Modelling, Fabrication and Characterisation of the EEPROM

Thesis submitted by

Anthony James Chester

for the degree of

Doctor of Philosophy

Department of Electrical Engineering

University of Edinburgh

Scotland

August 1993

Abstract

The Electrically Erasable Programmable ROM (EEPROM) is used in applications such as microcontrollers and mass storage media. Each of these markets is rapidly expanding. However, EEPROMs are particularly susceptible to reliability problems, since they must survive severe voltage and current stressing. This has a knock on effect, since operating speed must be reduced, to increase reliability. Motorola's implementation of the EEPROM is the Floating Gate Electron Tunnelling MOS (FETMOS), which has been adopted for study in this thesis.

An analytic model has been developed for the FETMOS, which encompasses transient response, threshold window and reliability. A good correlation has been shown between modelled data and experimental results, testifying to the model's accuracy. The effect of basic design parameters upon threshold window has been characterised, thus indicating how processing variations may be used to tailor the EEPROM threshold window. Equally, the model may be used to predict the effect of sizing down a circuit - this is important as integration densities escalate. Program endurance is the most pressing reliability issue. Modelling has indicated that, large improvements may be made in this by increasing the floating gate/drain overlap, with little effect on threshold window.

An novel experiment has then been devised to monitor the effect of floating gate/drain overlap and doping species, upon EEPROM reliability. For this, transistor arrays with a spectrum of well defined gate/drain offsets have been produced. The results of these are consistent with the model. It has also been found that the chemistry of the dopant has only a tangential effect upon reliability.

In conclusion, it is proposed that an increase in the tilt angle of the drain implant would be the most suitable method to increase the overlap, since no adjustment is then required in the thermal budget of the process. An increase in the drain doping density could also be used, to similar effect.

Declaration

I declare that the research documented in this thesis is original and entirely of my own making, except where it has been indicated to the contrary.

Anthony James Chester

Acknowledgements

I wish to express my thanks to all those people who helped in any way towards this PhD. In particular, I would like to thank my academic supervisor, Dr. A.J.Walton, both for offering his expert help and guidance throughout this Ph.D. program, and for allowing me autonomy in all my final decisions.

I would also like to thank my industrial supervisor, Paul Tuohy, for his excellent technical discussions, particularly with respect to EEPROM modelling.

Many thanks to all staff and students of the Edinburgh Microfabrication Facility, past and present, whose support has been greatly appreciated.

I conclude with an acknowledgement to the Science and Engineering Research Council and from Motorola, who provided financial assistance under a CASE award scheme.

Table of Contents

1. Introduction **1**

1.1 Floating Gate Technologies 3

1.1.1 The FETMOS Device 3

1.1.2 The FLOTOX Device 5

1.1.3 The Flash EEPROM 6

1.1.4 Contrast Between FETMOS/FLOTOX and Flash 8

1.2 Utility of the EEPROM 9

1.2.1 Program/Data Storage Media 9

1.2.2 Embedded Systems 10

1.2.3 Embryonic Technologies 10

1.3 Reliability 11

1.3.1 Shrinking Geometries 12

1.3.2 Reliability In General 13

1.3.3 Reliability and the EEPROM 14

1.4 Thesis Plan 14

2. EEPROM Physics and Reliability **18**

2.1 Fowler-Nordheim Tunnelling 18

2.1.1 Wave Particle Duality 18

2.1.2 Fowler-Nordheim Tunnelling in MOS Structures 20
2.1.3 Factors Which Affect Fowler-Nordheim Tunnelling 22
2.1.4 An Equation to Describe Fowler-Nordheim Tunnelling 25
2.2 Reliability Issues 26
2.2.1 Oxide Breakdown 26
2.2.2 Analysis of Reliability Data 31
2.2.3 Accelerated Testing 33
2.2.4 EEPROM Reliability Issues 34
2.3 Improving EEPROM Operation and Reliability 37

3. Derivation of a FETMOS Model **43**

3.1 Overview 43
3.2 Distribution of Injected Charge in a Capacitor System 44
3.3 Derivation of Equations to Describe the FETMOS Device 47
3.3.1 Equivalent Capacitive Circuit for the FETMOS Device . . . 47
3.3.2 Coupling Ratios 49
3.3.3 The Electric Field as a Function of Time 50
3.3.4 Threshold Voltage as a Function of Electric Field 56
3.3.5 Initial Electric Field 59
3.3.6 Summary of Equations 59
3.4 Calculation of FETMOS Parameters 61
3.4.1 Overview 61
3.4.2 Geometry of the FETMOS Device 62
3.4.3 Calculation of Capacitances 63
3.4.4 Measurement of Fowler-Nordheim Coefficients 64

3.4.5 Measurement of Threshold Voltage 67

3.4.6 Calculation of RC Time Constant, τ 69

3.5 Verification of the FETMOS Model Against Experimental Results . 70

3.5.1 Overview 70

3.5.2 Comparison of Model Predictions and Experimental Results 71

3.6 Conclusion 73

4. Analysis Using the FETMOS Model 76

4.1 Transient Analysis of the FETMOS Device 76

4.1.1 Threshold Window as a Function of Time 76

4.1.2 Electric Fields as a Function of Time 77

4.1.3 Current Densities as a Function of Time 78

4.1.4 Charge Densities as a Function of Time 79

4.2 A Methodology for Modelling EEPROM Endurance 80

4.3 Analysis of the Threshold Window and Endurance 84

4.3.1 Effect of Varying Floating Gate/Drain Overlap 84

4.3.2 Effect of Varying Effective Width 87

4.3.3 Effect of Varying Floating Gate Area 90

4.3.4 Effect of Varying Interlevel Oxide Thickness 93

4.3.5 Effect of Varying Floating Gate Length 96

4.3.6 Effect of Varying Gate Oxide Thickness 99

4.3.7 Effect of Varying Fowler-Nordheim Coefficients, A and B . . 102

4.4 Conclusion 107

5. Fabrication of EEPROM Structures **110**

5.1 The Progressional Offset Technique 110

5.2 Processing 112

5.2.1 Overview 112

5.2.2 The Bi-Poly Process 112

5.2.3 Growth of High Integrity Thin Oxide Films 115

5.2.4 Complete Process Flow 120

5.3 Circuit Design 131

5.3.1 Reticle Production 131

5.3.2 Transistor Design 131

5.3.3 Interconnection Scheme 135

5.3.4 Test Structures 136

5.4 Simulation 136

5.4.1 1 Dimensional Simulation of Arsenic Transistors 138

5.4.2 1 Dimensional Simulation of the Phosphorus Transistors . . 143

5.4.3 2 Dimensional Simulation of Transistors 146

5.5 Conclusion 148

6. Analysis of EEPROM Structures **152**

6.1 Process Quality 152

6.1.1 Wafer Yield 153

6.1.2 Gate Oxide Thickness 154

6.1.3 Threshold Voltage 154

6.1.4 Drain Characteristics 156

6.2 Assessment of POT Array Symmetry 156

6.3 Reliability Analysis 161

6.3.1 Test Methodology 161

6.3.2 Reliability of Progressional Offset Transistors 162

6.3.3 Average Charge to Breakdown 165

6.3.4 Calculation of Lateral Diffusion 167

6.3.5 Prediction of EEPROM Endurance from Experimental Results 169

6.4 Conclusion 171

7. Conclusion **175**

7.1 Modelling 176

7.1.1 Discussion of Results 176

7.1.2 Areas for Further Modelling 176

7.2 Experimental 179

7.2.1 Discussion of Results 179

7.2.2 Areas for Further Experimental Investigation 180

7.3 Concluding Remarks 182

Appendices

A. Summary of Paper i

B. Program in C to Model the EEPROM vii

C. Program to Measure Threshold Voltage xi

D. Process Simulation xvii

E. Program to Analyse POT Devices xx

Chapter 1

Introduction

The microelectronics industry ebbs and flows on a tide of technical innovation, particularly so in today's climate of economic aggression. As silicon processing technology becomes more sophisticated, so transistor geometries have shrunk and integration densities increased. The Dynamic Random Access Memory (DRAM) rides the crest of this technological wave, with the highest transistor count per chip. The quest for ever shrinking device geometries is lead primarily by insatiable thirst of the microcomputer for more memory. However, the DRAM does have a draw back in terms of its volatility, ie. it looses data when the power is turned off. For this reason magnetic storage media are required, although these have slow read/write and access times. The paradigm would be a *nonvolatile* semiconductor memory, and this is indeed a description of the Electrically Erasable Programmable Read Only Memory (EEPROM). The EEPROM is one of the more sophisticated types of semiconductor memories available, and has been adopted for study in this thesis. To recap, it offers:

1. Non-volatility. In common with magnetic storage media, the EEPROM retains data when the power supply is switched off.

2. Electrical programmability *and* erasability. As for the ubiquitous DRAM, data is updated electrically.

A broad spectrum of technologies may be included under the EEPROM banner. Amongst these are: MNOS devices (metal-nitride-oxide-semiconductor) for space

and military applications [1], devices using ferroelectric materials [2] [3], and novel micro-machined designs [4]. However, the most dominant technologies are the "Floating Gate" technologies, which are readily compatible with main stream MOS processes. This project will be limited solely to these devices. A generic floating gate EEPROM structure is illustrated in figure 1–1. This is similar to ordinary MOSFETs, but for the addition of a floating gate, which is electrically isolated. To program or erase the EEPROM charge is injected onto the floating gate, and (depending on its polarity) the underlying channel will go into either inversion or accumulation. Thus, by modulating the threshold voltage, a logic 1 or logic 0 is defined [1]. This is equivalent to the operation of EPROM. However, the EPROM must be illuminated in U.V. light for charge removal from the floating gate, whereas the EEPROM allows both program *and* erase electrically.

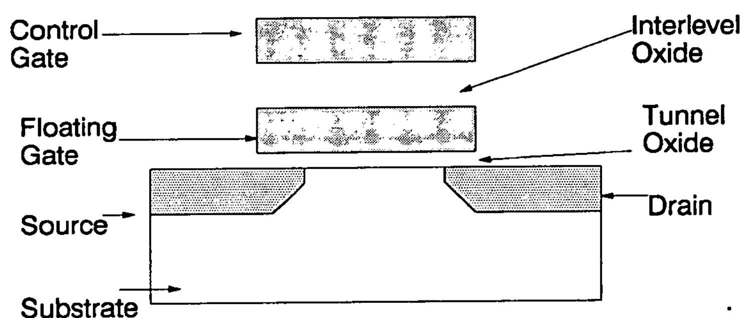

Figure 1–1: Cross Section Illustrating a Generic EEPROM Design.

[1]A programmed EEPROM may be defined as one with either positive or negative charge on the floating gate. In literature no rule has been adopted. In this thesis "programmed" describes an EEPROM with positive charge on the floating gate, and "erased" indicates injected electrons. This is irrespective of the nomenclature used in a referenced paper.

1.1 Floating Gate Technologies

Many designs have been realised using the floating gate [1] [5]. From these, three have been singled out for discussion.

1. The Floating Gate Electron Tunnelling MOS (FETMOS). This is the EEPROM fabricated by Motorola for their microcontrollers, at East Kilbride, and features most strongly in this work.

2. The Floating gate Tunnel Oxide (FLOTOX). This device is the most popular of the mature technologies [6] [7].

3. The flash EEPROM. These devices have a very "bright" future [8] [9].

Both FLOTOX and flash EEPROMs share features of the FETMOS [10]. Each technology will be described below, and contrasts will be made.

1.1.1 The FETMOS Device

A cross section of the FETMOS transistor is given in in figure 1–2 [11], the simplicity of which is evident. Table 1–1 summarises the program, erase and read voltages. The high voltages are usually produced "on chip" with charge pumping techniques [12].

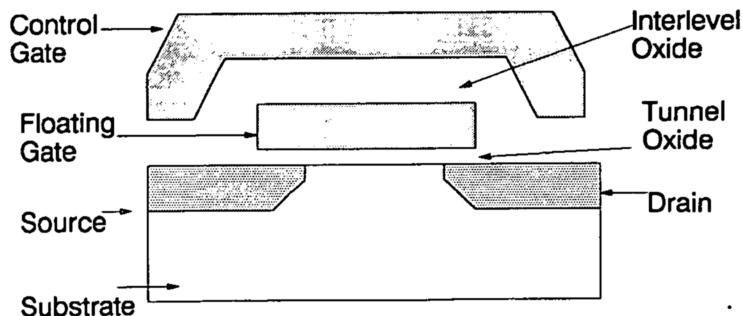

Figure 1–2: Cross Section of FETMOS Device. This is Realised in NMOS.

FETMOS Operational Voltages				
	Control Gate	Drain	Source	Substrate
Program	$0V$	$18V$	Floating	$0V$
Erase	$18V$	$0V$	$0V$	0V
Read	$0V$	$1V$	$0V$	$0V$

Table 1–1: FETMOS Operating Voltages.

For programming a positive potential is applied to the drain. Through capacitively coupling, an electric field is generated across the tunnel oxide [13]. Providing the field is high enough, electrons may then cross the tunnel oxide, from the floating gate to the drain. Electron transport is by Fowler-Nordheim tunnelling [11], which is discussed in more detail in chapter 2. Thus positive charge is left on the floating gate, which shifts the FETMOS threshold voltage negatively, to $\simeq -5V$ [14]. To ensure a large enough field $\simeq 7MVcm^{-1}$, a thin tunnel oxide of $110\AA$ is used. Notice that a transistor with a negative threshold will be effectively switched on. Hence the source is allowed to float, avoiding the onset of a channel current, which would otherwise retard the programming operation. Once sufficient charge has collected on the floating gate, the voltage across the tunnel oxide falls, and tunnelling will stop. Thus the process is self limiting.

Erase is a similar operation, but the electric field now falls between the substrate and floating gate, and electrons flow onto the floating gate. This gives a positive shift in the FETMOS threshold voltage, to $\simeq +5V$ [14]. Such a device is said to have a "Threshold Window" of $\pm 5V$.

To read stored data it is sufficient to apply $1V$ to the drain. Current flow (or absents) indicates the threshold voltage of the device, hence defining a logic 1 or 0. It would be possible to upset the threshold condition of the FETMOS during read, since the applied voltages could act as a "small program operation". This effect is called read disturb, and is minimised by limiting the read voltage to the lowest possible value [11]. One should also remember that, due to the thin gate

oxide, the drain voltage couples strongly to the floating gate. Thus the measured threshold voltage is sensitive to drain voltage.

In a complete memory cell, an enhancement select transistor is added in series with a FETMOS device. The select transistor provides cell isolation during read and allows program/erase of individual words of data.

1.1.2 The FLOTOX Device

A cross section of the FLOTOX transistor is given in in figure 1–3 [15], and operating voltages in table 1–2 [16] [7]. This is similar to the FETMOS, but the tunnel oxide of $\sim 110\AA$ is located above the drain.

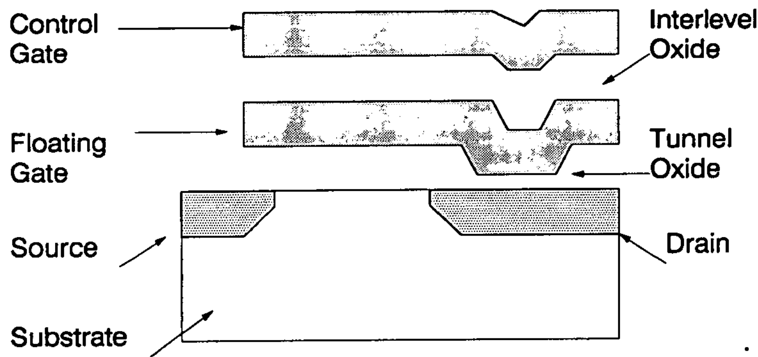

Figure 1–3: Cross Section of FLOTOX Device. This is Realised in NMOS.

FLOTOX Operational Voltages				
	Control Gate	Drain	Source	Substrate
Program	$0V$	$18V$	Floating	$0V$
Erase	$18V$	$0V$	$0V$	$0V$
Read	$0V$	$1V$	$0V$	$0V$

Table 1–2: FLOTOX Operational Voltages.

In programming, electrons pass from the floating gate to the drain by Fowler-Nordheim tunnelling, to give a negative threshold voltage shift of $\simeq -4V$ [16] [7].

For erase the electric field again falls between the drain and floating gate. However, electrons flow onto the floating gate, to leave a positive threshold voltage of $\simeq +8V$ [16].

A FLOTOX device is read in the same way as a FETMOS, and a select transistor is added to form a complete memory cell.

1.1.3 The Flash EEPROM

There are many suppliers of the flash EEPROM, and each has taken a slightly different approach to the device. The main suppliers are Intel, Seeq and Toshiba, all producing 1Mbit size arrays [9]. Figure 1–4 [10] illustrates the designs from these manufacturers. It is interesting to notice that the simplest device, by Intel, has the fastest access time, $120ns$, and the best reliability, 10^5 program/erase cycles. This compares with access times for Seeq and Toshiba of $200ns$ and $170ns$ respectively, and reliabilities of 10^3 and 10^2 program/erase cycles, respectively.

Flash EEPROMs are programmed by Fowler-Nordheim tunnelling [10] [2], in a similar manner to the FETMOS. However, memory cells are all programmed simultaneously, from which stems the name *flash*. In the SEEQ device for example, the drain is raised to $\simeq +19V$, while the control gate is grounded and the source is left floating [17]. Electrons pass from the floating gate to the drain, giving a negative threshold voltage shift.

Most flash EEPROMs are erased using a hot-electron injection technique [10]. An electrons is said to become hot when its drift velocity is comparable to its thermal velocity [18]. Thus, erasing requires a large voltage of $\simeq +20V$ [17] on the drain and control gate, while the source and substrate are grounded [10]. Channel hot electrons are created in the high electric field near the drain junction. Since

[2] For consistency within this thesis, program will describe the condition with stored positive charge on the floating gate. However, many flash EEPROM manufacturers consider program to denote stored electrons.

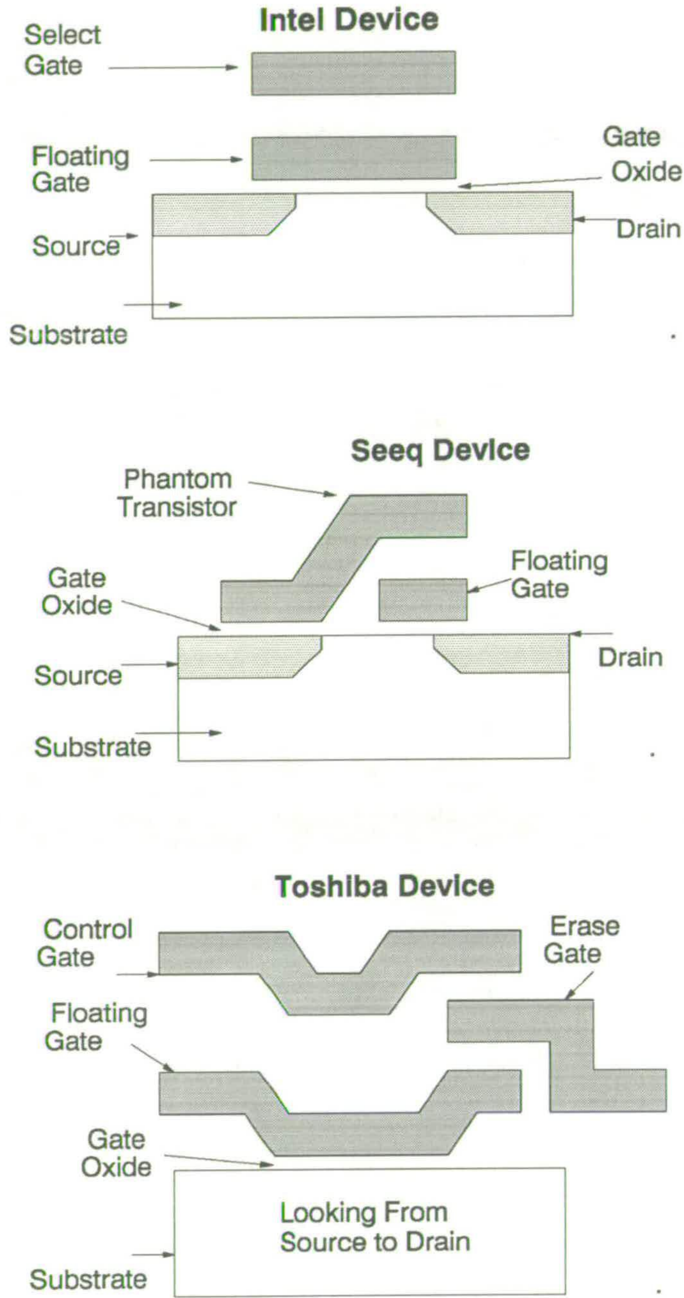

Figure 1–4: Flash EEPROM Designs. These are Realised in NMOS.

the oxide field favours injection, the electrons are then transported to the floating gate [1]. A positive threshold voltage shift results. Of course, channel hot electron injection can only bring electrons onto the floating gate.

The read operation resembles that of the FETMOS, where the threshold voltage defines either a logic 1 or 0.

1.1.4 Contrast Between FETMOS/FLOTOX and Flash

Memory technology in general is driven by cell size, and this will form the framework this discussion.

Flash EEPROM

To achieve a smaller cell in flash memories, the select transistor is omitted. Flash memories thus forgo the ability to be programmed in small sections, eg. in words. Instead they must be programmed in large blocks, this is known as bulk programming. From a practical view this is the main distinction between flash cells and FETMOS/FLOTOX cells.

FETMOS and FLOTOX

From the viewpoint of integration the FETMOS structure exhibits a number advantages over the FLOTOX [19]:

- A smaller cell area, which allows denser memory circuits to be produced.
- A simpler cell structure, which is compatible with easier scaling of dimensions [1].
- One less photomask layer is required during processing.
- The FLOTOX requires tighter lithographic control, for defining extremely small tunnel oxide dimensions.

A recognised drawback of the FETMOS is that it allows a higher substrate current during programming [20]. This is due to its low drain junction breakdown voltage [19]. Thus it draws more current than the FLOTOX during the program operation. Logic circuits including relatively small FETMOS arrays are generally unaffected by the higher current requirement [20]. However, for large arrays, special attention to charge pump design is needed. A number of other limitations have been cited for the FETMOS including:

- Threshold voltage shifts due to charge trapping in the tunnel oxide.
- Reduced endurance due to the inclusion of low integrity field oxide edges in the tunnel area.

However, the significance of these concerns remains a moot point [19].

1.2 Utility of the EEPROM

Let us now discuss the most popular uses for the EEPROM. There are two major driving forces in the development of EEPROM technology. One is for high density memories requiring low cell size and lowest cost per bit. The other requirement is in non-volatile low density memories, for micro-controllers and programmable logic type applications.

1.2.1 Program/Data Storage Media

EEPROM technologies of all kinds have been produced simply as memory chips [1] [11], as such they were intended mainly to displace the EPROM [1]. However, flash EEPROMs are now beginning to have a far reaching impact. With their high densities they are winning an increasing market share from magnetic tape and disc storage media.

Advantages of EEPROM over magnetic media include:

- Physical ruggedness.
- Light weight.
- Low power consumption.

Flash EEPROM replacements for the hard disc will be available very soon in the form of memory cards $\simeq$ 20Mbyte [8]. The effect on portable computers is to increase battery life from $\simeq$ 4 hours to between 30 and 60 hours. Weight is also reduced, from $\simeq$ 7 pounds to between 1 and 2 pounds [8]. The optimism felt by the industry for this technology may be summed up in a quote from Intel's marketing manager (made in 1990) [9]: " We understood DRAMs to be the smallest and simplest memory device available but now flash is *that* and non-volatile too. The Intel 1Mbit flash EEPROM is *smaller* than the NEC 1Mbit DRAM"...

1.2.2 Embedded Systems

EEPROMs may be embedded into logic circuits, such as microcontrollers. These are close cousins of the microprocessor, but have added features such as A to D conversion and EEPROM memory. In a microcontroller the EEPROM is often used to store the configuration of a machine, before power down. FETMOS devices are produced by Motorola on there HC11 micro-controller, in a block of 512 Bytes [21]. This has a large share in the lucrative automotive market, for self-tuning engine systems, where the EEPROM stores the engine configuration when the journey is over.

