.112 μm respectively withs values of 4.19E21 and 4.18E21. The effect of mask edges and lateral scattering is simulated by the incorporation of a new model based on work reported by Hobler, Langer and Selberherr. (Swansea-1986) This model consists of a sophisticated fitting procedure for lateral standard

deviation and straggle which is dependent on the vertical position as well as the energy of the implanted ions. The model was derived by fitting data obtained from monte-carlo simulations. It appears to give a much more accurate simulation of the two-dimensional implantation (and is in fact faster than the previous model). Figure 1 shows a contour plot of the above As implant adjacent to a mask edge, figures 2 and 3 show vertical and horizontal sections through the implant.

2.3 The Diffusion model.

A new diffusion model "model 2" has been written to give the user the opportunity to use a SUPREM II like model, as mentioned above the parameters for this model can be changed by editing the default file. The model has been shown to work well diring verification, particularily for the diffusion of As.

2.4 The Oxidation model

The old oxidation model was not particularily accurate in predicting oxide thicknesses. The new model uses an approach of combining the accuracy of a one-dimensional approach th calculate the oxide thichness well away from a mask edge coupled to the two-dimensional model to predict the shape of the oxide layer adjacent to a mask edge.

3.0 Verification and evaluation

3.1 G.E.C. Benchmark. The new SUPREM II diffusion model (model 2)in EQUIPS gives very good agreement with the GEC benchmark. Examination of figures 4-11 give a comparison of the two sets of results.

4.0 **Future development to the EQUIPS code.**

The experiments, that to date have not been very successful, to automate the mesh reduction and refinement during diffusion will continue since we feeel that most users do not want to use the interactive editing facility to create meshes for themselves at each stage of the process.

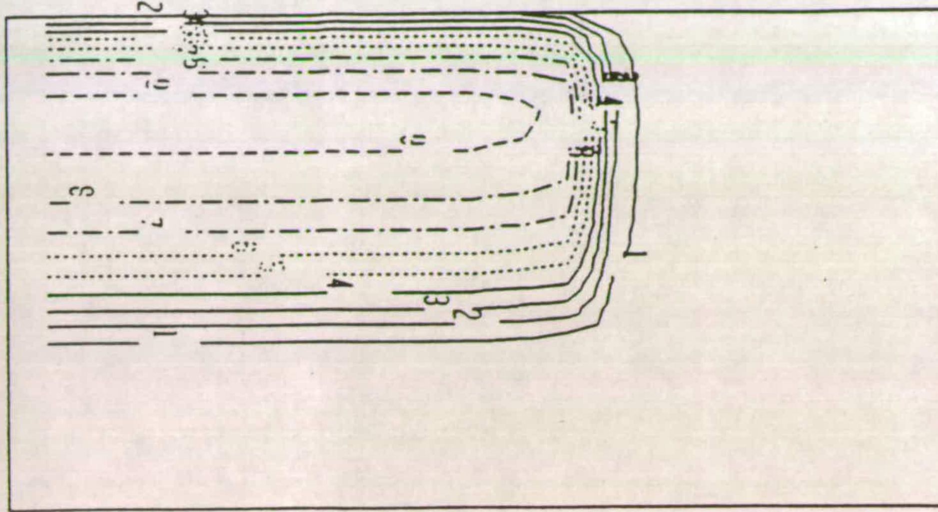


FIGURE 1 Contour plot of concentration after implant of 1e16 As at 100 keV into half masked Si substrate 0.5 microns wide and 0.4 microns deep.

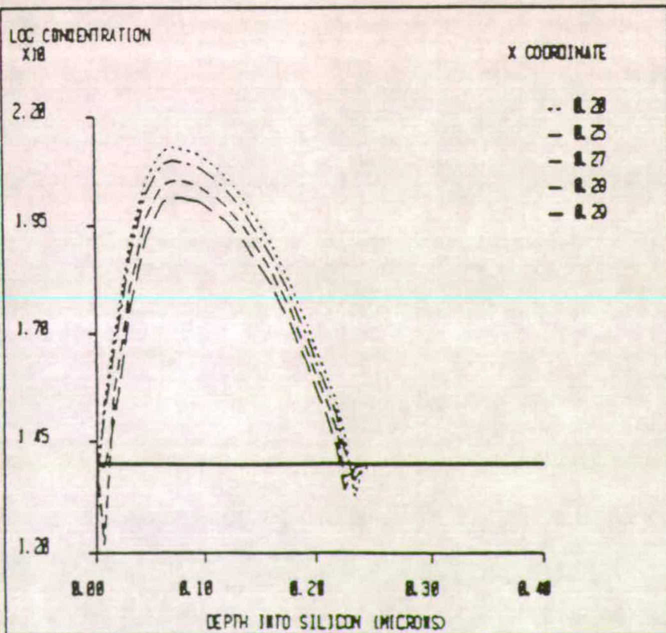


FIGURE 2 Concentration in vertical sections through substrate, after implant described in Fig. 6.1