1.2.3 Embryonic Technologies

EEPROMs offer advantages in many of tomorrow's technologies such as artificial intelligence, self adaptive systems [1]. The elegance of an EEPROM solution to a problem is epitomised in Neural Networks. Briefly, these are systems which mimic

the brain. Information is transmitted between neurons as discrete nerve impulses, with information encoded as an analogue potential: an "impulse density" [22]. Neurons receive impulses at "synaptic sites" and an arriving impulse generates an analogue potential, which is scaled in proportion to a "synaptic weight". Arriving potentials are summed , and when the summed potential exceeds a given threshold, nerve impulses are generated and transmitted to other neurons. At face value such a system may not appear very powerful, and an individual neuron does only very simple tasks. The high computing power of neural systems arise from the collective behaviour of large, highly interconnected, fine grain networks [23]. The learning function of neural networks originates from their ability to change the synaptic weights. Responses of the system may then optimised to solve a particular problem, eg. pattern recognition. In VLSI neural network circuitry the synaptic weight is commonly stored in a shift register [22], which is costly in terms of chip area. Alternatively however, an EEPROM may be used to store the synaptic weight, as an analogue charge on the floating gate [24]. Thus the number of transistors required per neuron is reduced, and a finer network produced. Naturally there are problems in such an implementation: program/erase characteristics of EEPROMs tend to vary across a wafer; and *reliable* EEPROMs are required, for which the program/erase characteristics remain stable with use [24]. [3]

1.3 Reliability

Since the first integrated circuits were made (around 1959) the maximum number of devices that can be successfully integrated into a single chip of silicon has risen steadily. This trend in increasing transistor counts was vocalised by Moore, who predicted in 1964 that the number of transistors would at least double every two years. The validity of this is born out in Figure 1–5 [25], which gives the transistor

---

[3] A "feedback-based" programming method has been used to over come these problem, though its slowness limits the usefulness of the EEPROM in the system [24].

counts per die for the densest examples of memory circuits, the DRAM. With more transistors available, it has become possible to produce logic circuits such as micro-controllers, with ever greater functionality. Indeed, as the level of integration on logic circuitry swells, so more of the transistor budget may be allocated to memory structures [25].

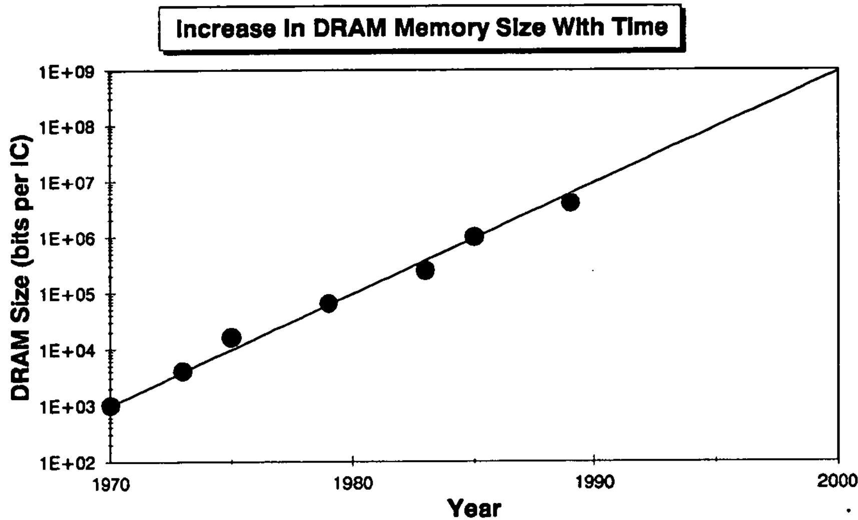

Figure 1–5: Increase in Integration Levels Over a Period of Years.

1.3.1 Shrinking Geometries

While there are many technologies available, MOS devices are at the leading edge of VLSI in terms of packing density, and will receive sole attention here. Integration levels may be increased in a number of ways [26]:

1. Increasing Die Area. Unfortunately this also makes the circuits more prone to defects, so reducing yield.

2. Design Improvements. Use of more efficient circuit architectures, which requires less components.

3. Circuit Layout. Careful attention to the layout can give a reduction in the area needed by a circuit.

4. Scaling down device feature size. In its simplest form this simply means reducing all dimensions by a factor α.

Scaling makes the largest contribution to the increasing complexity of integrated circuits [26]. Although there must be a limit beyond which further integration becomes uneconomic. In the the early 1980s it was argued that pushing minimum geometries below $0.5\mu m$ would yield only diminishing returns [27]. In the interim improvements in device design and fabrication technology have reduced this limit. Today's research have yielded silicon MOSFETS in the deep-sub-micron regime, with channel lengths of $0.15\mu m$ [28]. Meanwhile, state of the art circuits in commercial production have critical dimensions of $\sim 0.8\mu m$ [28]

1.3.2 Reliability In General

As transistor density escalates so do reliability problems. Integrated circuits generally have very long life expectancies $\sim$ decades. Never the less they do suffer from reliability problems. General reliability problems come in many guises, but two principal classes are:

1. Electromigration in metal interconnections [29]

2. Transistor threshold shift and gate oxide rupture, due to hot electron effects [30]. These can be minimised by careful engineering of the drain region.

In a logic circuit, the failure of only one device may cause the entire circuit to fail. Reliability may be measured in terms of FITs, where 1 FIT represents 1 failure in 10^9 device hours. (eg: if 10^9 transistors are tested for an hour, and one fails, we have 1 FIT). Currently acceptable failure rates are considered to be $\sim$ 100 FITs [31]. Clearly however, as transistor counts grow, so FIT rates must fall. A daunting 0.1 FIT rate has been proposed as a reliability goal for the year 2001, by The Semiconductor Research Council [32]. Thus, methods of improving device reliability are a major concern to the semiconductor industry as a whole.

1.3.3 Reliability and the EEPROM

Due to the tunnelling of electrons through the gate oxide, EEPROMs are especially susceptible to reliability problems. FETMOS devices have a maximum life expectancy of only 10^5 program/erase cycles [11], while flash EEPROMs have a life expectancy of $\sim 10^3$ cycles [10], depending on the manufacturer. In addition, EEPROMs are sensitive to fabrication conditions and may not display an acceptable threshold window, so lowering yield. Accurate control of the processing and reduction of impurities obviously help. Beyond this, two areas are of interest:

1. It would be useful to know exactly how a change in processing effects the threshold window and reliability of an EEPROM. For instance, a smaller floating gate may give a higher threshold voltage, or improved reliability. With this knowledge, some degree of process optimisation may be possible.

2. It would also be beneficial to design a device with a higher innate reliability.

These then, form the joint objectives of this thesis.

1.4 Thesis Plan

Before tackling either of the above problems however, it will be necessary to more closely examine the mechanics of the EEPROM. This will be the subject for Chapter 2. In chapter 3 a model is derived for the FETMOS device, and in chapter 4 this is used to analyse FETMOS threshold window and reliability. In chapter 5 the fabrication of novel test structures is described, which are equivalent to a set of FETMOS devices. In chapter 6 these test structures are used to analyse FETMOS reliability. Finally, in chapter 7 results are summarised, overall conclusions are drawn, and recommendations are made regarding improved FETMOS design.

Bibliography

[1] H.E.Meas, J.Witters, and G.Groeseneken. Trends in non-volatile memory devices and technologies. In *ESSDERC Bologna*, pages 743–754, 1987.

[2] D.Bondurant and F.Gnadinger. Ferroelectrics for nonvolatile RAMs. *IEEE Spectrum*, 26(7):30–33, 1989.

[3] B.C.Cole. Finally, it's ferroelectric. *Electronics*, pages 88–89, August 1989.

[4] B.Halg. On a micro-electro-mechanical nonvolatile memory cell. *IEEE Transactions On Electron Devices*, 37(10):2230–2236, October 1990.

[5] H.Haznedar. *Digital Microelectronics*, pages 476–496. Benjamin Cummings, 1991.

[6] R.Bez, D.Cantarelli, and P.Cappelletti. Experimental transient analysis of the tunnel current in EEPROMs. *IEEE Transactions On Electron Devices*, 37(4):1081–1086, 1990.

[7] E.S.Yang. *Microelectronic Devices*, chapter Charge-Coupled And Nonvolatile Memory Devices. McGraw-Hill Book Company, 1988.

[8] D.Manners. Intel chip ousts hard discs. *Electronics Weekly*, (1587):24, 1992.

[9] S.Parry. Flash of memory to fill the gap. *New Electronics (On Campus)*, 1(2):11, 1990.

[10] R.D.Pashley and S.K.Lai. Flash memories: The best of two worlds. *IEEE Spectrum*, 26(12):30–33, December 1989.

[11] C.Kuo, Y.R.Yeargain, and W.J.Downey. An 80ns 32K EEPROM using the FETMOS cell. *IEEE J.Solid State Circuits*, (5):821–827, October 1982.

[12] J.F.Dickson. On-chip high voltage generation in MNOS integrated circuits using an improved voltage multiplier technique. *IEEE Journal Of Solid State Circuits*, (3):374–378, 1976.

[13] J.Callder. Master's thesis, University Of Edinburgh, 1988.

[14] J.R.Yeargain and C.Kuo. A high density floating-gate EEPROM cell. In *IEEE IEDM*, pages 24–27, 1981.

[15] S.K.Lai, V.K.Dham, and D.Guterman. Comparison and trends in today's dominant EE technologies. In *IEEE IEDM*, pages 580–583, 1986.

[16] P.I.Suciu, B.P.Cox, D.D.Rinerson, and S.F.Cagnina. Cell model for EEPROM floating gate memories. In *IEEE IEDM*, pages 737–740, 1982.

[17] G.Samachisa, C-S.Su, and Y-K.Kao. A 128k flash EEPROM using double polysilicon technology. In *IEEE International Solid State Circuits Conference*, pages 76–77, 1987.

[18] B.G.Streetman. *Solid State Electronic Devices*, chapter 3. Energy Bands And Charge Carriers In Semiconductors. Prentice/Hall International Editions, 1980.

[19] C.Kuo, K.Fu, P.Kim, M.Chonko, and J.Jorvig. High FETMOS EEPROM cell using ONO inter-polysilicon dielectrics. pages 98–99.

[20] K.Y.Chang, S.Cheng, and K-M.Chang. An advanced high voltage CMOS process for custom logic circuits with embedded EEPROM. In *CICC*, 1988.

[21] *HCMOS Single-Chip Microcomputer*, chapter 1. Motorola Inc, 1985.

[22] A.Masaki, Y.Hirai, and M.Yamada. Neural networks in CMOS: A case study. *IEEE Circuits And Devices*, 6(4):12–17, 1990.

[23] H.P.Graf and L.D.Jackel. Analogue electronic neural network circuits. *IEEE Circuits And Devices*, 5(4):44–50, 1989.

[24] C-K.Sin, A.Kramer, and V.Hu. EEPROM as an analog storage device, with particular application in neural networks. *IEEE Transactions On Electron Devices*, 39(6):1410–1419, 1992.

[25] P.P.Gelsinger, P.A.Gargini, and G.H.Parker. Microprocessor circa 200. *IEEE Spectrum*, 26(10):43–47, October 1989.

[26] D.Gorham, J.Wood, and D.Butts. *Field Effect Devices And VlSI*, chapter 5 Steps Towards VLSI. The Open University Press, 1985.

[27] R.T.Bates. Nanoelectronics. *Solid State Technology*, 32(11):101–107, November 1989.

[28] R-H.Yan, K.F.Lee, and D.Y.Jeon. 89GHz ft room-temperature silicon MOSFETs. *IEEE Electron Device Letters*, 13(5):256–257, May 1992.

[29] E.A.Amerasekra and D.S.Campbell. *Failure Mechanisms In Semiconductors*, chapter 3. John Wiley And Sons, 1987.

[30] J.J.Sanchez, K.K.Hsueh, and T.A.DeMassa. Drain-engineered hot-electron-resistant device structures: A review. *IEEE Transactions On Electron Devices*, 36(6):1125–1132, June 1989.

[31] D.L.Crook. Evolution of reliability engineering. In *IEEE/IRPS*, pages 2–7, 1990.

[32] C.Hu. IC reliability simulation. *IEEE Journal Of Solid State Circuits*, 27(3):241–246, March 1992.

Chapter 2

EEPROM Physics and Reliability

2.1 Fowler-Nordheim Tunnelling

Fowler-Nordheim tunnelling describes a quantum mechanical effect, implicitly linked with theories produced by quantum physicists during the early twentieth century. Although Fowler-Nordheim (FN) tunnelling is often mentioned in literature, the underlying physics of the effect only ever receive a cursory consideration. In this thesis FN tunnelling is met both in the analysis of test structures and in the modelling of the EEPROM cell. As such, FN tunnelling is central to the fabric of the thesis and merits a closer examination. With a clear understanding of the FN tunnelling mechanism, the physical accuracy of any model for the EEPROM cell may be assessed with greater confidence. The relationship between the FN tunnel current and parameters within a test structure (eg. gate oxide thickness) may also be better understood. Quantum mechanics is a very broad subject and only the salient features relating to FN tunnelling will be examined.

2.1.1 Wave Particle Duality

Following Einstein's theory (in 1905) that light may exhibit both a wave and a particle nature, de Broglie (in 1924) extended the idea of dualism, to suggest that particles may also exhibit a wave nature. This is to say particles such as electrons also behave as waves [1]. Their wavelength is given by:

$$\lambda = \frac{h}{mv}$$

Where:

- λ = Wave length
- h = Planck's constant = $6.625 \times 10^{-34} Js$
- m= Mass of the particle (eg. rest mass of an electron = $9.1091 \times 10^{-31} Kg$)
- v= Velocity of the particle

Waves normally conjure up an image of a moving entity (eg. sea waves travelling towards the shore), and a *free* electron will fit such a description. However, an electron *bound* within a silicon atom will be stationary, and as such it can exist only as a standing wave. An analogy for this regime is that of a guitar string fixed at both ends, in which a stationary wave may also be produced.

All waves may be described by an appropriate set of equations (eg. Maxwell's equations for electromagnetic radiation). It was Schrödinger (in 1925) who developed an equation to describe the wave nature of particles [2]. Even though the mathematics is quite complex, it is interesting to include a version of Schrödinger's equation, if only to introduce the variable Ψ. Hence, the standing wave associated with an electron (for a one dimensional case) may be described by [2]:

$$-\frac{\hbar^2}{2m}\frac{d^2\Psi}{dx^2} + V\Psi = E\Psi$$

Where:

- Ψ = The quantity which varies in the wave.
- $\hbar = \frac{h}{2\pi}$ = Reduced Planck's constant = $1.054 \times 10^{-34} Js$
- m = Mass of the electron
- x = Position
- $\left(-\frac{\hbar^2}{2m}\frac{d^2\Psi}{dx^2}\right)$ = Kinetic energy of the electron.

- V = Potential energy of the electron
- E = Total energy of the electron

The concept of Ψ itself is an abstract one, but was given a tangible interpretation by Born, who proposed that $|\Psi|^2\,dx$ represents the probability of finding an electron between a distance x and $x + dx$ from an origin. As an example, consider an electron bound within a single hydrogen atom. For the electron in its lowest energy level $|\Psi|^2\,dx$ is illustrated in Figure 2–1 .

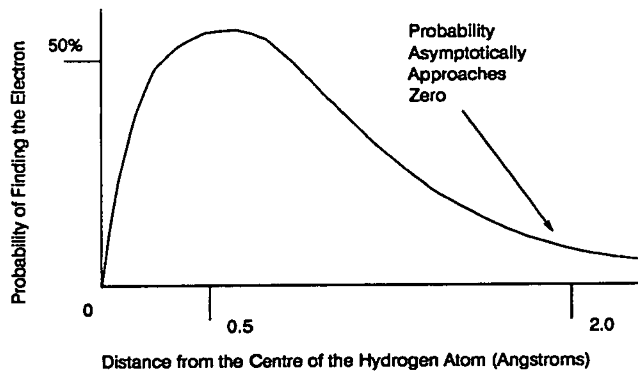

Figure 2–1: Probability of Finding an Electron at a Distance x From the Center of a Hydrogen Atom.

The salient feature to note is that at large radii, $|\Psi|^2\,dx$ decreases asymptotically, but never reaches zero. There is a possibility that the electron could be found anywhere, either inside or outside the atom. In fact, one can never be certain where an electron is at any time.

2.1.2 Fowler-Nordheim Tunnelling in MOS Structures

The energy band diagram for a MOS structure is given in figure 2–2 [3] This is equivalent to the region in a FETMOS device where the floating gate and drain overlap. Electrons in the polysilicon conduction band meet a large energy barrier at the polysilicon/oxide interface, whose height is $3.1eV$. Application of a small voltage to the drain, $\sim 1V$, will cause the energy bands to bend, with the

majority of the voltage falling across the oxide. However, in the dark and at room temperature, few electrons have sufficient energy to surmount the barrier, and current flow will be negligible. Remembering the function $|\Psi|^2$, it can be said that

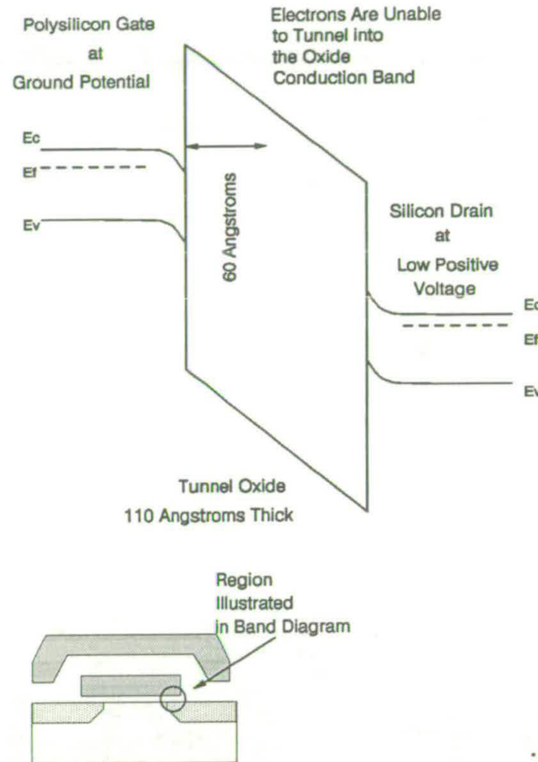

Figure 2–2: Energy Band Diagram for a MOS Structure, Under a Small Applied Bias. This is Equivalent to the Floating Gate/Drain Overlap Region of a FETMOS Device.

there is a possibility of the electrons leaving the polysilicon gate and reaching the drain. The probability of this increases for thinner tunnel oxides, and at $60\AA$ the probability is sufficiently high for the current to become significant. This process is known as quantum mechanical tunnelling. It places a theoretical limit on the minimum oxide thickness suitable for a floating gate EEPROM. Oxides below $60\AA$ would readily leak charge, giving data retention problems. However, for a $110\AA$ tunnel oxide the current flow is negligible.

As the drain voltage is raised so band bending becomes more pronounced. This represents a lowering in the energy level of the oxide conduction band. Once

band bending has become sufficient, electrons can cross into the oxide conduction band. This is Fowler-Nordheim tunnelling, which becomes significant for an oxide electric field of $\sim 7MVcm^{-1}$ or greater. Here, the polysilicon gate/oxide interface is referred to as the "injecting interface". Tunnelling electrons are then accelerated to the drain, and as they travel electrons loose energy in collisions with atoms in the oxide, this is illustrated in figure 2–3. Since the energy barrier between holes and the oxide valence band, $4.3eV$ [4], is larger than that between electrons and the oxide conduction band, $3.1eV$, the hole current is negligible in comparison. The electrical characteristic for a MOS structure under a voltage ramp is given in figure 2–4.

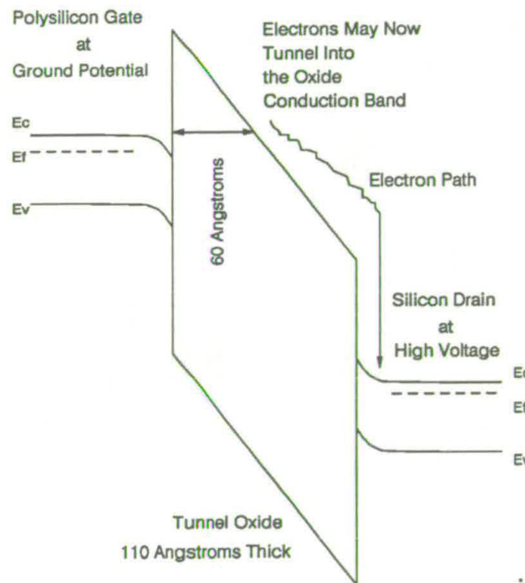

Figure 2–3: Energy Band Diagram Illustrating Fowler-Nordheim Tunnelling in a MOS Structure.

2.1.3 Factors Which Affect Fowler-Nordheim Tunnelling

Fowler-Nordheim tunnelling is an electrode limited process, as opposed to a bulk limited process [5]. Thus, the tunnelling current will be varied by phenomena at the injecting interface. Some effects are negligible, such as image force barrier lowering, which tends to "round off" of top of the polysilicon/oxide energy barrier

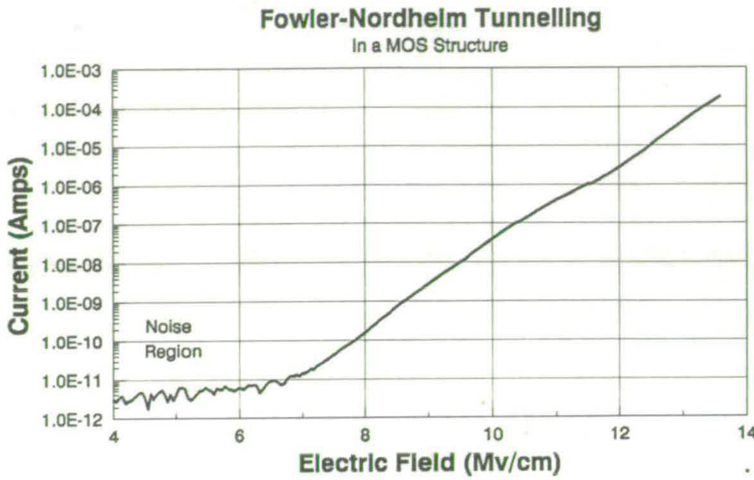

Figure 2–4: Electrical Characteristic for a MOS Structure Under a Voltage Ramp. This measurement was made on a MOS capacitor of area $2.5 \times 10^{-4} cm^2$

[6] [7]. This has little effect since electrons are mainly distributed at the base of the energy barrier, whereas rounding is confined to the top of the barrier. The energy of incident electrons may be increased by temperature or illumination, at higher energies these electrons have a shorter distance to tunnel. However, the majority of factors are process dependent, such as doping density of the polysilicon. This may be used to increase the number of electrons incident on the silicon surface, and hence the tunnel current. Indeed, the polysilicon/oxide interface itself is not well defined, rather there is a transition region of $\simeq 10\mathring{A}$ which consists of silicon rich oxide [8]. This region contains silicon "islands" at which the electric field is locally enhanced, increasing the tunnel current [9]. It has also been noticed, that the Fowler-Nordheim coefficients A and B increase as oxides become thicker, in the range $60\mathring{A}$ to $140\mathring{A}$, but saturate towards thicker oxides [10]. While there is no good argument to explain this, it may be that the thin transition region at the S_i/S_iO_2 interface becomes more influential for thinner oxides [10]. Defects of various descriptions will also effect tunnelling [11] [12], these are discussed further in chapter 5. However, two of the most significant influences on Fowler-Nordheim tunnelling are charge trapping, and field enhancement due to asperities.

Charge Generation and Trapping in the Oxide

A proportion of the electrons being accelerated towards the drain experience impact ionisation events in the oxide, which produce electron/hole pairs [13]. The holes are then accelerated towards the gate, and electrons continue towards the drain. A proportion of each charge species becomes trapped in the oxide, which affects the field at the injecting interface [14], as illustrated in figure 2–5. Hole trapping enhances the field at the injecting interface, whereas electron trapping reduces it.

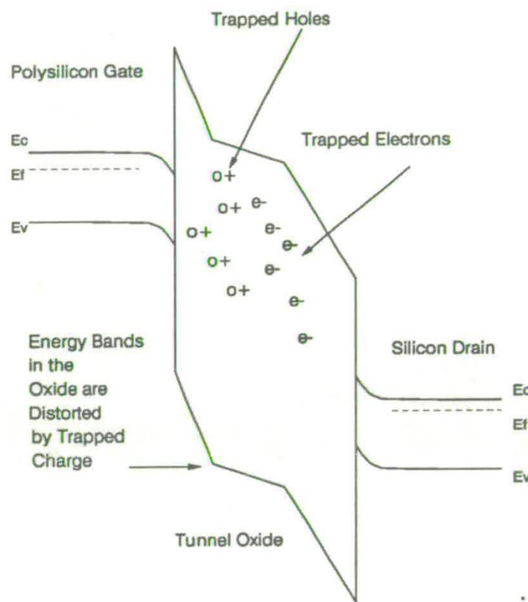

Figure 2–5: Energy Band Diagram for a MOS Structure After Charge Trapping.

Asperities at the Injecting Interface

The crystals which make up polysilicon form a rough surface with many ridges, or "asperities" [14], as illustrated in figure 2–6 [15]. These bend the electric field lines which become crowded at the asperities, since electric field lines always lie normal to the surface of a conductor [7]. Thus, asperities at the injecting interface will locally enhance the electric field [14], usually by a factor of 3 to 5 times [16]. Therefore the tunnelling current becomes locally enhanced, as will the associated

charge trapping. In some devices asperities are enhanced during processing, to give so called textured surfaces. These devices use much thicker tunnel oxides, of $\simeq 600\AA$ to $1000\AA$ [16], while still providing the required tunnel currents. Although a thin tunnel oxide is no longer necessary, charge trapping becomes enhanced, which reduces any benefits [17].

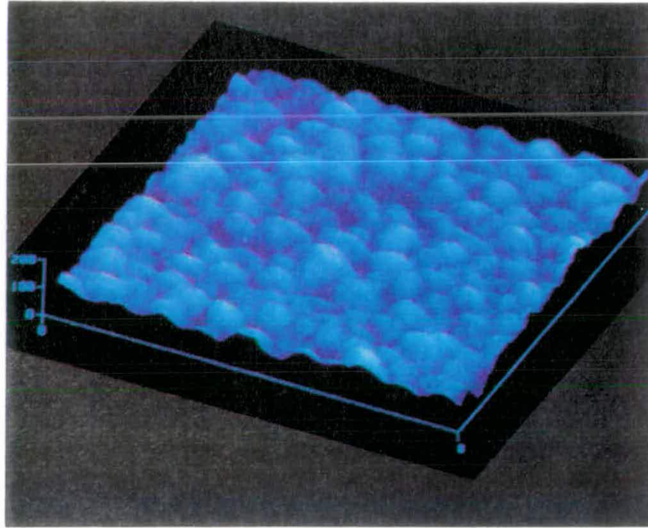

Figure 2–6: Quantum Tunnelling Micrograph, Illustrating Asperities on a Polysilicon Surface.

2.1.4 An Equation to Describe Fowler-Nordheim Tunnelling

Fowler-Nordheim tunnelling is described by equation (2.1):

$$J = AE^2 exp^{(\frac{-B}{E})} \tag{2.1}$$

Where:

- $J=$ Current density flowing through the oxide.
- $E=$ Electric field strength across the oxide.

- A and B = Fowler-Nordheim Coefficients.

The Fowler-Nordheim coefficients may calculated either experimentally [18], or theoretically [6] [5]. For theoretical calculations A and B are expressed in terms of fundamental parameters such as the height of the polysilicon/oxide energy barrier. To give greater accuracy modifying factors may also be added, to account for such effects as image barrier lowering [5]. Even so, no theoretical expression has been developed to include all factors which effect Fowler-Nordheim tunnelling, such as trapping of carriers in the oxide [13]. In addition, oxide processing has a significant effect on tunnelling characteristics. Accurate values for A and B should therefore be extracted from experimental data. This is considered in chapter 3.

2.2 Reliability Issues

2.2.1 Oxide Breakdown

Since EEPROM failures result almost exclusively from tunnel oxide degradation, we shall first review tunnel oxide reliability. As with all VLSI processes, manufacturers take much trouble to produce good quality oxides, and over the small area of one EEPROM cell the oxide should be of uniformly good integrity. Therefore, we will be concerned largely with defect free oxides in our review.

The reliability of thin oxide films is of great importance to the MOS semiconductor industry, since oxide failures make up a large proportion of yield loss. For this reason, a wealth of material has been published on the subject. Never the less, the mechanism of oxide failure is not well understood, and there are a number of comp eting theories in existence. To give a feel for the processes associated with oxide degradation, the impact ionisation model has been chosen for consideration [13] [19] [20]. This is a well established model for which supporting evidence is still being produced today [21]. The hole trapping phenomenon this proposes, would explain why radiation hard processing techniques, as used in this work, produce oxides of superior quality.

Methods of Oxide Stressing

An oxide must be stressed to assess its reliability, and the stressing techniques may be subdivided into two categories [22]:

1. TZDB: Time Zero Dielectric Breakdown. This is essentially field dependent stress.

2. TDDB: Time Dependent Dielectric Breakdown. This is essentially time dependent stress.

TZDB: A ramped voltage is applied to the oxide causing it to rupture. The steep rise in current observed at breakdown then gives a convenient and unambiguous signal that rupture has occurred. The quality of the oxide will be indicated by the electric field required for breakdown, given in $MVcm^{-1}$. TZDB can be categorised into three modes [11] [23]:

1. A mode. This is attributed to pin holes in the gate oxide because of the nearly zero breakdown field of $E_{BD} \leq 1MVcm^{-1}$.

2. B mode. This is caused by a defect, giving intermediate breakdown field of $1MVcm^{-1} < E_{BD} \leq 8MVcm^{-1}$. The upper limit of the B mode breakdown voltage is determined self-healing energy, which is necessary to explosively evaporate the gate polysilicon layer above the B mode defect. The lower limit for this breakdown is determined by thermal breakdown. Breakdown happens when the Joule heating due to conduction, for which power$= I^2R$, exceeds the rate of energy dissipation.

3. C mode. This failure mode is due to intrinsic breakdown of the oxide, and typically occurs for $E_{BD} \geq 8MVcm^{-1}$.

The field at which intrinsic breakdown occurs defines the "dielectric strength" of the oxide sample. Note that for oxides thinner than $\simeq 150\mathring{A}$, it becomes difficult to distinguish between B and C mode failures [23].

TDDB: A constant voltage, or constant current, is applied to the oxide until it ruptures. In essence, there is no qualitative difference between constant current or constant voltage stressing [13], and for this review constant voltage stressing will be considered, as this method of testing is used in this project. The applied voltage should be sufficient to produce a Fowler-Nordheim tunnel current in the oxide, while still remaining below the oxide's dielectric strength. Again, the steep rise in current observed at breakdown gives an unambiguous signal that rupture has occurred. In this case, however, oxide quality is indicated by the amount of charge which has passed through the oxide before breakdown. On a practical level, this will be the integral of current as a function of time, referred to as Q_{BD}.

Although one EEPROM program/erase operation is short lived ($\sim 10ms$), over the life time of an EEPROM many such operations will take place, and the net length of time spent in programming/erasing will be up to 10^5 times as long [24]. Thus constant current or voltage stressing will roughly emulate conditions during programming/erasing of an EEPROM [13]. TDDB has therefore been chosen as the principle method of investigating the relative reliabilities of devices fabricated in this work.

The Impact Ionisation Model

The Impact Ionisation Model sets out principally to explain TDDB, although literature suggests that breakdown mechanisms for TDDB and TZDB are the same [25]. Conceptually TDDB can be thought of as a two stage process:

1. Build up.
2. Runaway.

Figure 2–7 shows a graph of current against time (I/t), for a constant voltage TDDB test. Since the runaway stage takes only fractions of a second to be completed, it is the build up stage which determines the life time of the oxide.

During the build up stage, electrons will pass into the oxide conduction band by Fowler-Nordheim tunnelling. The salient feature of this build up stage is a

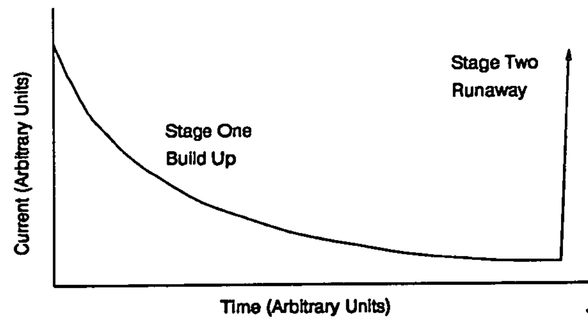

Figure 2–7: Current Against Time for a Constant Voltage TDDB Test.

decrease in tunnel current with time, which is widely accepted to be a result of electron trapping in the oxide. Electrons will become trapped over a wide range of distances from the injecting interface, or cathode. However, trapped electrons may be modelled as a charge sheet, whose centroid lies at X_N, as illustrated in figure 2–8 . The effect of trapped electrons is to reduce the field at the cathode, while increasing the anode field. This reduction in cathode field is responsible for the reduction in tunnel current. Notice also, that the slope of the TDDB curve never becomes level, indicating that electron trapping continues throughout the entire test, ie. the traps are never completely filled. This non-saturating behaviour is a general feature of TDDB testing, and is believed to be due to the generation of electron traps during the test, under the influence of the high field.

A proportion of the injected electrons gain sufficient energy to cause impact ionisation and generate electron/hole pairs [26]. Generated electrons then continue to the anode, while holes are swept back towards the cathode. A number of these holes become trapped in the oxide, also illustrated in figure 2–8 . Again, hole trapping proceeds over a range of distances from the cathode. This may also be modelled as a charge sheet, whose centroid lies at X_P. By enhancing the cathode field trapped holes tend to raise the magnitude of tunnel current, and so have a contrary effect to trapped electrons. Thus the question is raised, why should the effect of electron trapping dominate? It has been proposed that the hole trapping proceeds only over a small area of the oxide, and experiment has indicated that

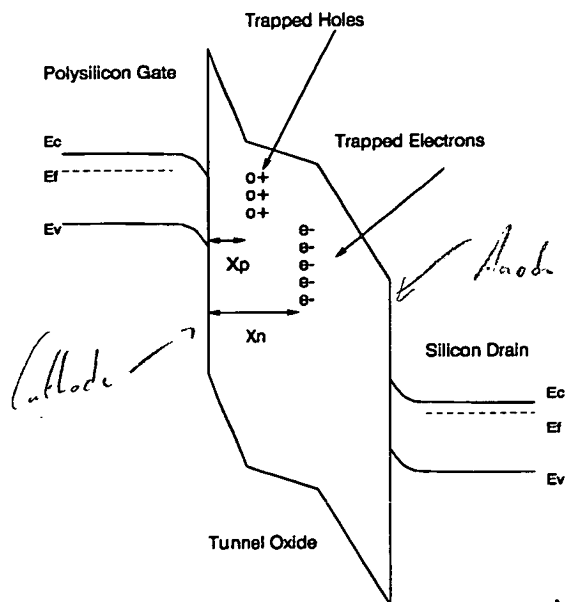

Figure 2–8: Energy Band Diagram to Illustrate Charge Trapping During Voltage Stressing of the Oxide.

hole trapping is localised to approximately 1 part in 10^6 of the total oxide area [13].

Even in a good quality oxide there will be a degree of inhomogeneity, hence we may subdivide areas into two categories, connected in parallel:

1. "Robust areas". These are relatively *un-susceptible* to hole trapping at the cathode.

2. "Weak areas". These are relatively *susceptible* to hole trapping at the cathode.

A positive feedback loop evolves in the weak areas, since the increased current produces a larger number holes through impact ionisation, which enhances the cathode field ..... and so on. Associated with this will be a localised increase in the rate of electron trapping, which locally raises the anode field. Given that the impact ionisation coefficient has a strong field dependence, this phenomenon will provide an added fillip to the positive feedback cycle. Once the localised current

density obtains a "critical value" runaway will begin, this is stage two of the TDDB process. Current instability then leads to electrical and thermal runaway, associated with catastrophic failure.

Early breakdowns, resulting from defects, may also be described by the impact ionisation model [13]. Such defects are assumed to suffer from a combination of one or more of the following ailments:

- A high density of hole traps.
- A large hole capture cross section.
- A lower effective tunnelling barrier height.

2.2.2 Analysis of Reliability Data

Consider a batch of wafers, on which thin S_iO_2 capacitors have been fabricated. The reliability of the batch as a whole, will be an aggregate of of the reliability of individual capacitors. Some capacitors will be more reliable than others, and this spread will have a random nature. Therefore, *many* capacitors should be tested to access the reliability of the batch, and statistical validity will improve as larger sample sizes are used.

Historically, the microelectronics industry has used the "bathtub curve" when discussing reliability [27]. The results which could be expected from a TDDB test over a large sample of oxide capacitors is illustrated in figure 2–9. This curve is characterised by three regions [28]:

1. A high initial failure rate, the so called "infant mortality" period.
2. A low but nonzero midlife failure rate, attributed to random failures. This region represents the useful life of the capacitor.
3. An increasing failure rate at end-of-life, due to intrinsic wearout mechanisms.

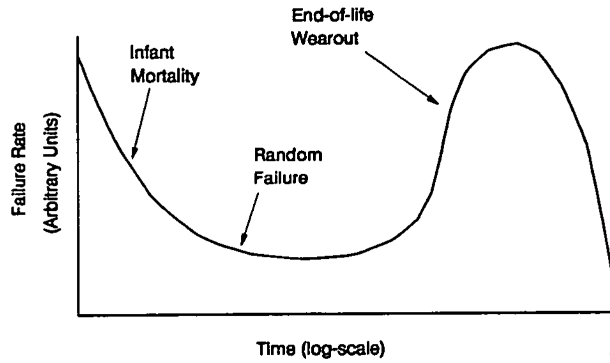

Figure 2–9: "Bathtub Curve".

The overall shape of the bathtub curve will vary, depending on the processing conditions of the oxide. The entire curve may be characterised using a Weibull distribution [29] [30], and equation (2.2) gives a simple form of this:

$$F(t) = 1 - e^{\left(-\frac{1}{\alpha}t^{\beta}\right)} \tag{2.2}$$

where:

- F = Cumulative probability of failure.
- t = Time to breakdown.
- α = Constant
- For $\beta < 1$ the failure rate decreases with time.
  For $\beta = 1$ the failure rate is constant.
  For $\beta > 1$ the failure rate increases with time.

Of course, low infant mortality and midlife failure rates are preferable. However, for this project intrinsic reliability problems are of chief interest, associated with end-of-life failure. For previous researchers the intrinsic breakdown of S_iO_2 has been shown to have a log-normal distribution [20]. This gives it the bell shaped Gaussian curve, observed when a logarithmic x-axis is used, as in figure 2–10. This distribution is characterised by two parameters:

1. The median time to failure. This is the time taken for half of the sample to fail.

2. The shape factor σ. A low value of σ indicates a failure distribution which is tightly grouped in time.

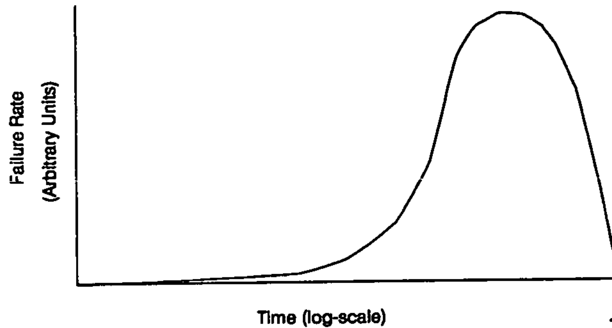

Figure 2–10: A Gaussian Distribution.

The validity of this model for the failure distribution can be checked by plotting the results on a log-normal graph. The graph can then be scrutinised, to ensure all points lie on a straight line. Significant nonlinearity may indicate that more than one failure mechanism is at play [29]. The shape factor σ will be the slope of this line.

2.2.3 Accelerated Testing

Clearly, the bathtub curve is produced by testing capacitors to destruction. However, if normal operating conditions were used for stressing, tests would take an inordinate length of time. A method for accelerating wearout is obviously required, and for this the semiconductor industry commonly uses increased temperature, voltage, current density or humidity [29]. The reliability under normal operation can then be calculated from the accelerated test, using an acceleration factor. The great majority of accelerated life tests for semiconductors use temperature acceleration. However, high temperature acceleration is not always the most appropriate stress, since "threshold triggered" mechanisms may be encountered [30]. These

are failure mechanisms which are only encountered above a certain threshold temperature. It is possible to detect the onset of a threshold triggered mechanism, by testing reliability at sequentially raised temperatures. This methodology is called "step-stress". It should also be remembered that while high stressing conditions will be suited to detecting wearout mechanisms with high activation energies, there is a risk that a low activation energy mechanism may be concealed. Here again step-stress methodology can be used to detect low activation energy mechanisms. Now, Fowler-Nordheim tunnelling is primarily driven by an increase in electric field, while the integrity of an oxide is judged by its charge to breakdown. Therefore, voltage acceleration has been chosen for use in this project.

2.2.4 EEPROM Reliability Issues

Long Term Charge Retention

For good long term reliability of the EEPROM it is essential that less than 10% of the stored charge leaks away in 10 years [31]. The key to avoiding leakage is the strong dependence of tunnel current on voltage across the oxide, as characterised by the Fowler-Nordheim tunnelling curve. The current rises by an order of magnitude for every $0.8V$ increase in applied voltage. Simple arithmetic can be used to prove the long term retention of any EEPROM, as follows [31]:

- Programming voltage $= 18V$.
- Average programming current $\simeq 1 \times 10^{-10}A$ [32].
- Average charge stored on floating gate $\simeq 8 \times 10^{-14}C$ [31].
- Read voltage $= 1V$.
- Ratio of read disturb current to programming current $= \frac{(18V-1V)}{0.8} \simeq 20$ orders of magnitude.
- Read disturb current $= \frac{1\times10^{-10}}{1\times10^{20}} = 1 \times 10^{-30}A$.

- Now imagine the EEPROM is read for 10 years (3×10^8 seconds).
  The charge lost will be $= 3 \times 10^8 \times 1 \times 10^{-30} = 3 \times 10^{-22} C$.

- As a percentage, only $\frac{3\times10^{-22}}{8\times10^{-14}} \times 100 = 3.7 \times 10^{-7}\%$ of the charge is lost.

Evidently then, charge leakage is negligible.

Endurance

During a life time of being continually programmed and erased, the threshold voltages of an EEPROM will vary. The program/erase endurance characteristics of a typical FETMOS device is given in figure 2–11, this shows program/erase threshold against the number of program/erase cycles [24].

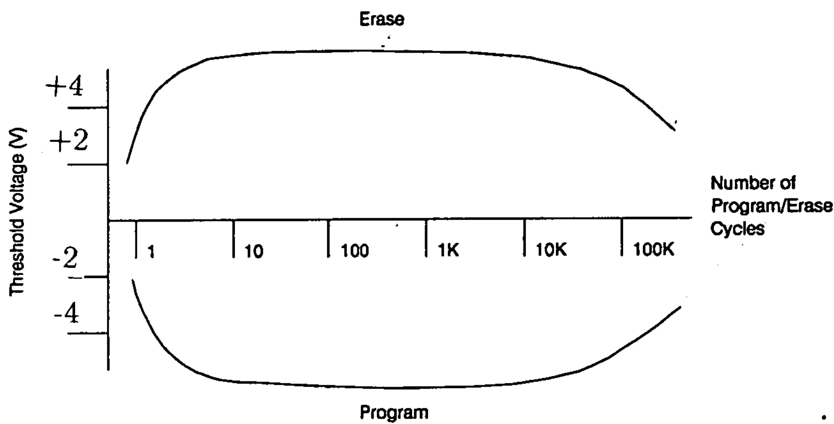

Figure 2–11: Typical Endurance Characteristic for a FETMOS Device.

The endurance characteristic can be divided into 4 regions:

1. Initially the width of the threshold window increases as positive charge is trapped during program/erase operations [24], this regime lasts for $\simeq 10$ cycles.

2. The width remains constant for $\simeq 1 \times 10^3$ cycles.

3. After $\simeq 1 \times 10^3$ cycles the effect of electron trapping during program/erase operations becomes evident. This trapped negative charge reduces the Fowler-Nordheim tunnel current and closes the threshold window, this process is called "trap-up".

4. After $\simeq 1 \times 10^5$ cycles the effect of electron trap-up has closed the threshold window to such an extent, that programmed and erased devices may no longer be distinguished. The EEPROM has now failed.

One should notice that an EEPROM failure occurs even before the tunnel oxide has ruptured. Programming is more susceptible to trap-up than erasing, since the program current is localised, whereas the erase current is distributed over the entire channel region [33].

Firm Errors

Ionising radiation incident on an EEPROM, may excite the stored electrons. Thus, providing them with sufficient energy to leave the floating-gate. Conversely, electrons may be provided with sufficient energy to pass into a floating gate where positive charge is stored. Over an extended period of time many such ionising events would cause the data to be lost. Outside space or military applications, the most common source for this radiation is actually the integrated circuit package [4].

Soft Errors

Ionising radiation may also upset the sense circuitry, which reads the EEPROM cells. A single ionising event is enough to cause a soft error, but the error can be corrected by simply re-reading the data [4] .

Logic Circuits with Embedded EEPROM

In a system containing embedded EEPROM, high voltage transistors are required for the programming/erasing circuitry. In order to reduce the number of masking

steps and make the process economic, high voltage transistors and logic transistors use the same gate oxide [34]. Clearly there is a trade off, since logic transistor require thin oxide $\sim 200\AA$ for speed and density, whereas high voltage transistors require thick gate oxide $400\AA$ for reliability. Thus the inclusion of EEPROM has an effect on the whole circuit.

2.3 Improving EEPROM Operation and Reliability

Two avenues lie open for investigation of the EEPROM, either computer simulation or the fabrication of a set of test devices. Once set up,a simulations allows a large amount of data to be produced relatively quickly. Thus, the effect of varying all parameters in an EEPROM can be assessed, and the most significant identified. Once an important parameter has been isolated, eg. gate/drain overlap, a set of test structures may be fabricated to assess model predictions.

Therefore, it has been decided to produce a model for the FETMOS, in terms of its basic design parameters, such as oxide thickness and gate/drain overlap. The effect of these parameters on FETMOS operation and reliability can then be assessed. No model for the FETMOS currently exists, nor does a suitable methodology for modelling reliability. The development of a new model is discussed in chapters 3 and 4.

Already it has been observed at Motorola that FETMOS devices fabricated with phosphorus drains, are more reliable than those fabricated with arsenic. Phosphorus has a greater mobility in silicon, than does arsenic, and produces a larger gate/drain overlap. Although one may conjecture that improved reliability is due to increased tunnel area, there is nothing to prove that chemistry does not account for improved reliability. Conjecture is an insubstantial foundation on which to base costly commercial semiconductor technologies. This is especially true in today's climate of economic aggression [35]. It would be interesting to clarify the reason for this reliability improvement scientifically. To this authors

knowledge, the relationship between doping or gate/drain overlap, and device reliability has not been investigated in previous research. The fabrication of devices to test this will be the subject of chapters 5 and 6.

Bibliography

[1] M.N.Rudden and J.Wilson. *Elements Of Solid State Physics.* John Wiley and Sons, 1984.

[2] A.P.French and E.F.Taylor. *An Introduction To Quantum Physics.* Van Nostrand Reinhold (UK) Co. Ltd., 1979.

[3] S.M.Sze. *Physics of semiconductor devices*, chapter 7, MIS diode and charge coupled devices. John wiley and sons, 1981.

[4] H.Haznedar. *Digital Microelectronics*, pages 476–496. Benjamin Cummings, 1991.