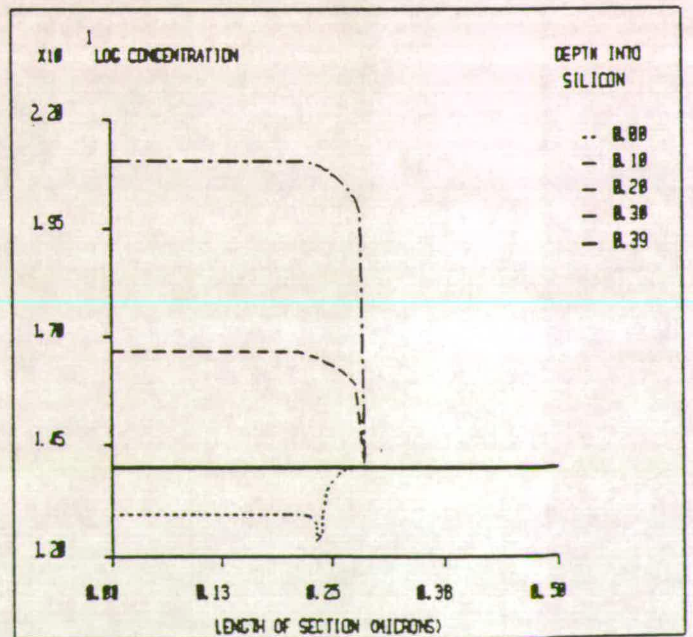


FIGURE 3 Concentration in horizontal sections through substrate, after implant described in Fig. 6.

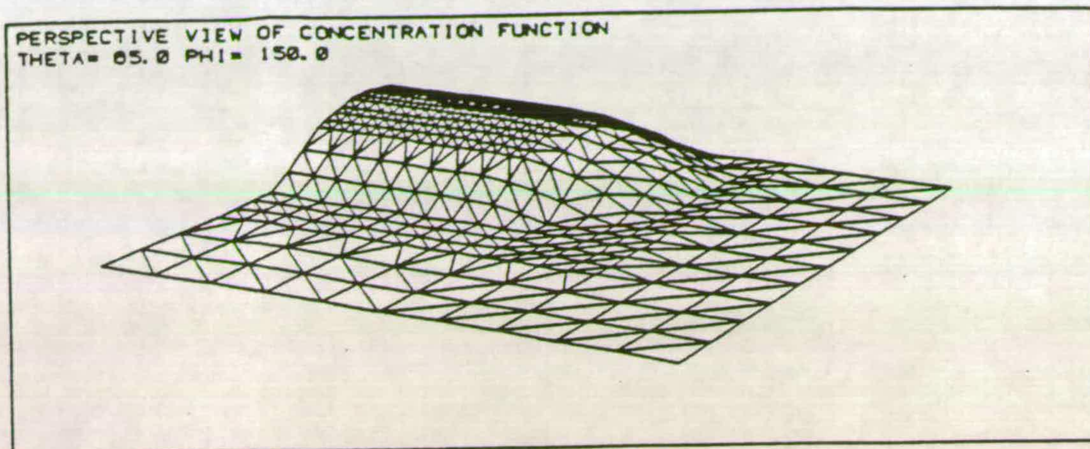


FIGURE 4 G.E.C. benchmark. Implant on an E.Q.U.I.P.S. mesh.