[5] M.Lenzlinger and E.H.Snow. Fowler-Nordheim tunneling into thermally grown SiO_2. *J.Appl.Phys*, 40(1):278–283, January 1969.

[6] Z.A.Weinberg. On tunneling in metal-oxide-silicon structures. *J.Appl.Phys*, 53(7):5052–5056, July 1982.

[7] M.A.Plonus. *Applied Electromagnetics*, chapter 2, Conductors and charges. McGraw-hill book company, 1978.

[8] S.S.Cohen. Electrical properties of post-annealed thin SiO_2 films. *J.Electrochem.soc Solid State Science And Technology*, 130(4):929–932, April 1983.

[9] D.J.Dimaria, T.N.Theis, and J.R.Kirtley. Electron heating in silicon dioxide and off-stoichiometric silicon dioxide films. *IEEE Transactions On Electron Devices*, 57(4):1214–1238, February 1985.

[10] C.Chang, R.W.Brodersen, and C.Hu. Direct and Fowler-Nordheim tunneling in thin gate oxide MOS strucure. In *Insulating Films On Semiconductors*, pages 176–179. Elsevier Science Publishers B.V (North-Holland), 1983.

[11] H.Shirai, K.Kanya, and A.Yamaguchi. Effect of oxide-induced stacking faults on dielectric breakdown characteristics of thermal silicon dioxide. *J. Appl. Phys.*, 66(11):5651–5653, December 1989.

[12] C.Hashimoto, S.Muramoto, N.Shiono, and O.Nakajima. A method of forming thin and highly reliable gate oxides. *Journal Of The Electrochemical Society*, 127(1):129–135, 1980.

[13] Ih-C.Chen, S.E.Holland, and C.Hu. Electrical breakdown in thin gate and tunneling oxides. *IEEE Transactions On Electron Devices*, 32(2):413–422, February 1985.

[14] C-Y.Wu and C-F.Chen. Transport properties of thermal oxide films grown on polycrystalline silicon - modeling and experiments. *IEEE Transactions On Electron Devices*, 34(7):1590–1601, July 1987.

[15] I.Smith and R.Howland. Applications of scanning probe microscopy in the semiconductor industry. *Solid state technology*, 33(12):53–56, December 1990.

[16] S.K.Lai, V.K.Dham, and D.Guterman. Comparison and trends in today's dominant EE technologies. In *IEEE IEDM*, pages 580–583, 1986.

[17] H.E.Meas, J.Witters, and G.Groeseneken. Trends in non-volatile memory devices and technologies. In *ESSDERC Bologna*, pages 743–754, 1987.

[18] A.Kolodny, S.T.K.Nieh, B.Eitan, and J.Shappir. Analysis and modelling of floating-gate EEPROM cells. *IEEE Transactions On Electron Devices*, 33(6):835–844, 1986.

[19] C-F.Chen, C-Y.Wu, and M-K.Lee. The dielectric reliability of intrinsic thin SiO_2 films thermally grown on a heavily doped si substrate- characterization

and modeling. *IEEE Transactions On Electron Devices*, 34(7):1540–1552, 1987.

[20] Y.Hokari, T.Baba, and N.Kawamura. Reliability of 6-10nm thermal SiO_2 films showing intrinsic dielectric integrity. *IEEE Transactions On Electron Devices*, 32(11):2485–2491, November 1985.

[21] E.Rosenbaum, R.Rofan, and C.Hu. Effect of hot-carrier injection on n- and p-MOSFET gate oxide. *IEEE Electron Device Letters*, 12(11):599–601, November 1991.

[22] B.Root and M.Davis. *Reliability Testing Seminar 10-14 June 1991.* Masic Europe - Sienna Technologies.

[23] K.Yamabe and K.Taniguchi. Time dependent dielectric breakdown of thin thermally grown SiO_2 films. *IEEE Transactions On Electron Devices*, 32(2):423–428, February 1985.

[24] C.Kuo, Y.R.Yeargain, and W.J.Downey. An 80ns 32K EEPROM using the FETMOS cell. *IEEE J.Solid State Circuits*, (5):821–827, October 1982.

[25] S.K.Haywood, M.M.Heyns, and R.F.DeKeersmaeker. The statistics of dielectric breakdown in mos capacitors under static and dynamic voltage stress. In *Applied Surface Science, Insulating Films On semiconductors*, volume 30, pages 325–331, 1987.

[26] S.D.Brorson, D.J.DiMaria, and M.V.Fischetti. Direct measurement of the energy distrigution of hot electrons in silicon dioxide. *J.Appl.Phys*, 58(3):1302–1313, August 1985.

[27] E.A.Amerasekera and D.S.Campbell. *Failure Mechanisms In Semiconductor Devices.* J.Wiley and Sons, 1987.

[28] R.J.Allen and W.J.Roesch. Reliability prediction: The applicability of high temperature testing. *Solid State Technology*, 33(9):103–108, 1990.

[29] S.M.Sze, editor. *VLSI Technology*, chapter 14. McGraw-Hill International Editions, 1988.

[30] S.K.Haywood, M.M.Heyns, and R.F.DeKeersmaecker. The statistics of dielectric breakdown in MOS capacitors under static and dynamic voltage stress. In *Applied Surface Science*, pages 325–331, May 1987.

[31] W.S.Johnson, G.L.Kuhn, A.L.Renninger, and G.Perlegos. 16-K EE-PROM relies on tunneling for Byte-erasable program storage. *Electronics*, pages 113–117, February 1980.

[32] P.I.Suciu, B.P.Cox, D.D.Rinerson, and S.F.Cagnina. Cell model for EEPROM floating gate memories. In *IEEE IEDM*, pages 737–740, 1982.

[33] J.S.Witters, G.Groesenken, and H.E.Maes. Degradation phenomena of tunnel oxide floating gate EEPROM devices. In *IMEC*, pages 167–170, 1987.

[34] K.Y.Chang, S.Cheng, and K-M.Chang. An advanced high voltage CMOS process for custom logic circuits with embedded EEPROM. In *CICC*, 1988.

[35] G.Kaplan. Europower 92, how a united Europe plans to exploit technology to improve its trading posture. *IEEE Spectrum*, page 20, 1990.

Chapter 3

Derivation of a FETMOS Model

3.1 Overview

In the words of Einstein "A model should be as simple as possible, but no simpler". In adopting this approach to modelling, it is hoped to place interesting effects in the lime-light, without clouding results with more subtle (but unimportant) phenomena. Reliability is a key area [1], and it is proposed to model this, in terms of fundamental physical parameters, such as gate oxide thickness and gate length. It will then be possible to vary each parameter, and find the most influential. In this way a means to enhance FETMOS endurance can be found. The dependence of the threshold window upon fundamental parameters is also of interest. Even in a well established process the threshold window varies [1], and may stray outside limits required by accompanying circuitry. It is therefore desirable that the model can predict parameter variations, which can be used to restore the threshold window, to its original value. Finally, the model should also provide the capability to probe internal currents and fields, during program and erase operations.

Endurance is a key reliability issue for the EEPROM. However, no suitable methodology currently exists to access this. No model exists for the FETMOS either, although models have been developed for a variety of other EEPROM devices. Hence, a new methodology for calculating reliability and a new model for the FETMOS are required. Previously, models have been produced for both

[1] As noted in discussion with Motorola.

avalanche type electron injection devices [2], and Fowler-Nordheim type injection devices [3,4,5,6,7]. These models are all based on the capacitive equivalent circuit for the cell. However, only two of the models, [6] and [7], account for parasitic resistances and the ramped nature of the program/erase voltage. Inclusion of these parasitic values allows an accurate transient analysis to be made. This will be important in deriving a methodology to calculate reliability. Suciu's model for the FLOTOX [7] was deemed to be the best of these, since it includes parasitic factors most elegantly into the model equations. The FETMOS can be described by a capacitor network, which is equivalent to the FLOTOX. Therefore, Suciu's FLOTOX model may be used as the basis for the FETMOS model.

However, before deriving the model it will be useful to consider a simple capacitor network containing injected charge. Thus, assumptions made during the derivation will be have a solid foundation.

3.2 Distribution of Injected Charge in a Capacitor System

The FETMOS device is to be studied in terms of a lumped capacitor network, containing injected charge. Thus, information is required regarding the charge distribution. How much injected charge resides on each capacitor? Furthermore, will an applied voltage cause the injected charge to re-distribute itself? These questions can be answered by considering the simple case of two series capacitors.

Figure 3-1 shows two series capacitors *without* any injected charge. An applied voltage induces an equal charge on each capacitor. We have:

$$
\begin{aligned}
Q_1 &= Q_{1o} \\
Q_2 &= Q_{2o} \\
V_a &= V_1 + V_2 \qquad (3.1) \\
V_a &= \frac{Q_{1o}}{C_1} + \frac{Q_{2o}}{C_2}
\end{aligned}
$$

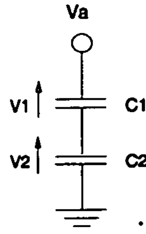

Figure 3–1: Two Series Capacitors, With an Applied Voltage.

$$V_a - \frac{Q_{1o}}{C_1} - \frac{Q_{2o}}{C_2} = 0 \tag{3.2}$$

Where:

- Q_1 = Net charge on C_1.
- Q_2 = Net Charge on C_2.
- Q_{1o} = Induced charge on C_1.
- Q_{2o} = Induced charge on C_2.
- V_a = Applied voltage.
- V_1 = Voltage across C_1.
- V_2 = Voltage across C_2.

Now consider the same system, *with* injected charge. The charge on C_1 will have changed by a factor Q_{1c}, and the charge on C_2 will have changed by a factor Q_{2c}. Here, Q_{1c} and Q_{2c} are arbitrary factors, whose value is unknown. Thus:

$$Q_1 = Q_{1o} + Q_{1c}$$

$$Q_2 = Q_{2o} + Q_{2c}$$

Dividing through by capacitance gives:

$$\frac{Q_1}{C_1} = \frac{Q_{1o}}{C_1} + \frac{Q_{1c}}{C_1}$$

$$V_1 = \frac{Q_{1o}}{C_1} + \frac{Q_{1c}}{C_1} \quad (3.3)$$

And:

$$\frac{Q_2}{C_2} = \frac{Q_{2o}}{C_2} + \frac{Q_{2c}}{C_2}$$

$$V_2 = \frac{Q_{2o}}{C_2} + \frac{Q_{2c}}{C_2} \quad (3.4)$$

Equations (3.3) and (3.4) may be substituted into equation (3.1), to give:

$$V_a = \frac{Q_{1o}}{C_1} + \frac{Q_{1c}}{C_1} + \frac{Q_{2o}}{C_2} + \frac{Q_{2c}}{C_2}$$

$$V_a - \frac{Q_{1o}}{C_1} - \frac{Q_{2o}}{C_2} = \frac{Q_{1c}}{C_1} + \frac{Q_{2c}}{C_2} \quad (3.5)$$

From equation (3.2), we know that the left hand side of equation (3.5) is zero. Thus:

$$0 = \frac{Q_{1c}}{C_1} + \frac{Q_{2c}}{C_2}$$

$$\frac{Q_{1c}}{C_1} = -\frac{Q_{2c}}{C_2}$$

$$\frac{Q_{1c}}{Q_{2c}} = -\frac{C_1}{C_2}$$

This equation is *unrelated* to the applied voltage. It is deduced that Q_{c1} and Q_{c2} are only due to the injected charge, which distributes itself in the ratio of the capacitors. Hence:

$$\frac{Q_{1i}}{Q_{2i}} = -\frac{C_1}{C_2} \quad (3.6)$$

Where:

- Q_{1i} = Amount of injected charge on C_1.
- Q_{2i} = Amount of injected charge on C_2.

The injected charge will occupy the bottom plate of C_1 and the top plate of C_2. This effectively adds charges of opposite polarity to each capacitor, and the equation therefore contains a minus sign. Note that these equations would not be

valid for a system in which one of the nodes was floating. If in equation (3.4) Q_{2c} is now replaced by Q_{2i}, we have:

$$V_2 = \frac{Q_{2o}}{C_2} + \frac{Q_{2i}}{C_2}$$

$$V_2 = V_{2o} + V_{2i} \tag{3.7}$$

Where:

- V_{2o} = Induced voltage on C_2.
- V_{2i} = Voltage on C_2 due to injected charge.

Hence, the voltages due to induced and injected charge add linearly in the capacitor system.

3.3 Derivation of Equations to Describe the FETMOS Device

3.3.1 Equivalent Capacitive Circuit for the FETMOS Device

Equivalent capacitive circuits are required for both the program and erase operations. To simplify the analysis only capacitances having the most significant effect on FETMOS operation are included [7], see figures 3–2 and 3–3. In the program operation charge may not flow onto C_{fs}, as the source is floating. It may therefore be argued that C_{fs} should be excluded from the modelling of the program operation. However $C_{fs} = 1.2\ fF$ which is insignificant compared to other parameters, eg. $C_{fg} = 54.6\ fF$ It will therefore be included in this derivation, to simplify equations.

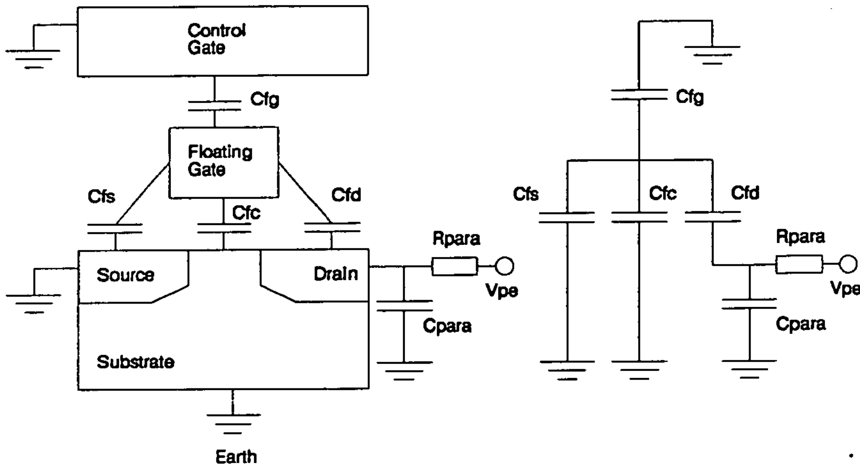

Figure 3–2: Equivalent Capacitor Circuit for a FETMOS Device, during the Program Operation.

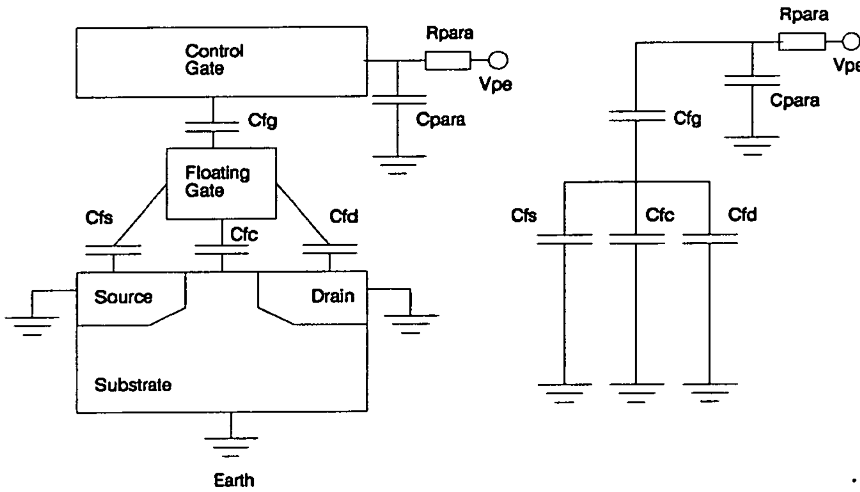

Figure 3–3: Equivalent Capacitor Circuit for FETMOS Device, during the Erase Operation.

In figures 3–2 and 3–3:

- C_{fg} = Capacitance between the control gate and floating gate.
- C_{fd} = Capacitance between the drain and floating gate.
- C_{fs} = Capacitance between the source and floating gate.
- C_{fc} = Capacitance between the channel and floating gate.
- C_{para} = Parasitic capacitance.
- R_{para} = Parasitic resistance.
- V_{pe} = Applied program/erase step pulse.

3.3.2 Coupling Ratios

The program/erase voltage, V_{pe}, generates a high electric field across the tunnel oxide. The proportion of V_{pe} which falls across the tunnel oxide is defined as the coupling ratio [6]. Let:

$$C_t = C_{fg} + C_{fs} + C_{fc} + C_{fd}$$

Then the erase coupling ratio is given by:

$$\frac{C_{fg}}{C_t}$$

The program coupling ratio is given by:

$$\frac{C_t - C_{fd}}{C_t}$$

The higher the coupling ratio, the higher the field across the tunnel oxide.

3.3.3 The Electric Field as a Function of Time

A set of equations are required, which give the electric field, E, across the tunnel oxide, as a function of time.

The Effect of the Voltage Ramp and Parasitics

Rather than model the voltage on the drain or gate as a unit step, an exponential rise is used. This waveshape represents realistically changing voltages in a memory circuit [8], and is included by adding an RC time constant τ. We have:

$$\tau = R_{para}C_{para}$$

For the program operation the potential of the drain, V_d, is given by:

$$V_d = V_{pe}(1 - exp^{-\frac{t}{\tau}})$$

$$V_d = V_{pe} - V_{pe}exp^{-\frac{t}{\tau}} \tag{3.8}$$

For the erase operation the potential of the gate, V_g, is given by:

$$V_g = V_{pe}(1 - exp^{-\frac{t}{\tau}})$$

$$V_g = V_{pe} - V_{pe}exp^{-\frac{t}{\tau}} \tag{3.9}$$

The Potential of the Floating Gate

For a tunnel oxide of thickness X_o, the voltage across it is given by $E\,X_o$. Thus, for the program operation the potential of the floating gate, V_f, is given by:

$$V_f = V_d - EX_o \tag{3.10}$$

For the erase operation the potential of the floating gate is given by:

$$V_f = EX_o \tag{3.11}$$

The Tunnelling Current

Charge conservation tells us that [7]:

$$I = \frac{dQ_i}{dt}$$

Also:

$$I = PJ$$

Hence, for the program operation:

$$\frac{dQ_i}{dt} = PJ \tag{3.12}$$

For the erase operation

$$\frac{dQ_i}{dt} = -PJ \tag{3.13}$$

Where:

- I = Tunnelling current.
- J = Tunnelling current density.
- P = Tunnelling area.
- Q_i = Net injected charge on the floating gate.

Fowler-Nordheim Tunnelling Equation

The tunnelling current may be described by the Fowler-Nordheim equation (3.14) where A and B are constants [3].

$$J = AE^2 exp^{-\frac{B}{E}} \tag{3.14}$$

Net Injected Charge

The equivalent circuits of figures 3–2 and 3–3 may each be reduced to the form of two series capacitors, as in figure 3–4. The amount of charge injected into the system, is equal to the difference in the charge on each capacitor.

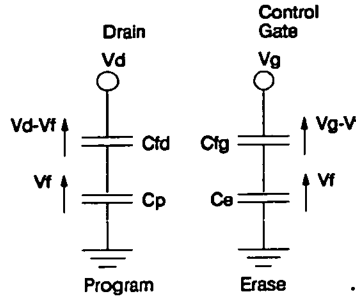

Figure 3–4: Reduced Forms of the Equivalent Capacitive Circuits for the FETMOS Device.

Where:

- $C_p = C_{fg} + C_{fs} + C_{fc} =$ "Program Capacitance".
- $C_e = C_{fd} + C_{fs} + C_{fc} =$ "Erase Capacitance".
- $C_t = C_{fd} + C_{fs} + C_{fg} + C_{fc} =$ "Total capacitance".
- Q_i =Net injected charge.
- Q_{cfd} =Charge on C_{fd}.
- Q_{cfg} =Charge on C_{fg}.
- Q_{cp} =Charge on C_p.
- Q_{ce} =Charge on C_e.

In the program operation positive charge is injected onto the floating gate. This is given by:

$$Q_i = Q_{cp} - Q_{cfd} \tag{3.15}$$

Now:

$$Q_{cp} = V_f C_p$$

$$Q_{cfd} = (V_d - V_f)C_{fd}$$

Substituting for Q_{cp} and Q_{cfd} in equation (3.15) gives:

$$Q_i = V_f C_p - (V_d - V_f)C_{fd}$$

$$Q_i = V_f(C_p + C_{fd}) - V_d C_{fd}$$

$$Q_i = V_f C_t - V_d C_{fd} \tag{3.16}$$

In the erase operation negative charge is injected onto the floating gate. This is given by:

$$Q_i = Q_{ce} - Q_{cfg} \tag{3.17}$$

Now:

$$Q_{ce} = V_f C_e$$

$$Q_{cfg} = (V_g - V_f)C_{fg}$$

Substituting for Q_{cfg} and Q_{ce} in equation (3.17) gives:

$$Q_i = V_f C_e - (V_g - V_f)C_{fg}$$

$$Q_i = V_f(C_e + C_{fg}) - V_g C_{fg}$$

$$Q_i = V_f C_t - V_g C_{fg} \tag{3.18}$$

Programming Field as a Function of Time

These equations are now *combined* to give programming field as a function of time. Recalling equation (3.16):

$$Q_i = V_f C_t - V_d C_{fd}$$

Substituting for V_f using equation (3.10) gives:

$$Q_i = (V_d - EX_o)C_t - V_d C_{fd}$$

Substituting for V_d using equation (3.8) gives:

$$Q_i = \left(\left(V_{pe} - V_{pe}exp^{-\frac{t}{\tau}}\right) - EX_o\right)C_t - \left(V_{pe} - V_{pe}exp^{-\frac{t}{\tau}}\right)C_{fd}$$

$$Q_i = V_{pe}C_t - V_{pe}exp^{-\frac{t}{\tau}}C_t - EX_oC_t - V_{pe}C_{fd} + V_{pe}exp^{-\frac{t}{\tau}}C_{fd}$$

$$Q_i = V_{pe}(C_t - C_{fd}) - V_{pe}exp^{-\frac{t}{\tau}}(C_t - C_{fd}) - EX_oC_t$$

Now from equation (3.12), where P_a is program tunnelling area:

$$\frac{dQ_i}{dt} = P_aJ$$

Substituting for Q_i gives:

$$\frac{d}{dt}\left(V_{pe}(C_t - C_{fd})\right) - \frac{d}{dt}\left(V_{pe}exp^{-\frac{t}{\tau}}(C_t - C_{fd})\right) - \frac{d}{dt}\left(EX_oC_t\right) = P_aJ$$

Differentiating with respect to t gives:

$$V_{pe}exp^{-\frac{t}{\tau}}\left(\frac{C_t - C_{fd}}{\tau}\right) - \frac{dE}{dt}X_oC_t = P_aJ$$

Substituting for J using equation (3.14) gives:

$$V_{pe}exp^{-\frac{t}{\tau}}\left(\frac{C_t - C_{fd}}{\tau}\right) - \frac{dE}{dt}X_oC_t = P_aAE^2exp^{-\frac{B}{E}}$$

Rearranging gives:

$$\frac{dE}{dt} = V_{pe}exp^{-\frac{t}{\tau}}\left(\frac{C_t - C_{fd}}{X_oC_t\tau}\right) - \frac{P_aAE^2}{X_oC_t}exp^{-\frac{B}{E}}$$

$$\frac{dE}{dt} = \frac{V_{pe}}{X_o\tau}exp^{-\frac{t}{\tau}}\left(1 - \frac{C_{fd}}{C_t}\right) - \frac{P_aAE^2}{X_oC_t}exp^{-\frac{B}{E}} \qquad (3.19)$$

This is equivalent to Suciu's equation, derived for *discharging* the FLOTOX cell [2]. It is a first order non-linear differential equation, and a program was written to solve this numerically, see appendix B. The 4th order Runge-Kutta method was used for this solution [9].

---

[2]Due to a typographical error, a τ was missed out in reference [7].

Erasing Field as a Function of Time

These equations are combined to give the erasing field, as a function of time. Recalling equation (3.18):

$$Q_i = V_f C_t - V_g C_{fg}$$

Substituting for V_f using equation (3.11) gives:

$$Q_i = E X_o C_t - V_g C_{fg}$$

Substituting for V_g using equation (3.9) gives:

$$Q_i = E X_o C_t - \left(V_{pe} - V_{pe} exp^{-\frac{t}{\tau}}\right) C_{fg}$$

$$Q_i = E X_o C_t - V_{pe} C_{fg} + V_{pe} exp^{-\frac{t}{\tau}} C_{fg}$$

Now from equation (3.13), where E_a is the erase tunnelling area:

$$\frac{dQ_i}{dt} = -E_a J$$

Substituting for Q_i gives:

$$\frac{d}{dt}\left(-V_{pe} C_{fg}\right) + \frac{d}{dt}\left(V_{pe} exp^{-\frac{t}{\tau}} C_{fg}\right) + \frac{d}{dt}\left(E X_o C_t\right) = -E_a J$$

Differentiating with respect to t gives:

$$-V_{pe} exp^{-\frac{t}{\tau}} \frac{C_{fg}}{\tau} + \frac{dE}{dt} X_o C_t = -E_a J$$

Substituting for J using equation (3.14) gives:

$$-V_{pe} exp^{-\frac{t}{\tau}} \frac{C_{fg}}{\tau} + \frac{dE}{dt} X_o C_t = -E_a A E^2 exp^{-\frac{B}{E}}$$

Rearranging gives:

$$\frac{dE}{dt} = \frac{V_{pe} C_{fg}}{X_o C_t \tau} exp^{-\frac{t}{\tau}} - \frac{E_a A E^2}{X_o C_t} exp^{-\frac{B}{E}} \qquad (3.20)$$

This is equivalent to Suciu's equation for *charging* [7]. Again, a program was written to solve this numerically, see appendix B.

3.3.4 Threshold Voltage as a Function of Electric Field

Producing a General Equation for the Threshold Voltage

A general expression for threshold voltage will be derived, which may then be applied to the program and erase operations separately. Figure 3–5 gives the equivalent capacitive circuits, for the program and erase operations.

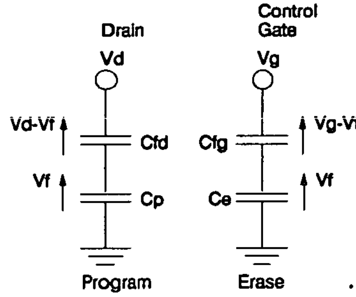

Figure 3–5: Equivalent Capacitive Circuits for the FETMOS Device.

For a FETMOS device with a voltage on the drain, the capacitors will each have the same induced charge, hence:

$$Q_{cfd} = Q_{cp}$$

$$C_{fd}(V_d - V_f) = C_p V_f$$

$$V_f = V_d \left(\frac{C_{fd}}{C_{fd} + C_p} \right) = V_d \left(\frac{C_{fd}}{C_t} \right) \tag{3.21}$$

For a FETMOS device with a voltage on the control gate, the capacitors will have induced charge:

$$Q_{cfg} = Q_{ce}$$

$$C_{fg}(V_g - V_f) = C_e V_f$$

$$V_f = V_g \left(\frac{C_{fg}}{C_e + C_{fg}} \right) = V_g \left(\frac{C_{fg}}{C_t} \right) \tag{3.22}$$

Voltages add linearly in a FETMOS device. Thus, combining equations (3.21) and (3.22) gives:

$$V_f = V_d \frac{C_{fd}}{C_t} + V_g \frac{C_{fg}}{C_t}$$

Add to this the effect of injected charge and we have:

$$V_f = V_d \frac{C_{fd}}{C_t} + V_g \frac{C_{fg}}{C_t} + V_{fo}$$

$$0 = V_d \frac{C_{fd}}{C_t} + V_g \frac{C_{fg}}{C_t} + V_{fo} - V_f \tag{3.23}$$

Where V_{fo} is the potential of the floating gate due to injected charge. In the threshold regime we may re-write equation (3.22) as:

$$V_{ft} = V_{to} \left(\frac{C_{fg}}{C_t} \right) \tag{3.24}$$

Where V_{ft} is the potential of the floating gate when the FETMOS is in the threshold regime, and V_{to} is the threshold voltage of the FETMOS *without* any injected charge. If injected charge is added by Fowler-Nordheim tunnelling, a term must be added to equation (3.24), and we have:

$$V_{ft} = V_t \left(\frac{C_{fg}}{C_t} \right) + V_{fo} \tag{3.25}$$

Where V_t is the threshold voltage of the FETMOS device. Equating equations (3.24) and (3.25) gives:

$$V_t \left(\frac{C_{fg}}{C_t} \right) + V_{fo} = V_{to} \left(\frac{C_{fg}}{C_t} \right)$$

$$V_t \left(\frac{C_{fg}}{C_t} \right) + V_{fo} - V_{to} \left(\frac{C_{fg}}{C_t} \right) = 0 \tag{3.26}$$

Equating equations (3.23) and (3.26) gives:

$$V_t \left(\frac{C_{fg}}{C_t} \right) + V_{fo} - V_{to} \left(\frac{C_{fg}}{C_t} \right) = V_d \left(\frac{C_{fd}}{C_t} \right) + V_{fo} - V_f + V_g \left(\frac{C_{fg}}{C_t} \right)$$

$$V_t - V_{to} = V_d \left(\frac{C_{fd}}{C_{fg}} \right) - V_f \left(\frac{C_t}{C_{fg}} \right) + V_g$$

$$V_t = V_d \left(\frac{C_{fd}}{C_{fg}} \right) - V_f \left(\frac{C_t}{C_{fg}} \right) + V_g + V_{to} \tag{3.27}$$

Equation (3.27) provides a general expression for threshold voltage, which may be applied to the program and erase cases separately.

Erased Threshold Voltage

During the erase operation $V_d = 0$, and equation (3.27) becomes:

$$V_t = -V_f\left(\frac{C_t}{C_{fg}}\right) + V_g + V_{to}$$

Substituting for V_f using equation (3.11) gives:

$$V_t = -EX_o\left(\frac{C_t}{C_{fg}}\right) + V_g + V_{to}$$

Substituting for V_g using equation (3.9) gives:

$$V_t = -EX_o\left(\frac{C_t}{C_{fg}}\right) + V_{pe}\left(1 - exp^{\frac{-t}{\tau}}\right) + V_{to} \qquad (3.28)$$

The electric field E may be calculated from equation (3.20) and substituted into equation (3.28) to give V_t, after any duration of erasing pulse.

Programmed Threshold Voltage

During the program operation $V_g = 0$, and equation (3.27) becomes:

$$V_t = V_d\left(\frac{C_{fd}}{C_{fg}}\right) - V_f\left(\frac{C_t}{C_{fg}}\right) + V_{to}$$

Substituting for V_f using equation (3.10) gives:

$$V_t = V_d\left(\frac{C_{fd}}{C_{fg}}\right) - (V_d - EX_o)\left(\frac{C_t}{C_{fg}}\right) + V_{to}$$

$$V_t = V_d\left(\frac{C_{fd}}{C_{fg}}\right) - V_d\left(\frac{C_t}{C_{fg}}\right) + EX_o\left(\frac{C_t}{C_{fg}}\right) + V_{to}$$

$$V_t = V_d\left(\frac{C_{fd} - Ct}{C_{fg}}\right) + EX_o\left(\frac{C_t}{C_{fg}}\right) + V_{to}$$

Substituting for V_d using equation (3.8) gives:

$$V_t = V_{pe}\left(1 - exp^{\frac{-t}{\tau}}\right)\left(\frac{C_{fd} - Ct}{C_{fg}}\right) + EX_o\left(\frac{C_t}{C_{fg}}\right) + V_{to} \qquad (3.29)$$

Again the electric field E may be calculated from 3.19 and substituted into equation (3.29) to give V_t, after any duration of programming pulse.

3.3.5 Initial Electric Field

The electric field across the tunnel oxide prior to a program or erase operation, E_i, forms a boundary condition for the solution of $\frac{dE}{dt}$. If V_{ti} is the threshold voltage at the beginning of a program/erase operation, then for the erase case equation (3.28) gives:

$$V_t = -EX_o\left(\frac{C_t}{C_{fg}}\right) + V_{pe}\left(1 - exp^{\frac{-t}{\tau}}\right) + V_{to}$$

At $t = 0$, $(1 - exp^{\frac{-t}{\tau}}) = 0$. This gives:

$$V_t = -EX_o\left(\frac{C_t}{C_{fg}}\right) + V_{to}$$

$$V_{ti} = -E_iX_o\left(\frac{C_t}{C_{fg}}\right) + V_{to}$$

$$E_i = \left(\frac{C_{fg}}{X_oC_t}\right)(V_{to} - V_{ti}) \tag{3.30}$$

For the program case equation (3.29) gives:

$$V_t = V_{pe}\left(1 - exp^{\frac{-t}{\tau}}\right)\left(\frac{C_{fd} - Ct}{C_{fg}}\right) + EX_o\left(\frac{C_t}{C_{fg}}\right) + V_{to}$$

Again at $t = 0$, $(1 - exp^{\frac{-t}{\tau}}) = 0$, hence:

$$V_t = EX_o\left(\frac{C_t}{C_{fg}}\right) + V_{to}$$

$$V_{ti} = E_iX_o\left(\frac{C_t}{C_{fg}}\right) + V_{to}$$

$$E_i = -\left(\frac{C_{fg}}{X_oC_t}\right)(V_{to} - V_{ti}) \tag{3.31}$$

3.3.6 Summary of Equations

We require a set of equations to model the FETMOS in terms of fundamental design parameters, such as effective width. The electric field E, across the tunnel oxide is given by equations (3.19) and (3.20), for programming and erasing respectively.

$$\frac{dE}{dt} = \frac{V_{pe}}{X_o\tau}exp^{-\frac{t}{\tau}}\left(1 - \frac{C_{fd}}{C_t}\right) - \frac{P_aAE^2}{X_oC_t}exp^{-\frac{B}{E}}$$

$$\frac{dE}{dt} = \frac{V_{pe}C_{fg}}{X_oC_t\tau}exp^{-\frac{t}{\tau}} - \frac{E_aAE^2}{X_oC_t}exp^{-\frac{B}{E}}$$

These first order non-linear differential equations may be solved numerically. The necessary boundary conditions are the initial time $\simeq 0$ seconds, and the initial electric field E_i across the tunnel oxide. Equations (3.31) and (3.30) give E_i for programming and erasing respectively.

$$E_i = -\left(\frac{C_{fg}}{X_oC_t}\right)(V_{to} - V_{ti})$$

$$E_i = \left(\frac{C_{fg}}{X_oC_t}\right)(V_{to} - V_{ti})$$

Tunnelling current density is given by:

$$J = AE^2exp^{-\frac{B}{E}}$$

The threshold voltage V_t, may be calculated by substituting the electric field E into equations (3.29) and (3.28) for programming and erasing respectively.

$$V_t = V_{pe}\left(1 - exp^{\frac{-t}{\tau}}\right)\left(\frac{C_{fd} - Ct}{C_{fg}}\right) + EX_o\left(\frac{C_t}{C_{fg}}\right) + V_{to}$$

$$V_t = -EX_o\left(\frac{C_t}{C_{fg}}\right) + V_{pe}\left(1 - exp^{\frac{-t}{\tau}}\right) + V_{to}$$

Finally program tunnel area P_a, erase tunnel area E_a, and capacitances, can be calculated as follows;

$$P_a = GD_{over} \times W_{eff}$$

$$E_a = L_g \times W_{eff}$$

$$C_{fg} = \frac{A_{fg}\varepsilon_{oxide}\varepsilon_o}{X_{int}}$$

$$C_{fd} = C_{fs} = \frac{W_{eff}GD_{over}\varepsilon_{oxide}\varepsilon_o}{X_o}$$

$$C_{fc} = \frac{W_{eff}(L_g - 2 \times GD_{over})\varepsilon_{oxide}\varepsilon_o}{X_o}$$

Where:

- GD_{over} is the floating gate/drain overlap.

- A_{FG} is the floating gate area.
- L_g is the gate length.
- W_{eff} is the effective width.
- X_o is the tunnel oxide thickness.
- X_{int} is the inter-level oxide thickness,.
- ε_{oxide} is the relative permittivity of S_iO_2.
- ε_o is the primary electric constant.

Thus we have a set of equations for FETMOS analysis, in terms of basic design parameters.

3.4 Calculation of FETMOS Parameters

3.4.1 Overview

A set of parameters must be derived for the FETMOS device. On wafers containing the HC11 microcontroller [10], test structures are provided within the scribe grid. Of these, test structure "*SGPC8#*" [3] provides FETMOS devices with access to the gate, drain, source and substrate. In addition, "*SGPC8#*" provides a $2.5 \times 10^{-4} cm^2$ tunnel oxide capacitor, for Fowler-Nordheim measurements. Many of the parameters are fixed, eg. dimensions drawn on reticle, while other vary during processing, eg. tunnel oxide thickness. Now, tunnelling during program/erase operations is strongly dependent on oxide thickness, X_o. Hence a sample of devices with uniform X_o was needed, to allow accurate fitting of the model to experimental

[3]Described in Motorola internal documentation.

data. The uniformity of X_o, and indeed of the Fowler-Nordheim tunnelling coefficients, can be optimised by taking results from a single wafer . This wafer was taken from a batch showing no EEPROM defects, and having a typical threshold window.

3.4.2 Geometry of the FETMOS Device

The FETMOS dimensions are as follows:

- Gate width as drawn on the reticle...$W_d = 2.5\mu m$

- Field oxide thickness...$X_f = 0.6\mu m$

- Effective gate width...$W_{eff} = W_d - (2 \times X_f) = 1.3\mu m$
  The width is reduced due to "Bird's Beaking" [11] - the encroachment of field oxide on either side of the gate.

- Floating gate/drain overlap...$GD_{over} = 0.3\mu m$
  The source/drain implantation diffuses laterally during fabrication, to give the gate-drain overlap. This will be in the order of $0.3\mu m$ [12].

- Gate length as drawn on the reticle...$L_g = 2.8\ \mu m$

- Channel length...$L_{chan} = L_g - (2 \times GD_{over}) = 2.2\mu m$
  This is approximately equal to the gate length, but the gate-drain and gate-source overlaps must be subtracted.

- Tunnel oxide thickness...$X_o = 108\ \AA$
  This was measured using an automatic ellipsometer, with an accuracy better than $\pm 10\AA$ [13].

- Interlevel oxide thickness...X_{int} =400 $\AA$
  This is also measured using automatic ellipsometery.

- Floating gate area...$A_{fg} = 50\mu m^2$

  An optical micrograph was taken of the floating gate and surrounding area. The dimension of a large feature (the distance between contact holes) was measured using a Vicker Photoplan. This provided calibration, from which the dimensions and area of the floating gate were calculated.

- Tunnel area during programming...$P_a = GD_{over} \times W_{eff} = 0.39\mu m^2$

  During programming, tunnel current flows between the gate and drain.

- Tunnel area during erasing...$E_a = L_g \times W_{eff} = 3.64\mu m^2$

  During erasing, tunnel current flows across the entire length of the gate.

3.4.3 Calculation of Capacitances

Capacitances may be calculated from the dimensions of the FETMOS device.

$$C_{fg} = \frac{A_{fg}\varepsilon_{oxide}\varepsilon_o}{X_{int}}$$

$$C_{fg} = \frac{50\mu^2 \times 3.9 \times 8.85 \times 10^{-12}}{400\mathring{A}} = 43.1fF$$

$$C_{fd} = \frac{W_{eff}GD_{over}\varepsilon_{oxide}\varepsilon_o}{X_o}$$

$$C_{fd} = \frac{1.3\mu \times 0.3\mu \times 3.9 \times 8.85 \times 10^{-12}}{108\mathring{A}} = 1.2fF$$

$$C_{fs} = C_{fd} = 1.2fF$$

$$C_{fc} = \frac{W_{eff}L_{chan}\varepsilon_{oxide}\varepsilon_o}{X_o}$$

$$C_{fc} = \frac{1.3\mu \times 2.2\mu \times 3.9 \times 8.85 \times 10^{-12}}{108\mathring{A}} = 9.1fF$$

$$C_t = C_{fg} + C_{fd} + C_{fs} + C_{fc} = 54.6fF$$

3.4.4 Measurement of Fowler-Nordheim Coefficients

The Fowler-Nordheim coefficients, A and B, may be extracted from I-V characteristics of the tunnel oxide. The two major problems in determining these characteristics are oxide thickness measurement and charge trapping [6]. Oxide thickness was measured using an automatic ellipsometer, with an accuracy better than $\pm 10\AA$. The I-V characteristics will vary, depending upon how much charge has been trapped in the oxide. In common with previous authors, $0.1Ccm^{-2}$ of charge were passed through the oxide before coefficient measurement [6]. This is sufficient to allow saturation of positive charge trapping, and the rate of electron trapping will remain relatively small during any subsequent endurance analysis. An electric field of $11MVcm^{-1}$ was used to force this charge, which is of the same order as the field during program/erase [7]. Figure 3–6 gives the experimental set-up, including a Hewlett-Packard 4145B Semiconductor Parameter Analyser, controlled from a Hewlett-Packard Series 300 computer. To reduce noise, devices were packaged, and held in a Hewlett-Packard 16085 Test Fixture [14]. Ten sites on "$SGPC8\#$" were tested, and an example of the I-V characteristics is given in figure 3–7.

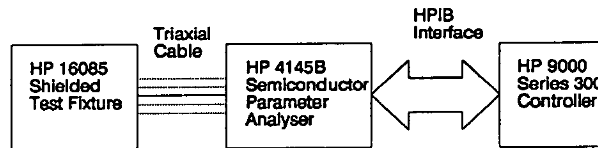

Figure 3–6: Block Scheme of Experimental Set-Up.

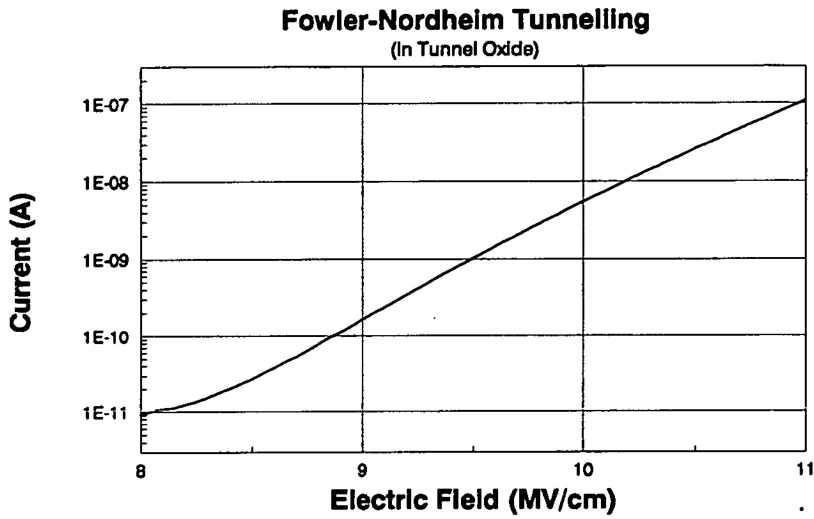

Figure 3–7: Fowler-Nordheim Current in an Oxide Capacitor of Area $2.5 \times 10^{-4} cm^2$.

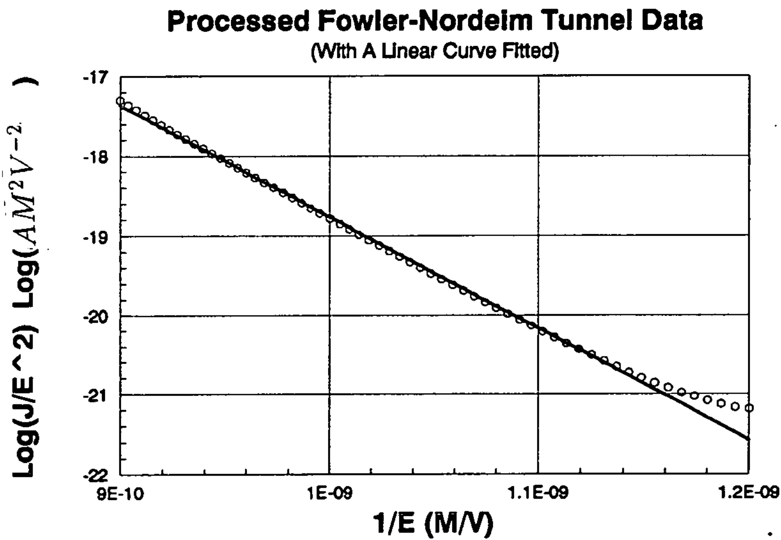

Figure 3–8: Manipulated Data from Fowler-Nordheim Tunnelling Plot of Oxide.

The resulting data was then manipulated, to produce a graph of $log_{10}J/E^2$ against 1/E, as shown in figure 3–8. To this a linear graph may be fitted, with:

$$\text{Slope...}M = -1.22 \times 10^8\, Vcm^{-1}$$

$$\text{“X – intercept”...}C = -5.65\, log(AV^{-2})$$

This may be related to the equation for Fowler-Nordheim tunnelling as follows:

$$J = AE^2 exp^{\frac{-B}{E}}$$

$$\frac{J}{E^2} = Aexp^{\frac{-B}{E}}$$

Taking logarithms of each side:

$$log_{10}\left(\frac{J}{E^2}\right) = log_{10}(A) + log_{10}\left(exp^{\frac{-B}{E}}\right)$$

$$log_{10}\left(\frac{J}{E^2}\right) = \frac{-B}{E} log_{10}(exp^1) + log_{10}(A)$$

Comparing this to the general expression for a straight line ($y = mx + c$) we have:

$$\text{“X – intercept”...}C = log_{10}(A)$$

$$\text{Slope...}M = -B\, log_{10}(exp^1)$$

Thus:

$$A = 10^C = 10^{-5.65} = 2.2 \times 10^{-6}\, AV^{-2}$$

$$B = -\frac{M}{log_{10}(exp^1)} = -\frac{1.22 \times 10^{10}}{log_{10}(exp^1)} = 2.8 \times 10^8\, Vcm^{-1}$$

These values compare with A=$1.88 \times 10^{-6} AV^{-2}$ and B=$2.55 \times 10^8 Vcm^{-1}$, calculated by previous authors [6]. Since, current flows in different directions during the program and erase operations, two sets of tunnel coefficients are required (one for each interface). To account for this, the value of A used for erase modelling is doubled, as by previous authors [7]. Hence:

- For programming: $A_{prg} = 2.2 \times 10^{-6}\, AV^{-2}$ and $B_{prg} = 2.8 \times 10^8\, Vcm^{-1}$
- For erasing: $A_{ers} = 4.4 \times 10^{-6}\, AV^{-2}$ and $B_{prg} = 2.8 \times 10^8\, Vcm^{-1}$

3.4.5 Measurement of Threshold Voltage

The gate voltage at which the transistor first begins to conduct may be given as the threshold voltage [15]. A less ambiguous criterion would be to say that the channel region, at the silicon surface, should as strongly n-type as the substrate is p-type [16]. This requires a bending of the Fermi level from its position above the intrinsic Fermi level in the substrate, to a position of equal distance below the intrinsic Fermi level, at the silicon surface. This is the widely used "$2\Phi_f$" criterion, which may also be used to define "strong inversion" [16], where Φ_f is the potential between the intrinsic and extrinsic Fermi levels.