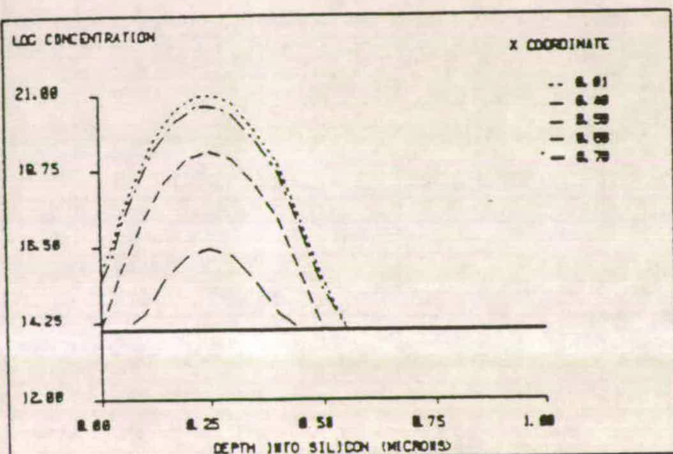


FIGURE 5 G.E.C. benchmark.
Vertical sections through implant
of Figure. 4

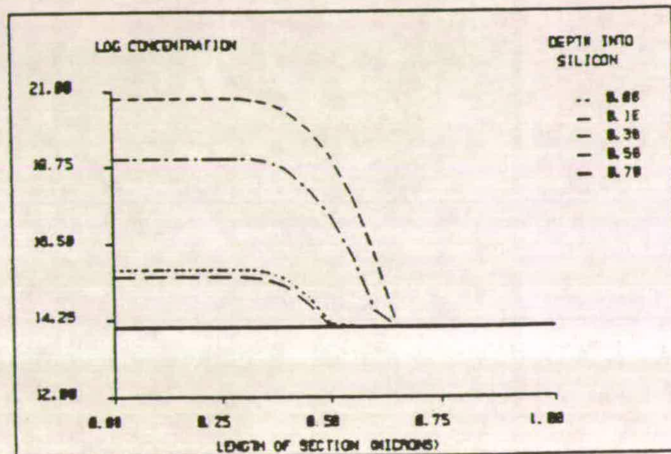


FIGURE 6 G.E.C. benchmark.
Horizontal sections through implant
of Figure. 4

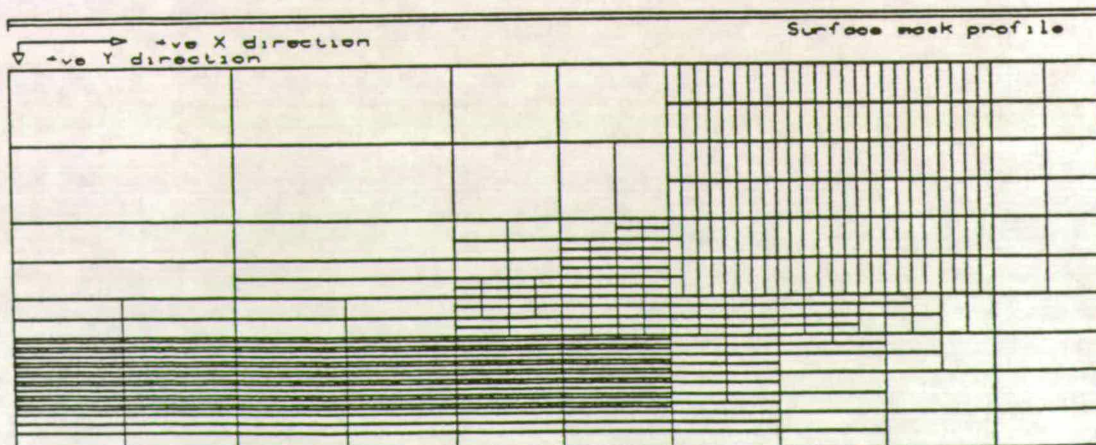


FIGURE 7 G.E.C. benchmark. Mesh used to produce the diffusion solution.

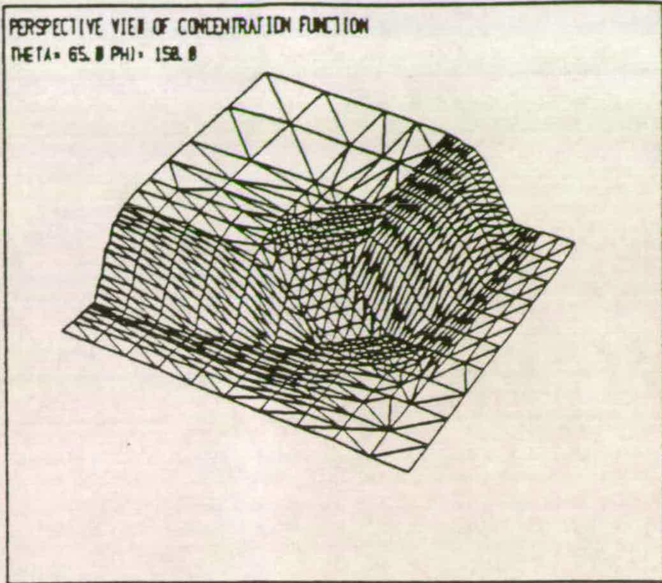


FIGURE 8 G.E.C. benchmark. Diffusion of implant. Projection plot.

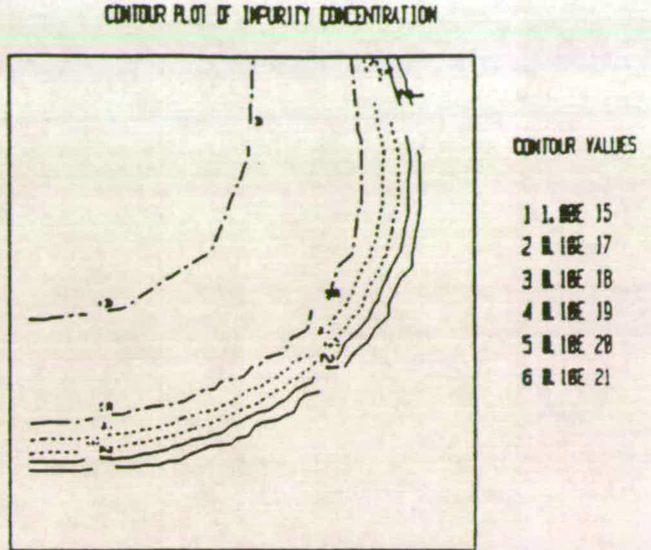


FIGURE 9 G.E.C. benchmark. Diffusion of implant. Contour plot.

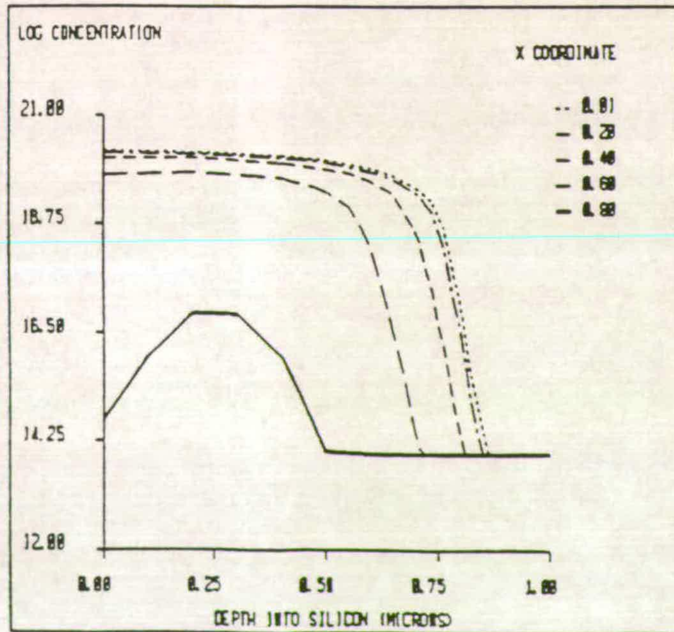


FIGURE 10 G.E.C. benchmark. VERTICAL sections through diffused profile of Figures 8 and 9

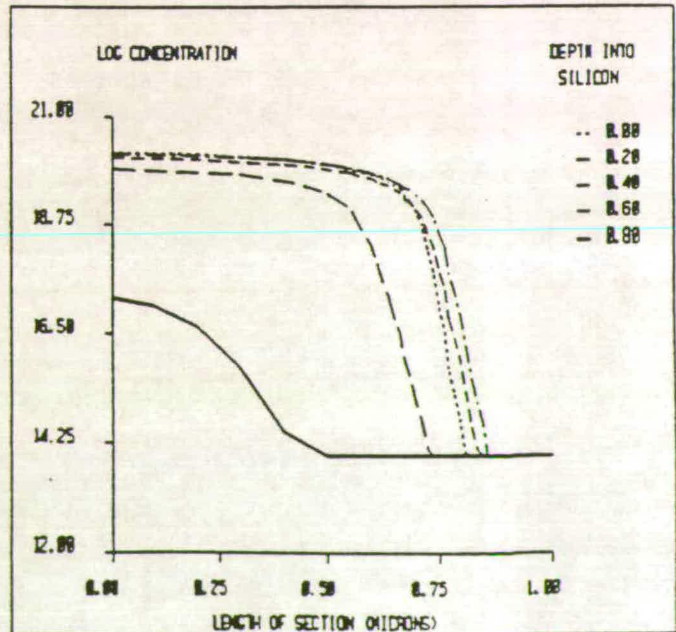


FIGURE 11 G.E.C. benchmark. HORIZONTAL sections through diffused profile of Figures 8 and 9