Threshold voltage measurement is complicated by the physics of the MOS device, which does not turn on abruptly. The simplest technique, is to measure the gate voltage required for a predetermined drain current to flow, arbitrarily taken ~1μA [17]. However, the parameter values used must give the minimum error between model predictions and experiment. Therefore, it was decided to assume the SPICE standard transistor model for threshold voltage determination [17]. For measurement of V_{to}, this may be implemented by applying a low voltage to the drain, $\simeq 0.1V$ [17], and sweeping the gate from -6V to 6V. Figure 3–9 illustrates such a plot. A tangent is fitted to this curve where the slope is maximum. Projecting this tangent back to the horizontal axis, we find a value of gate voltage at which the gate current is zero, V_{g0}. The threshold voltage is then given by equation (3.32) [17]:

$$V_t = V_{g0} - \frac{V_{ds}}{2} \tag{3.32}$$

The principle advantage of this technique is that the value of V_t can be considered to be independent of the device width or length. It is also includes a degree of physical meaning, since it is derived from the equation for a MOSFET in the linear regime [18]:

$$I_{ds} = \mu C_o \frac{b}{a} V_{ds} \left(V_{gs} - V_t - \frac{V_{ds}}{2} \right)$$

$$\frac{I_{ds}\, a}{\mu\, C_o\, b\, V_{ds}} = \left(V_{gs} - V_t - \frac{V_{ds}}{2} \right)$$

If $I_{ds} = 0$, this can be reduced to the form of equation (3.32), where:

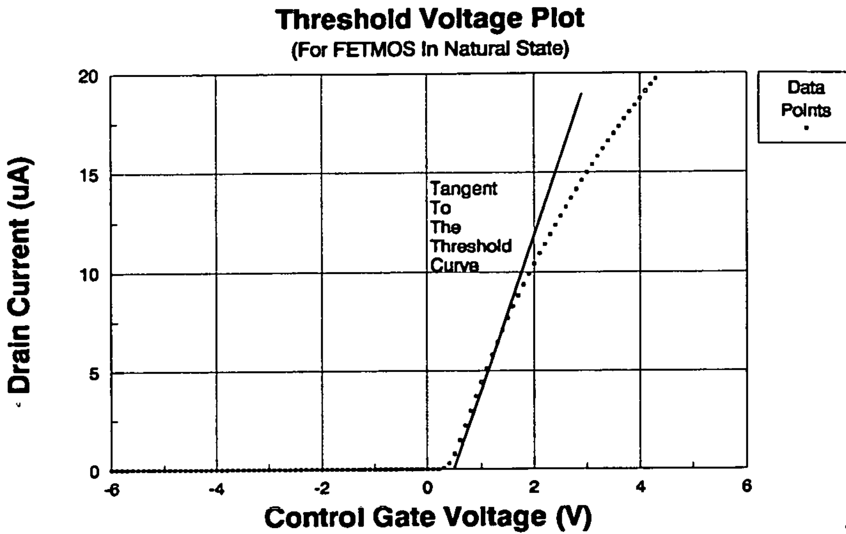

Figure 3–9: Drain Current as a Function of Control Gate Voltage, for Calculation of Threshold Voltage.

- I_{ds} = Drain current.
- μ = Surface mobility.
- C_o = Gate oxide capacitance.
- $\frac{b}{a}$ = Aspect ratio (effective width/ channel length).

Measurements were made using a Hewlett-Packard 4145B Semiconductor Parameter Analyser and a Wentworth Laboratories Manual Probe Station, as illustrated in figure 3–10. Note that when measuring a *programmed* or *erased* device, it is important to limit read disturb error. This can be minimised by beginning each measurement at $0V$, and sweeping to either $-10V$ or $+10V$, for programmed or erased devices respectively. A program was written to carry out the measurement and extract threshold voltages, using the SPICE algorithm. This was written in HP BASIC 5.1 and is given in appendix C. To ensure that hole trapping in the oxide had saturated, each FETMOS was cycled 20 times prior to measurement, this feature was included in the BASIC program. Results were taken from 10 random sites across a single wafer, yielding average threshold voltages of:

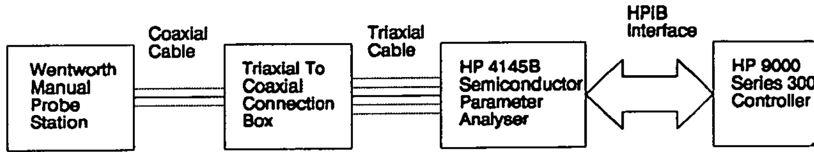

Figure 3–10: Block Scheme of Experimental Set-Up.

- $V_{to} = 0.5V$
- Programmed threshold voltage, $V_{tp} = -7.5V$
- Erased threshold voltage, $V_{te} = 5.5V$

These values compare favourably with designed threshold voltages of $V_{to} = 0V$, $V_{tp} = -5V$ and $V_{te} = +5V$ [19].

3.4.6 Calculation of RC Time Constant, τ

It was intended to compare model results with experimentally obtained threshold values. The RC time constant of the electrical measurement set up was therefore used. The largest resistance in the circuit is that of the FETMOS source/drain extension regions:

- Sheet resistance = 50 $\Omega/\square$
  As given Motorola internal documentation.
- Area = $1.3\mu m \times 10\mu m$ = 8 squares.
  Dimensions were taken from an optical micrograph.

This gives a source/drain extension resistance of $50 \times 8 = 800\Omega$. The largest parasitic capacitance in the circuit was that of the chuck, estimated to be $\sim 0.2\mu F$. This gave an estimated time constant of $800 \times 0.2\mu F \sim 0.1ms$, as compared to $0.4ms$ used by previous authors [7]. Figure 3–11 illustrates the variation in modelled threshold window, as a function of τ. The values are insensitive to τ,

provided τ lies within the range $20\mu s$ to $1ms$. This indicates that there is a wide leeway in acceptable values of τ.

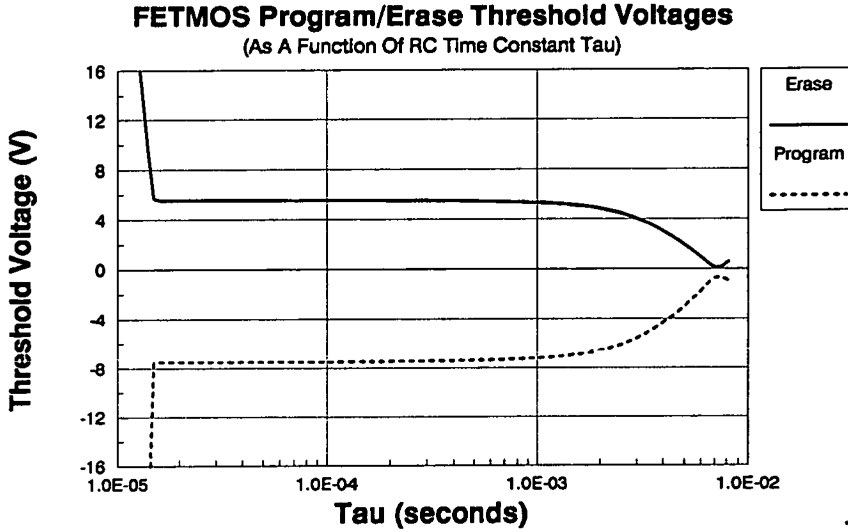

Figure 3–11: Threshold Window as a Function of the RC time constant τ, for a programming time of $10ms$.

3.5 Verification of the FETMOS Model Against Experimental Results

3.5.1 Overview

The methodology used to produce this model has already been successfully applied to the FLOTOX device [7]. Equally, a good agreement between experimental and modelled data might be expected here. Solution of the model equations using the parameters calculated above, gave $V_{te} = 5.536\ V$ and $V_{tp} = -7.448\ V$, compared to average experimental values of $V_{te} = 5.8\ V$ and $V_{tp} = -7.5\ V$. The error in the modelled results is $< 5\%$, testifying to the models validity. Further evidence is supplied in chapter 4, where the tunnel current during program/erase operations, is seen to have the same form as that obtained experimentally [20].

3.5.2 Comparison of Model Predictions and Experimental Results

The model is intended for investigation of the threshold window, as a function of varying parameters. A test was therefore made, to compare *predicted* trends in the threshold window, with experimental values. To provide a large number of well defined data points, the program/erase voltage (V_{pe}) was used as a variable. A minimum V_{pe} of 14V was used. This lies at the lower limit of FETMOS operation, where the threshold window has nearly closed. The standard operating voltage of 18V was set as the upper limit. Voltages above this were rejected, as they would lead to the onset of reliability problems - such as rupture of the tunnel oxide, and tunnelling in the interlevel dielectric.

A 10ms program/erase time was used, in common with commercial circuits containing the FETMOS [19]. Devices were given 10 program/erase cycles before hand, to saturate hole traps. Results were taken at 10 sites over a single wafer, and threshold voltages were calculated assuming the standard SPICE transistor model [17]. Measurements were made using a Hewlett-Packard 4145B Semiconductor Parameter Analyser and a Wentworth Laboratories Manual Probe Station, as illustrated in figure 3–12. A program was written in HP Basic 5.1 to measure the threshold window for a range of program/erase voltages, this is given in appendix C. A spread is observed in experimental data, as illustrated in figure 3–13. This is due to a variation in parameters such as oxide thicknesses, impurity doping levels and effective dimensions. Even so, trends in modelled program and erase thresholds, largely match experimental results.

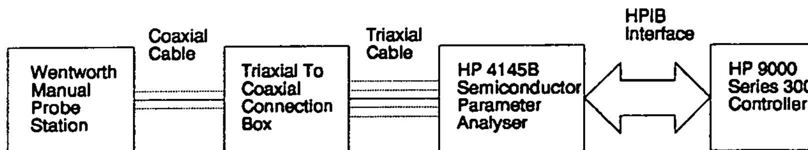

Figure 3–12: Block Scheme of Experimental Set-Up.

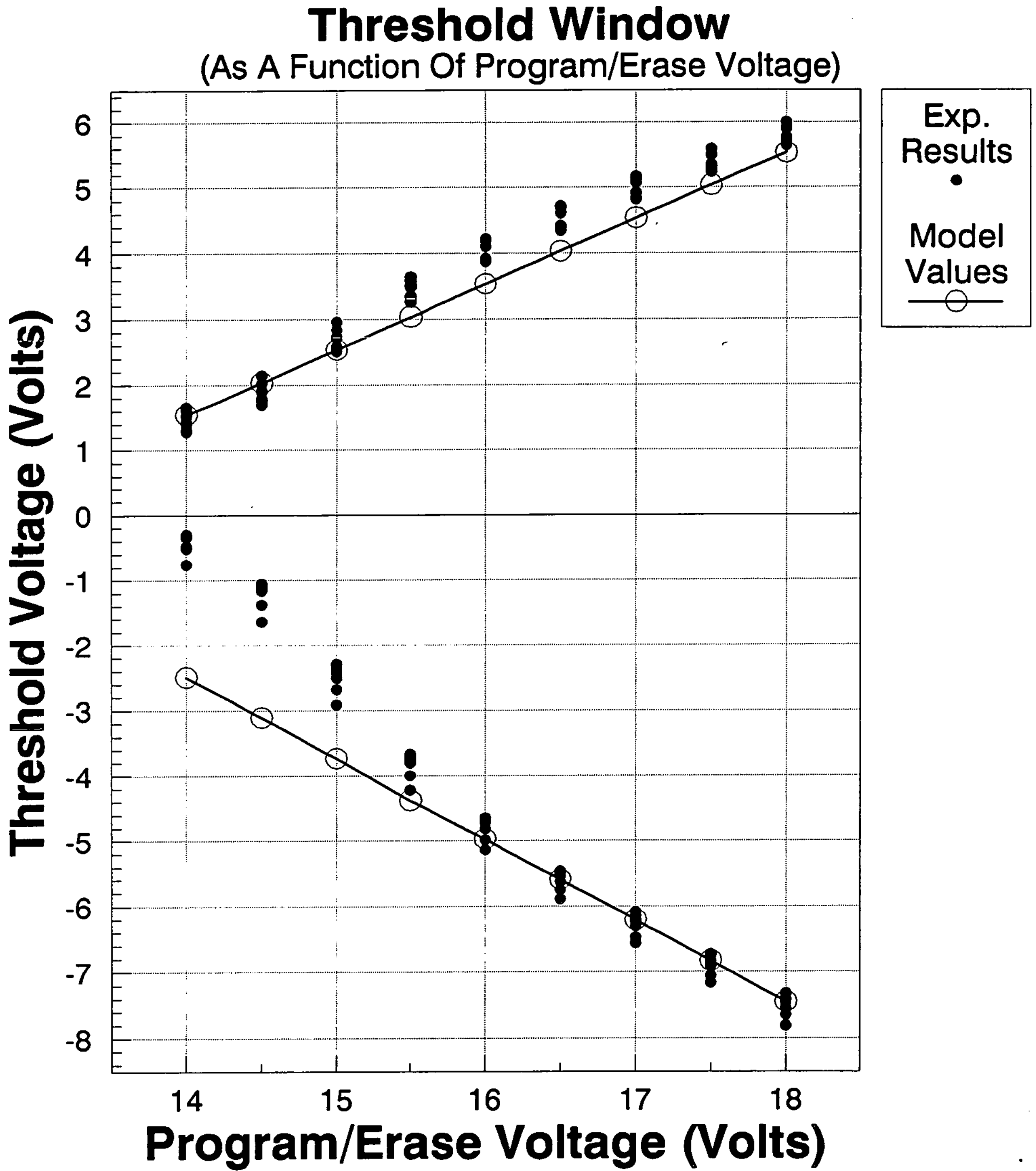

Figure 3–13: FETMOS Program/Erase Threshold Window.

For the erase operation all experimental and modelled data matched closely. This would indicate that model parameters are voltage independent, at least for the erase case. The program operation also fitted model predictions well, between 18V and 15.5V. However, for programming voltages below 15.5V, experimental threshold voltages fall more steeply than modelled ones. It seems that a new phenomena, not included in the model, now becomes visible. This effect has also been noted in the FLOTOX device [6], and has been attributed to deep depletion under the gate and tunnel oxide. However, it is considered to lie sufficiently far outside the normal operating regime of the FETMOS, to be neglected from this model.

3.6 Conclusion

An analytic model has been developed for the operation of the FETMOS device. Basic parameters such as effective width and parasitic resistance are used to describe the FETMOS, and currents are modelled with the Fowler-Nordheim tunnelling equation. The model is verified against experimental data, and is found to be in good agreement. This may now be used to investigate the internal working of the FETMOS, and a methodology developed to calculate FETMOS reliability. These avenues are dealt with in chapter 4.

Bibliography

[1] C.Hu. IC reliability simulation. *IEEE Journal Of Solid State Circuits*, 27(3):241–246, March 1992.

[2] H.Iizuka, F.Masuoka, T.Sato, and M.Ishikawa. Electrically alterable avalanche injection type MOS read-only memory with stacked gate structure. *IEEE Transactions On Electron Devices*, 23(4):379–387, 1976.

[3] S.T.Wang. Charge retension of floating gate transistors under applied bias conditions. *IEEE Transactions On Electron Devices*, 27(1):297–299, January 1980.

[4] R.D.Jolly, H.R.Grinolds, and R.Groth. A model for conduction in floating gate EEPROMs. *IEEE Transactions On Electron Devices*, 31(6):767–772, 1984.

[5] A.Bhattacharyya. Modelling of write/erase and charge retension characteristics of floating gate EEPROM devices. *Solid State Electronics*, 27(10):899–906, 1984.

[6] A.Kolodny, S.T.K.Nieh, B.Eitan, and J.Shappir. Analysis and modelling of floating-gate EEPROM cells. *IEEE Transactions On Electron Devices*, 33(6):835–844, 1986.

[7] P.I.Suciu, B.P.Cox, D.D.Rinerson, and S.F.Cagnina. Cell model for EEPROM floating gate memories. In *IEEE IEDM*, pages 737–740, 1982.

[8] J.Lee and V.K.Dham. Design considerations for scaling EEPROM FLOTOX cell. In *IEEE IEDM*, pages 589–592, 1983.

[9] K.A.Stroud. *Further Engineering Mathematics*, chapter 6. MacMillan, 1986.

[10] *HCMOS Single-Chip Microcomputer*, chapter 1. Motorola Inc, 1985.

[11] S.M.Sze, editor. *VLSI Technology*, chapter 3. McGraw-Hill International Editions, 1988.

[12] K.Y.Chang, S.Cheng, and K-M.Chang. An advanced high voltage CMOS process for custom logic circuits with embedded EEPROM. In *CICC*, 1988.

[13] *E-Probe 2000 User's Reference*, 1987.

[14] *4145B Semicinductor Parameter Analyser Operation And Service Manual*, 1986.

[15] D.Gorham, J.Wood, and D.Butts. *Field Effect Devices And VlSI*, chapter 2 Field-Effect Transistors. The Open University Press, 1985.

[16] B.G.Streetman. *Solid State Electronic Devices*, chapter 8. Field-Effect Transistors. Prentice/Hall International Editions, 1980.

[17] A.J.Walton and A.Gribben. A review of parametric testing. In *SEMICON Birmingham*, pages 39–63, 1987.

[18] D.Gorham, J.Wood, and D.Butts. *Field Effect Devices And VlSI*, chapter 4 The Operation Of The MOSFET. The Open University Press, 1985.

[19] C.Kuo, Y.R.Yeargain, and W.J.Downey. An 80ns 32K EEPROM using the FETMOS cell. *IEEE J.Solid State Circuits*, (5):821–827, October 1982.

[20] R.Bez, D.Cantarelli, and P.Cappelletti. Experimental transient analysis of the tunnel current in EEPROMs.. *IEEE Transactions On Electron Devices*, 37(4):1081–1086, 1990.

Chapter 4

Analysis Using the FETMOS Model

4.1 Transient Analysis of the FETMOS Device

Inclusion of a time constant τ, allows accurate modelling of transient response, for the FETMOS device. This helps to increase the accuracy of subsequent reliability analysis, as will be seen.

4.1.1 Threshold Window as a Function of Time

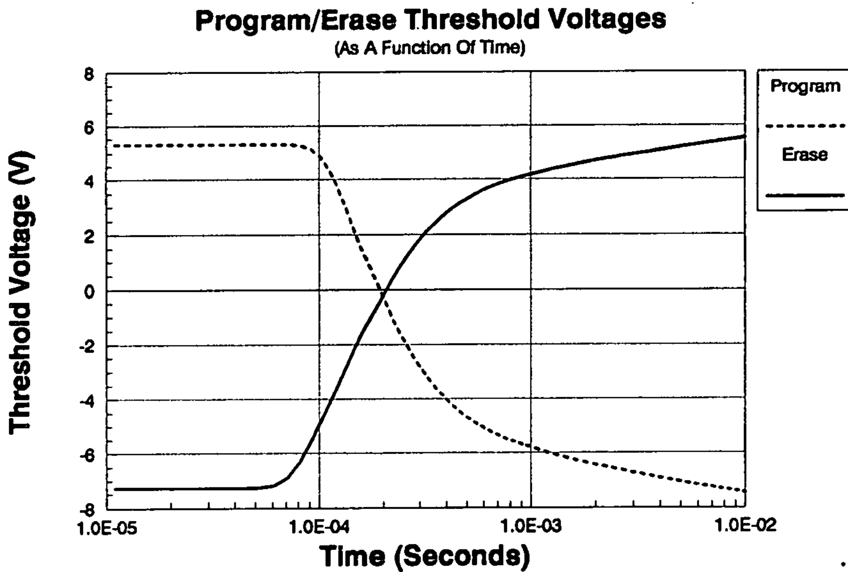

Figure 4–1: Program and Erase Threshold Voltages.

The variation in the program and erase threshold voltages are given as a function of time, in figure 4–1. These curves have the same form as those derived

experimentally [1]. Before programming, a device will be in the erased state, with a threshold voltage of 5.536V. Conversely, before erase a device will have a threshold voltage of -7.448V. For the first 50μs there is no change in the state of the FETMOS, since capacitances are charging up. Once the RC time constant of 100μs has been reached, the rate of change of threshold voltage arrives at its maximum. At 1ms the threshold voltages move more slowly to their final values, and at 10ms the program/erase operations are stopped.

4.1.2 Electric Fields as a Function of Time

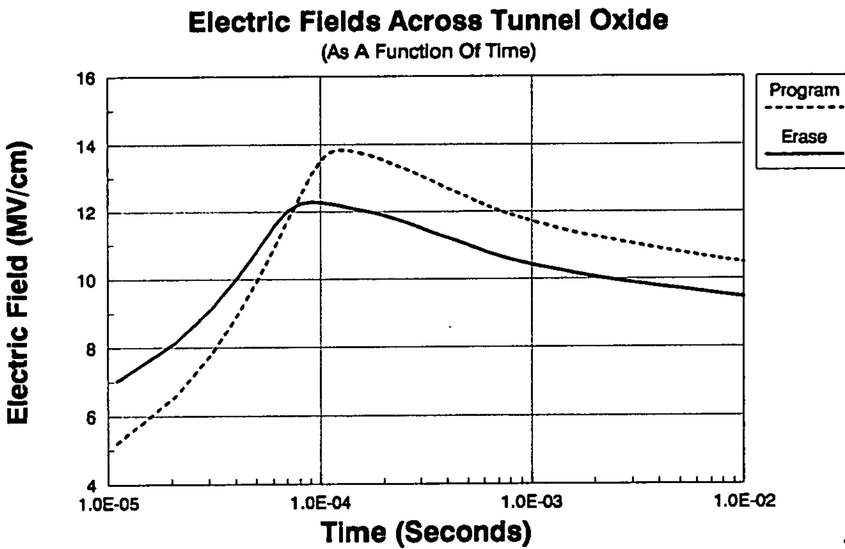

Figure 4–2: Electric Fields Across the Tunnel Oxide, During Program and Erase. Both Shown Positive For Comparison Sake .

Electric fields across the tunnel oxide are illustrated in figure 4–2, as a function of time. At the beginning of each operation, the initial electric fields are given by equations 3.30 and 3.31 . The initial program field is given by:

$$E_i = -\left(\frac{C_{fg}}{X_o C_t}\right)(V_{to} - V_{ti})$$

$$E_i = -\left(\frac{41.3fF}{108\mathring{A}\ 54.6fF}\right)(0.5 - 5.536) = 3.5MVcm^{-1}$$

The initial erase field is given by:

$$E_i = \left(\frac{C_{fg}}{X_o C_t}\right)(V_{to} - V_{ti})$$

$$E_i = \left(\frac{41.3fF}{108\AA\ 54.6fF}\right)(0.5 - 7.448) = -4.87MVcm^{-1}$$

In each case the field rises to a peak value at around $100\mu s$, then falls away. Since the program coupling ratio is larger than the erase coupling ratio, a larger voltage is coupled across the oxide during programming. Hence, the programming field reaches the highest peak value. Any oxide rupture which occurs, would be expected to coincide with the peak field. Thus, the program operation appears to be most prone to such failure. We have a peak field during programming of $13.87MVcm^{-1}$, and a peak field during erasing of $12.30MVcm^{-1}$.

4.1.3 Current Densities as a Function of Time

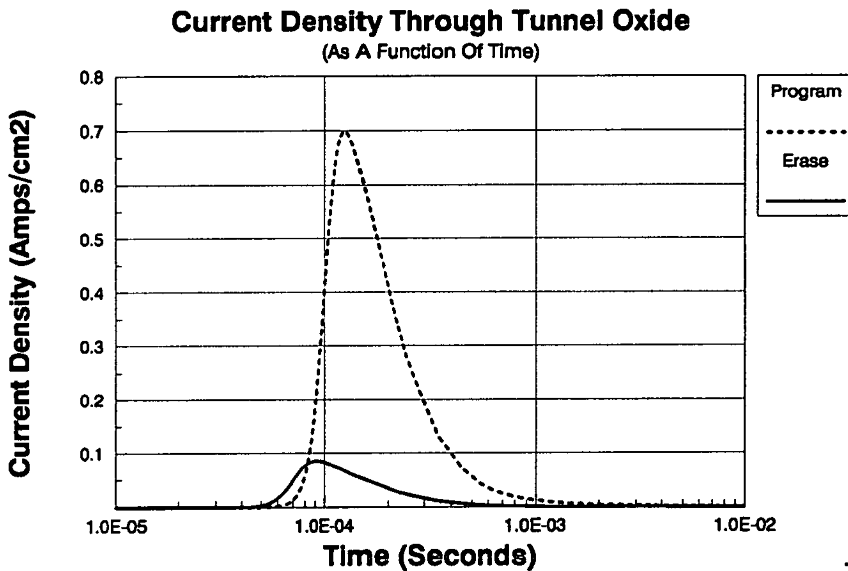

Figure 4–3: Current Densities in the Tunnel Oxide, During Programming And Erasing. Both Shown Positive for Comparison Sake.

Current densities in the tunnel oxide are illustrated in figure 4–3, as a function of time. These curves have the same form as tunnel currents observed experimentally [1], which helps to confirm the model's validity. Here again, the program current density is highest, since the program operation produces the highest electric fields. However, the tunnelling area is smallest for the program operation. Thus, the final charge packet on the floating gate, is similar in both the program

and erase case. The charge packet may be calculated using equation (4.1) [2]:

$$Q_i = C_{fg}(V_{to} - V_{tp}) \tag{4.1}$$

For the program case:

$$Q_i = 43.1fF(0.5 - -7.448) = 3.43 \times 10^{-13}C = 2.13 \times 10^6 \; holes$$

For the erase case:

$$Q_i = 43.1fF(0.5 - 5.536) = -2.17 \times 10^{-13}C = 1.35 \times 10^6 \; electrons$$

The total charge to pass through the tunnel oxide during programming, is the sum of negative charge, which must be removed, plus the positive charge which must be added. This is given by equation 4.2

$$3.43 \times 10^{-13} + 2.17 \times 10^{-13} = 5.6 \times 10^{-13}C \tag{4.2}$$

This figure is the same for erase, although for erase charge tunnels over a wider area.

4.1.4 Charge Densities as a Function of Time

By integrating the current density as a function of time, the charge density to pass through the oxide can be calculated. This is illustrated in figure 4–4. Integration of the current density over a complete program or erase operation gives the net charge density to pass through the oxide. For a program operation net charge density, $Q_{dp} = 1.441 \times 10^{-4}Ccm^{-2}$; and for an erase operation net charge density, $Q_{de} = 0.1522 \times 10^{-4}Ccm^{-2}$. These figures are used in assessing EEPROM reliability.

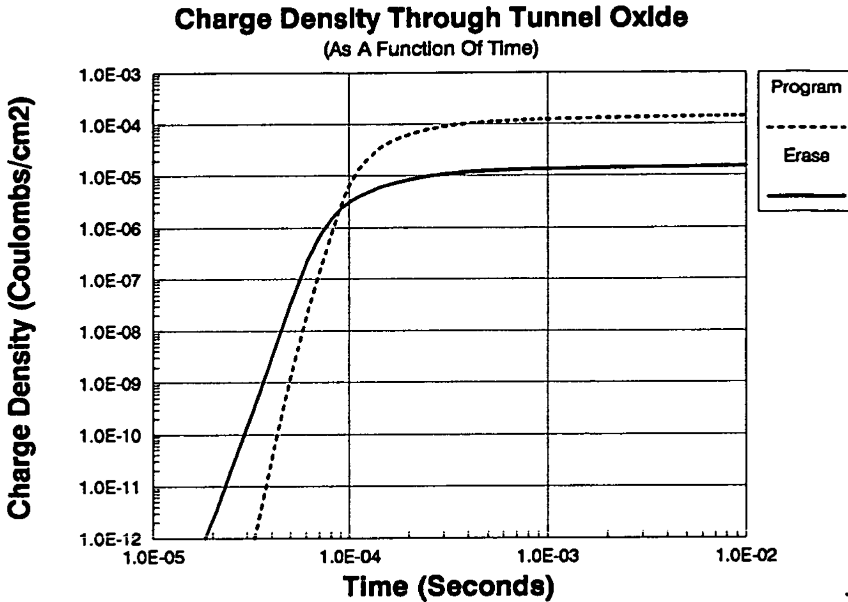

Figure 4–4: Charge Density to Pass Through the Tunnel Oxide, During Programming and Erasing. Both Shown Positive for Comparison Sake.

4.2 A Methodology for Modelling EEPROM Endurance

Three phenomena determine EEPROM program/erase endurance [3] [4]:

1. Trap Up, causing threshold window closure.

2. Time Dependent Dielectric Breakdown (TDDB) of the tunnel oxide.

3. Time Zero Dielectric Breakdown (TZDB) of the tunnel oxide.

A percentage of the electrons passing through the tunnel oxide become trapped there [5]. These reduce the electric fields at the injecting interfaces, so reducing the tunnel currents and associated threshold shifts. Eventually the threshold window closes to such an extent that the state of the EEPROM may no longer be read. This is the trap up failure mode [4]. Trap up represents less of a problem for *erase* endurance than for *program* endurance, since electron injection is distributed

uniformly over the whole gate area during erase [6]. Hence, a method will be presented for investigating the programming endurance, although this method is equally applicable to the erase case.

Assuming EEPROM failure is due to trap up during programming, the endurance may be described by equation 4.3:

$$Q_f = Q_p \times N_{cycles} \tag{4.3}$$

Where:

- N_{cycles} = The number of program/erase cycles an EEPROM can withstand, before the programmed threshold window closes. This defines the program endurance of the device, and is typically $\sim 10^5$ cycles [7].

- Q_p = The charge which passes though the tunnel oxide, during each program operation. In equation 4.3, it is assumed that Q_p is constant throughout the lifetime of the EEPROM. However, Q_p will depend upon the trapping in the oxide. During the first 10 cycles of EEPROM operation positive charge trapping will cause Q_p to increase [7]. Thereafter electron trapping will dominate, and Q_p will slowly reduce. An *average* value of $5.6 \times 10^{-13}C$ was calculated in section 4.3.1, which would be true of a midlife value.

- $Q_f = 5.6 \times 10^{-13}C \ \times \ 10^5 = 5.6 \times 10^{-8}C$
  This is the total charge to pass through the tunnel oxide, during all program operations, before failure. This depends upon oxide integrity, and a typical value has been given.

Let:

$$Q_{dp} = \frac{Q_p}{P_a} \tag{4.4}$$

$$Q_{df} = \frac{Q_f}{P_a} \tag{4.5}$$

Then, equation 4.3 can be re-written as:

$$Q_{df} \times P_a = Q_{dp} \times P_a \times N_{cycles} \tag{4.6}$$

Area cancels out, to give:

$$Q_{dp} = \frac{Q_{df}}{N_{cycles}} \tag{4.7}$$

Where:

- P_a = Tunnelling area during programming.
- Q_{dp} = The net charge density which flows during *one* program operation. (This is the charge fluence, for *one* program operation [8].)
- Q_{df} = The net charge density which flows after $N_{cycle} \sim 10^5$ program operations.

Thus, it has been shown that Q_{dp} is inversely proportional to the endurance, N_{cycles}. For instance, if Q_{dp} is halved then the endurance is doubled, since N_{cycles} must be doubled to maintain the equality of 4.7. It may be said that endurance variations are the *reciprocal* of Q_{dp} variations. Percentage variations in Q_{dp} will be calculated using equation 4.8:

$$R_{qdp} = \left(\frac{Q_{dp}}{Q_{dp\ standard}}\right) \times 100 \tag{4.8}$$

Where:

- $Q_{dp} = Q_{dp}$ for an EEPROM in which a parameter, such as gate/drain overlap, is varied.
- $Q_{dp\ standard} = Q_{dp}$ for an EEPROM with standard parameter values.
- R_{qdp} = Relative Q_{dp}.

The mechanisms which lead to time dependent dielectric breakdown are related to those which give trap up [3]. Thus, the relative susceptibility of a device to time dependent breakdown can also be described by variations in Q_{dp}. On the other hand, Time Zero Dielectric Breakdown (TZDB) is caused by electric field stress. Susceptibility to this is determined by the oxide dielectric strength, and the peak program or erase field. The percentage change in peak field, caused by allowing

parameters to vary, may also calculated. However, the growth of high integrity oxides, and production prescreening, ensure that TZDB is insignificant compared to trap up.

Finally, it should be appreciated that charge trapping is a complex phenomena. In chapter 2 the impact ionisation model was discussed [3]. However, there are conflicting views [9] [8], and the mechanisms involved are not properly understood. It seems likely that after a program or erase operation, the trapped charge relaxes. Thus, some trapped positive and negative charge is lost, or detrapped [10]. Higher current density (or higher oxide field) is expected to cause more impact ionisation inside the oxide, and increase the hole trapping rate [8].

Were the above phenomena *understood*, one could derive a function for charge trapping. This could be included in an EEPROM model, and absolute EEPROM endurance could be calculated. For instance, the threshold window after 10^4 program/erase cycles could be found. Even then, oxide reliability is acutely sensitive to processing conditions, such as preoxidation clean [11] [12] and purity of gas supplies. Temperature also is known to accelerate degradation phenomena [13]. EEPROM operating temperature depends upon the frequency of the logic circuitry, in which it is embedded. These parameters would need to be accounted for in such a model.

Given the wide disagreement concerning oxide degradation mechanisms, one should question the wisdom of including them in an EEPROM model. One should also question the merit, since the absolute endurance is already *known*, from experimental measurements [7]. What is of interest is the variation in endurance, caused by change in any parameter. In essence this is a sensitivity analysis, and such a methodology has been proposed in this thesis. The author believes this to be the best *engineering* solution to the problem.

4.3 Analysis of the Threshold Window and Endurance

4.3.1 Effect of Varying Floating Gate/Drain Overlap

A number of interesting effects are seen when parameters are allowed to vary. Figure 4–5 illustrates the variation in program threshold voltage, as a function of floating gate/drain overlap. Varying overlap has two opposing effects:

1. By reducing the area available for tunnelling, smaller overlaps give less charge flow, which reduces the threshold voltage.

2. By increasing the program coupling ratio, smaller overlaps increase the tunnelling field, which will increase the charge flow and threshold voltage.

The two effects counteract one another, the first one dominates below $0.31\mu m$ overlap, and the second one dominates above $0.31\mu m$ overlap. This said, the threshold voltage remains relatively constant above $0.31\mu m$. Note, that varying overlap has *no* effect on the erase operation.

Figure 4–6 gives the variation in net charge density for the program operation [1]. It is seen that increasing floating gate/drain overlap improves reliability substantially, ie by $\simeq 30\%$ for a $0.1\mu m$ increase in overlap. The peak field and threshold voltage are also included in the figure, both of which remain constant as overlap is increased. Hence, the threshold window is *unaffected* by increasing overlap. (The erase operation is immune to overlap variations). Therefore, any increase in overlap promises to pay substantial dividends, in terms of reliability improvement.

It is proposed that overlap could be increased by increasing the tilt angle of the drain implantation. No increase would then be needed to the thermal budget

---

[1] A decrease in net charge density is equivalent, to an increase in endurance.

of the process. This is significant, since the trend in VLSI processes is towards low thermal budgets. Overlap could also be increased to by an increasing the drain doping density, or the use of a more diffusive dopant. Table 4–1 summarises the effects of a variation in floating gate/drain overlap.

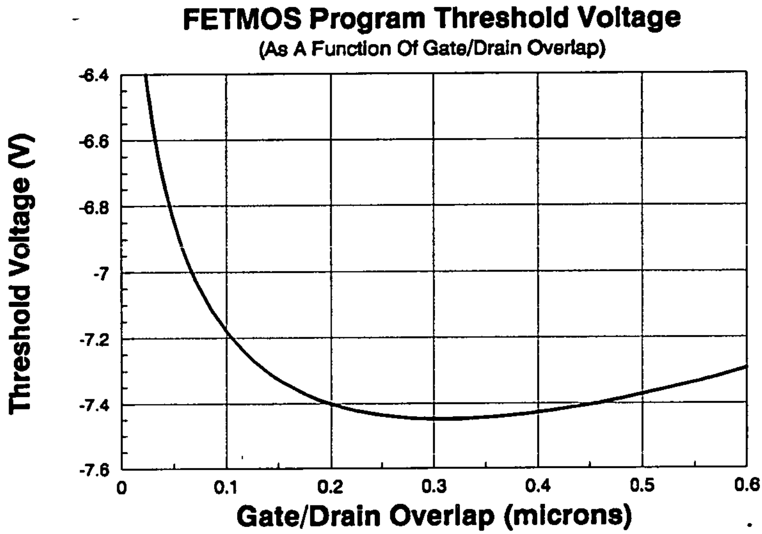

Figure 4–5: FETMOS Program Threshold Voltage as a Function of GD_{over}.

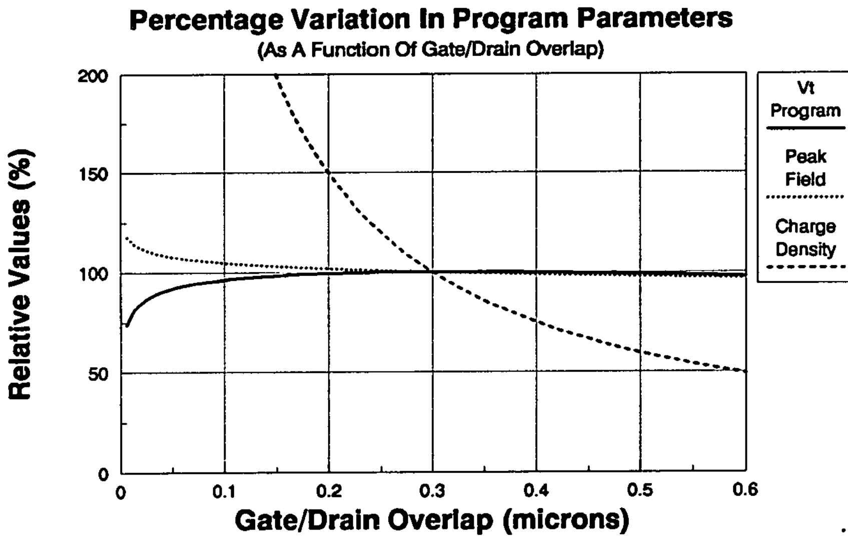

Figure 4–6: Percentage Variation in Program Parameters as a Function of GD_{over}. Note, a Decrease in Charge Density is Equivalent to an Increase in Endurance.

Parameter	Reduce Floating Gate/Drain Overlap	Increase Floating Gate/Drain Overlap
C_{fg}	$\otimes$	$\otimes$
C_{fd}	$\Downarrow$	$\Uparrow$
C_{fs}	$\Downarrow$	$\Uparrow$
C_{fc}	$\Uparrow$	$\Downarrow$
Program Coupling Ratio	$\Uparrow$	$\Downarrow$
Program Tunnel Area	$\Downarrow$	$\Uparrow$
Program Endurance	$\Downarrow$	$\Uparrow$ (Steep Rise)
V_{tp}	$\Downarrow$	$\Downarrow$ (Shallow Fall)
Erase Coupling Ratio	$\otimes$	$\otimes$
Erase Tunnel Area	$\otimes$	$\otimes$
Erase Endurance	$\otimes$	$\otimes$
V_{te}	$\otimes$	$\otimes$

Table 4–1: Effect of Floating Gate/Drain Overlap Variation, About the Target Value of $0.3\mu m$. Symbols Represent:Increase $\Uparrow$, Decrease $\Downarrow$, No Change $\otimes$.

4.3.2 Effect of Varying Effective Width

Figure 4–7 illustrates the variation in threshold window, as a function of effective width, W_{eff}. A decrease in this reduces the tunnel areas and capacitances C_{fd}, C_{fs} and C_{fc}, but has no effect on C_{fg}. A decrease in W_{eff} has two opposing effects:

1. By reducing the area available for tunnelling, a smaller W_{eff} gives less charge flow, which reduces program and erase threshold voltages.

2. By increasing the coupling ratios, a smaller W_{eff} increases charge flow, which increases program and erase threshold voltages.

The two effects counteract one another, this situation is similar to a variation in floating gate/drain overlap. In the erase case, the increase in coupling ratio dominates, and the threshold voltage rises as W_{eff} is reduced. In the program case, for which the coupling ratio is much larger, the reduction in tunnelling area dominates. Thus program threshold voltage falls, as W_{eff} is reduced. Since both coupling ratios increase as W_{eff} is reduced, the program and erase endurance both decrease, as illustrated in figures 4–8 and 4–9.

W_{eff} may be varied either in the reticle design, or by growing a thicker field oxide, so increasing the degree of birds beaking. Table 4–2 summarises the effect of variations in W_{eff}.

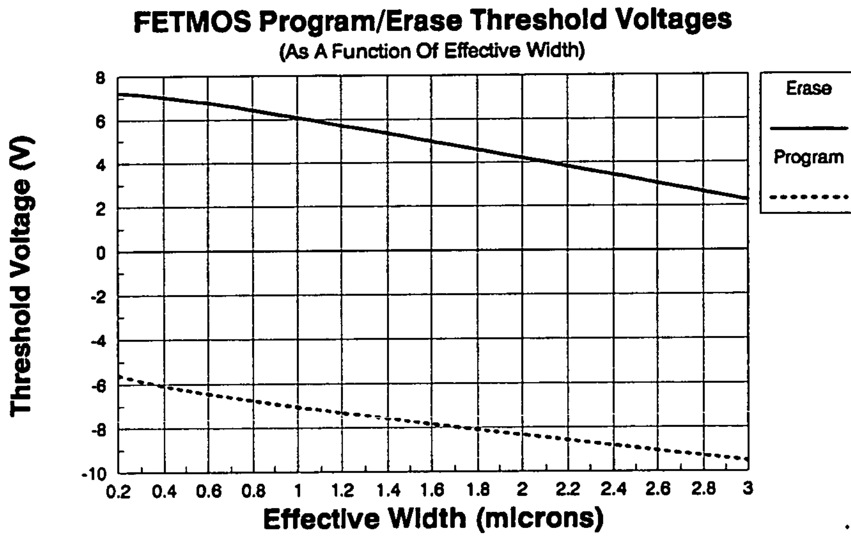

Figure 4–7: FETMOS Threshold Window as a Function of W_{eff}.

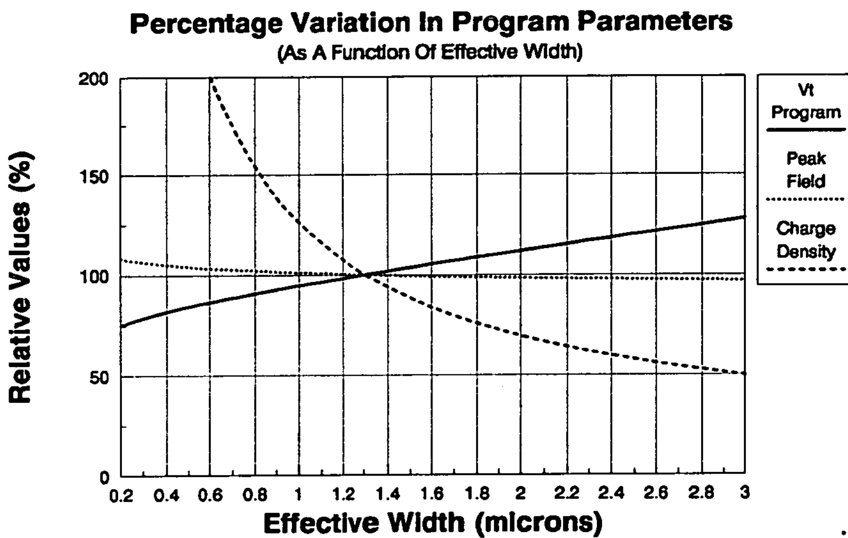

Figure 4–8: Percentage Variation in Program Parameters as a Function of W_{eff}. Note, a Decrease in Charge Density is Equivalent to an Increase in Endurance.

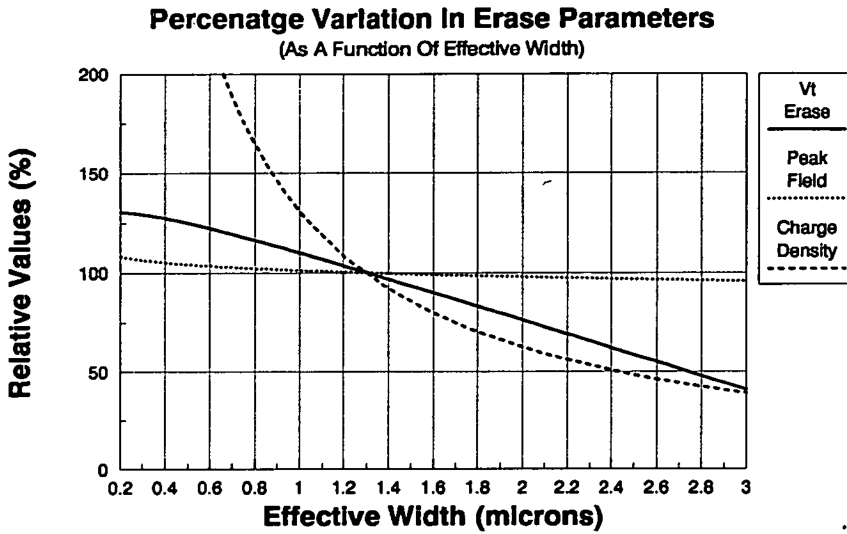

Figure 4–9: Percentage Variation in Erase Parameters as a Function of W_{eff}. Note, a Decrease in Charge Density is Equivalent to an Increase in Endurance.

Parameter	Reduce Effective Width	Increase Effective Width
C_{fg}	⊗	⊗
C_{fd}	⇓	⇑
C_{fs}	⇓	⇑
C_{fc}	⇓	⇑
Program Coupling Ratio	⇑	⇓
Program Tunnel Area	⇓	⇑
Program Endurance	⇓	⇑
V_{tp}	⇓	⇑
Erase Coupling Ratio	⇑	⇓
Erase Tunnel Area	⇓	⇑
Erase Endurance	⇓	⇑
V_{te}	⇑	⇓

Table 4–2: Effect of Effective Width Variation, About the Target Value of 1.3μ. Symbols Represent:Increase ⇑, Decrease ⇓, No Change ⊗.

4.3.3 Effect of Varying Floating Gate Area

Figure 4–10 illustrates the variation in program and erase threshold voltages, as a function of floating gate area. An increase in floating gate area will increase C_{fg}, which has two opposing effects:

1. As C_{fg} increases so the erase and program coupling ratios are increased, therefore the charge flow and threshold voltages increase.

2. Increasing C_{fg} causes a redistribution of injected charge on the floating gate, and the threshold voltage will reduce [14], according to the equation:

$$V_t = V_{to} - \frac{Q_i}{C_{fg}}$$

These two effects work in opposition to one another. In the erase case, the increase in coupling ratio dominates, and threshold voltage rises as floating gate area is increased. In the program case, for which the coupling ratio is much larger, the redistribution of charge dominates, hence the threshold voltage falls as floating gate area is increased.

Figure 4–11 illustrates program endurance and program threshold voltage, and suggests that both can be enhanced by reducing A_{fg}. Figure 4–12 gives erase endurance and erase threshold voltage. Although decreasing A_{fg} improves the erase endurance, it will also reduce the erase threshold voltage. Thus there is a trade off when reducing A_{fg}, between:

1. Enhanced endurance and program threshold voltage.

2. Reduced erase threshold voltage.

It is important therefore, to consider endurance and threshold window together, when assessing the effect of any parameters. A_{fg} may be varied either during the reticle design, or by varying exposure time during photo-lithography. Table 4–3 summarises the effects of a variation in A_{fg}.

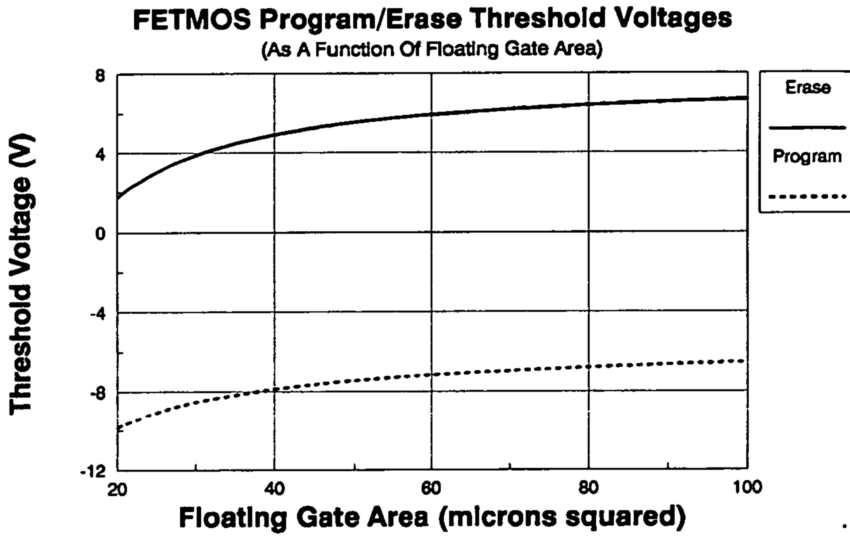

Figure 4–10: FETMOS Threshold Window as a Function of A_{fg}.

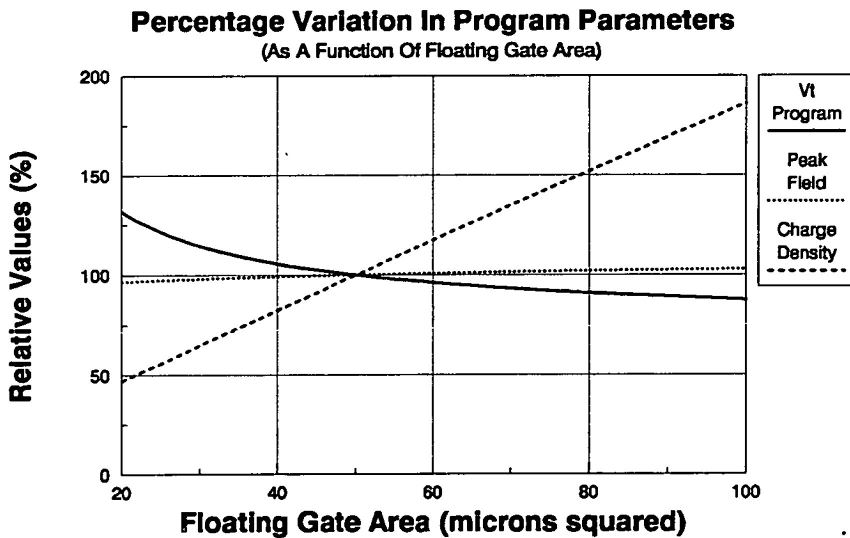

Figure 4–11: Percentage Variation in Program Parameters as a Function of A_{fg}. Note, a Decrease in Charge Density is Equivalent to an Increase in Endurance.

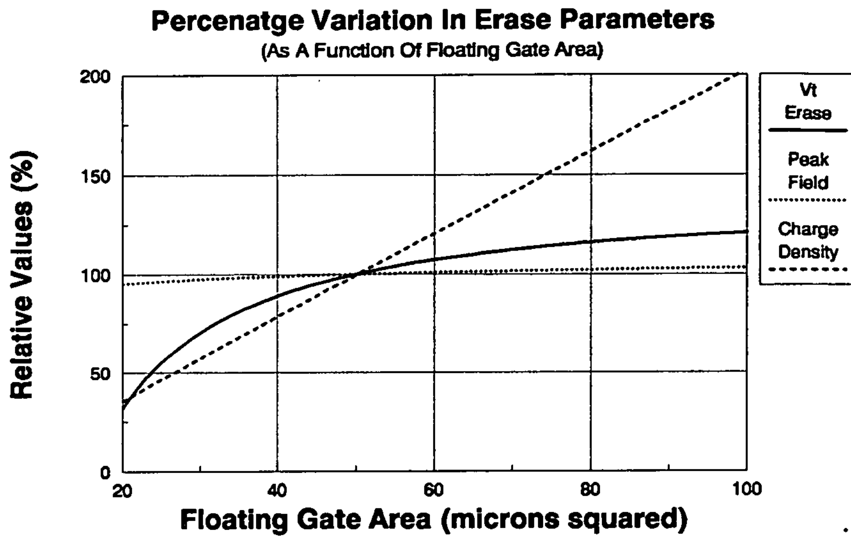

Figure 4–12: Percentage Variation in Erase Parameters as a Function of A_{fg}. Note, a Decrease in Charge Density is Equivalent to an Increase in Endurance.

Parameter	Reduce Floating Gate Area	Increase Floating Gate Area
C_{fg}	⇓	⇑
C_{fd}	⊗	⊗
C_{fs}	⊗	⊗
C_{fc}	⊗	⊗
Program Coupling Ratio	⇓	⇑
Program Tunnel Area	⊗	⊗
Program Endurance	⇑	⇓
V_{tp}	⇑	⇓
Erase Coupling Ratio	⇓	⇑
Erase Tunnel Area	⊗	⊗
Erase Endurance	⇑	⇓
V_{te}	⇓	⇑

Table 4–3: Effect of Floating Gate Area Variation, About the Target Value of $50\mu m^2$. Symbols Represent:Increase ⇑, Decrease ⇓, No Change ⊗.

4.3.4 Effect of Varying Interlevel Oxide Thickness

A reduction in interlevel oxide thickness, X_{int}, increases C_{fg}. Thus, the effect of reducing X_{int} is directly equivalent to increasing the floating gate area. Figure 4–13 illustrates the threshold window as a function of X_{int}, while figures 4–14 and 4–15 illustrate the endurance. It should be remembered, that thin interlevel oxides can give rise to long term retention problems. This is due to the onset of tunnelling at asperities. Table 4–4 summarises the effects of a variation in interlevel oxide thickness.

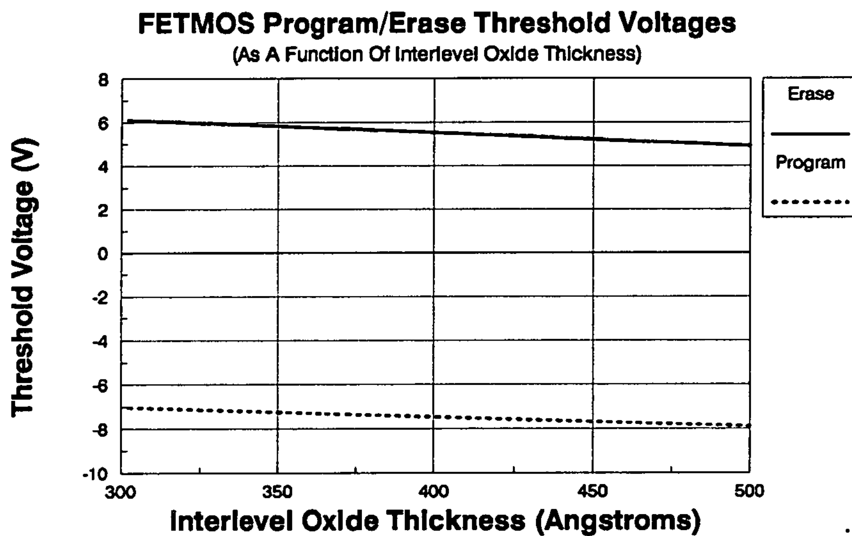

Figure 4–13: FETMOS Threshold Window as a Function of X_{int}.

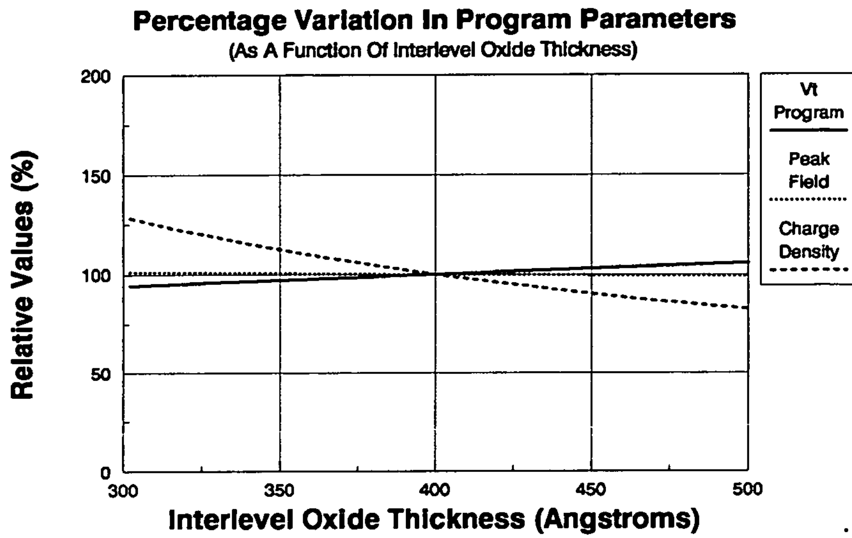

Figure 4–14: Percentage Variation in Program Parameters as a Function of X_{int}. Note, a Decrease in Charge Density is Equivalent to an Increase in Endurance.

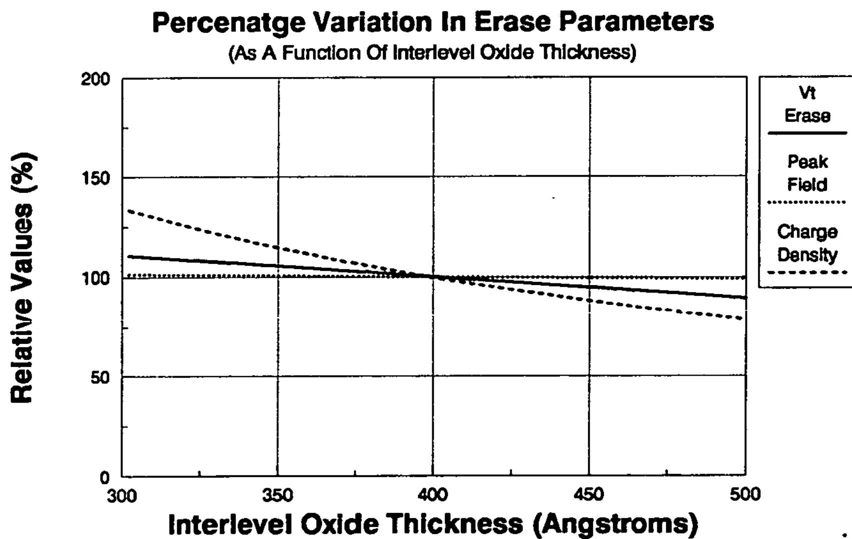

Figure 4–15: Percentage Variation in Erase Parameters as a Function of X_{int}. Note, a Decrease in Charge Density is Equivalent to an Increase in Endurance.

Parameter	Reduce Interlevel Oxide Thickness	Increase Interlevel Oxide Thickness
C_{fg}	⇑	⇓
C_{fd}	⊗	⊗
C_{fs}	⊗	⊗
C_{fc}	⊗	⊗
Program Coupling Ratio	⇑	⇓
Program Tunnel Area	⊗	⊗
Program Endurance	⇓	⇑
V_{tp}	⇓	⇑
Erase Coupling Ratio	⇑	⇓
Erase Tunnel Area	⊗	⊗
Erase Endurance	⇓	⇑
V_{te}	⇑	⇓

Table 4–4: Effect Of Interlevel Oxide Thickness Variation, About The Target Value Of 400Å. Symbols Represent:Increase ⇓, Decrease ⇑, No Change ⊗.

4.3.5 Effect of Varying Floating Gate Length

Figure 4–16 illustrates the variation in program and erase threshold voltages, as a function of floating gate length, L_g. An increase in L_g increases C_{fc} and the channel length. This has opposite effects on the program and erase operations:

1. Increasing L_g increases the program coupling ratio, and leaves the tunnel area unchanged. Therefore, the charge flow and program threshold voltage increase. Program coupling ratio is given by equation 4.9 [2]:

$$\frac{C_t - C_{fd}}{C_t} \tag{4.9}$$

2. Increasing L_g increases the erase tunnel area but reduces the erase coupling ratio. Now, the reduction in coupling ratio dominates. Thus, charge flow and erase threshold voltage both decrease. Erase coupling ratio is given by equation 4.10 [2]:

$$\frac{C_{fg}}{C_t} \tag{4.10}$$

In figure 4–17 it is seen that as L_g increases, so program threshold voltages increases, but endurance falls. In figure 4–18 it is seen that as L_g increases, erase threshold voltages falls, but endurance increases.

Gate length may be varied either in the reticle design, or during lithography. Over-exposure or under-exposure during photolithography, will modulate the amount of photoresist developed. This varies the amount of polysilicon removed by subsequent etching. Table 4–5 summarises the effects of varying L_g.

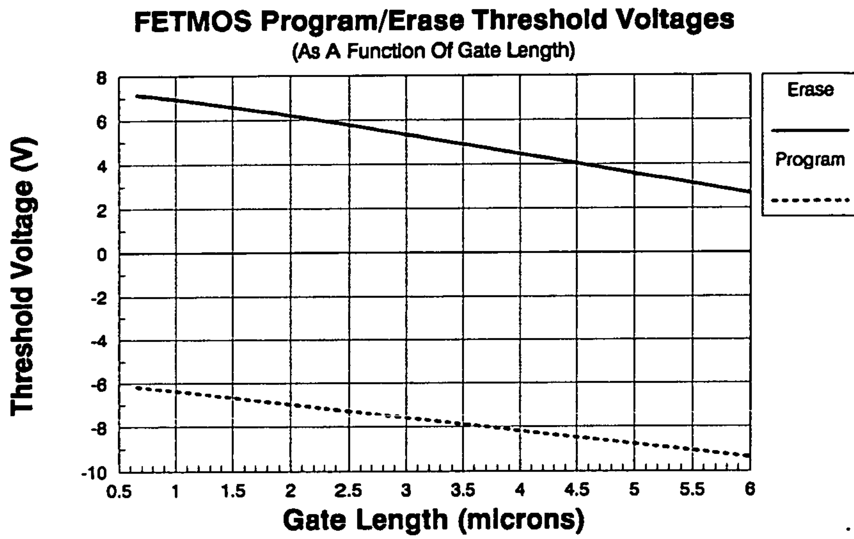

Figure 4–16: FETMOS Threshold Voltage as a Function of L_g.

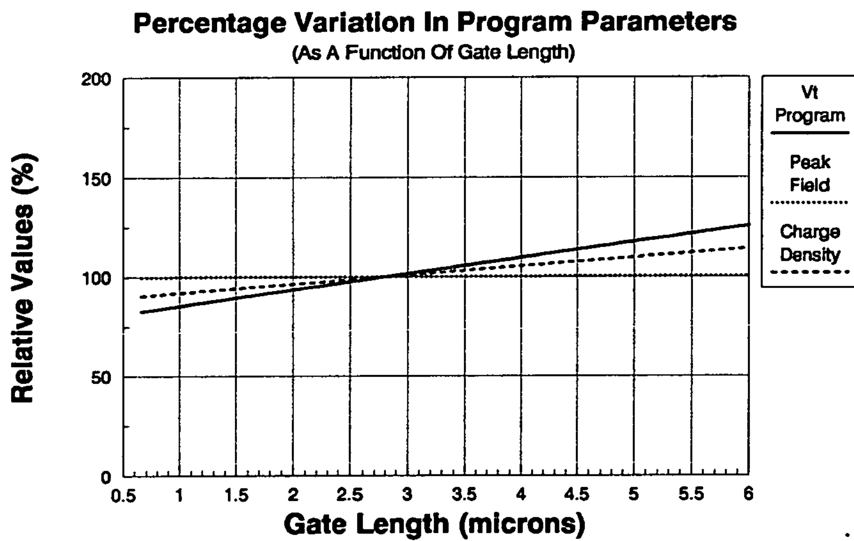

Figure 4–17: Percentage Variation in Program Parameters as a Function of L_g. Note, a Decrease in Charge Density is Equivalent to an Increase in Endurance.

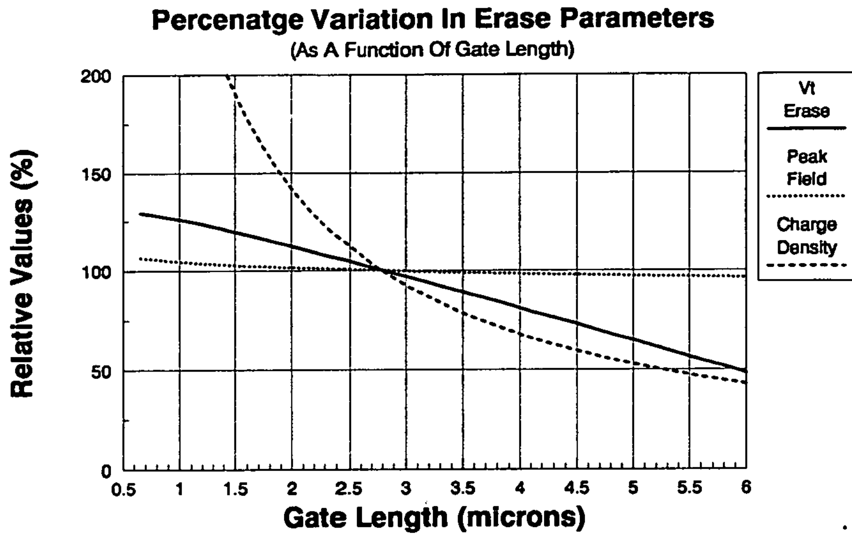

Figure 4–18: Percentage Variation in Erase Parameters as a Function of L_g. Note, a Decrease in Charge Density is Equivalent to an Increase in Endurance.

Parameter	Reduce Gate Length	Increase Gate Length
C_{fg}	⊗	⊗
C_{fd}	⊗	⊗
C_{fs}	⊗	⊗
C_{fc}	⇓	⇑
Program Coupling Ratio	⇓	⇑
Program Tunnel Area	⊗	⊗
Program Endurance	⇑	⇓
V_{tp}	⇓	⇑
Erase Coupling Ratio	⇑	⇓
Erase Tunnel Area	⇓	⇑
Erase Endurance	⇓	⇑
V_{te}	⇑	⇓

Table 4–5: Effect of Gate Length Variation, About The Target Value of $2.8\mu m$. Symbols Represent:Increase ⇑, Decrease ⇓, No Change ⊗.

4.3.6 Effect of Varying Gate Oxide Thickness

Although the optimum thickness for good gate oxide integrity is 110Å [15], there will always be some variation during fabrication. Figure 4–19 illustrates the program and erase threshold voltages as a function of oxide thickness, X_o. Increasing X_o has two opposing effects:

1. An increase in X_o reduces capacitances C_{fd}, C_{fc} and C_{fs}, but leaves C_{fg} unchanged. This increases coupling ratios, and hence charge flow and threshold voltages.

2. An increase in X_o reduces the electric field across the tunnel oxide, since:

$$E = \frac{V}{X_o}$$

   This reduces charge flow and threshold voltages.

The second effect dominates, hence program and erase voltages are both reduced by thicker oxides. There is a corresponding increase in endurance, as illustrated in figures 4–20 and 4–21. In table 4–6 the effects of gate oxide variation are summarised. Note, a change in oxide thickness could also produce a change in the Fowler-Nordheim coefficients [16].

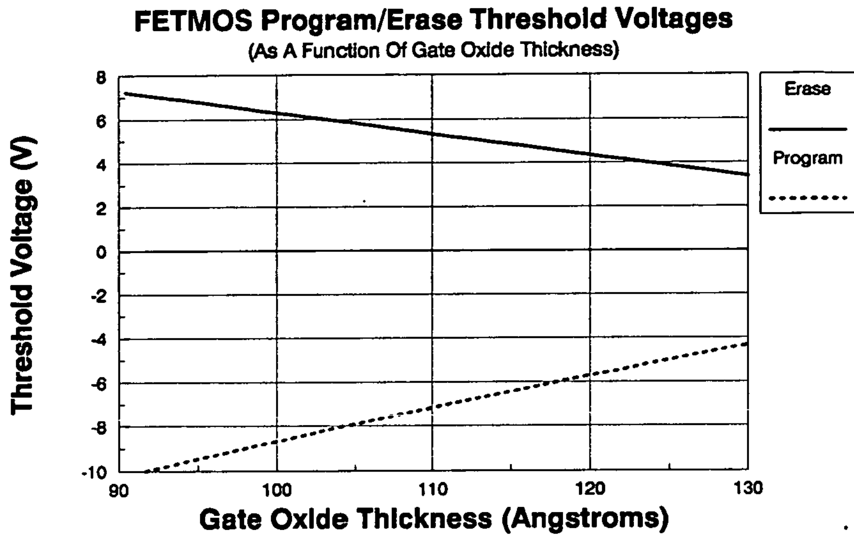

Figure 4–19: FETMOS Threshold Window as a Function of X_o.

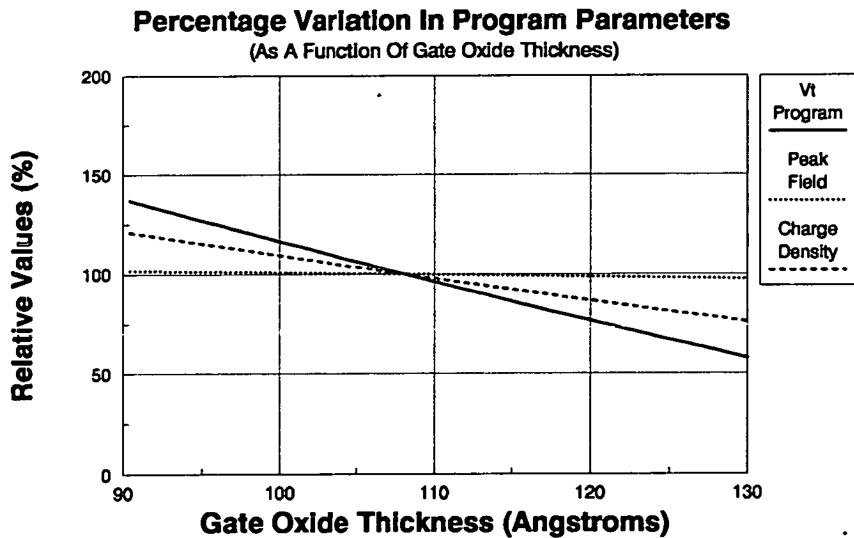

Figure 4–20: Percentage Variation in Erase Parameters as a Function of X_o. Note, a Decrease in Charge Density is Equivalent to an Increase in Endurance.

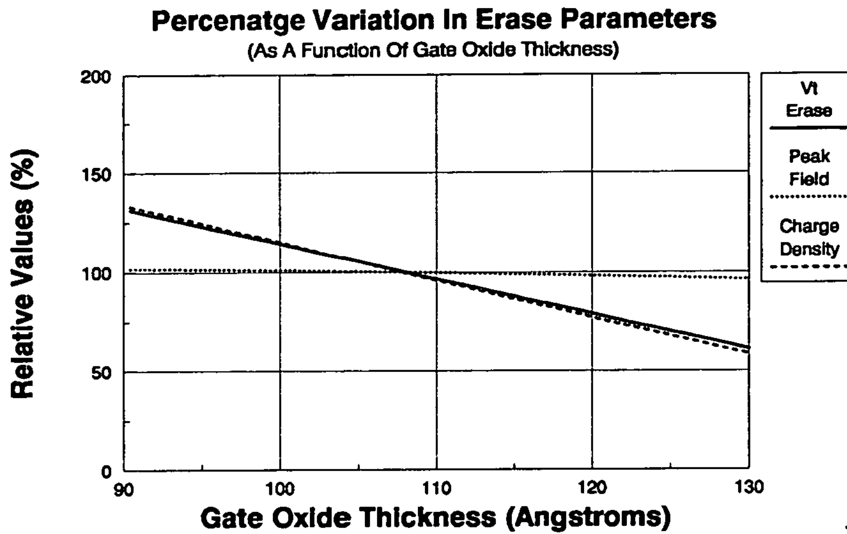

Figure 4–21: Percentage Variation in Program Parameters as a Function of X_o. Note, a Decrease in Charge Density is Equivalent to an Increase in Endurance.

Parameter	Thinner Gate Oxide	Thicker Gate Oxide
C_{fg}	⊗	⊗
C_{fd}	⇑	⇓
C_{fs}	⇑	⇓
C_{fc}	⇑	⇓
Program Coupling Ratio	⇓	⇑
Program Area	⊗	⊗
Program Endurance	⇓	⇑
V_{tp}	⇑	⇓
Erase Coupling Ratio	⇓	⇑
Erase Area	⊗	⊗
Erase Endurance	⇓	⇑
V_{te}	⇑	⇓

Table 4–6: Effect of Gate Oxide Variation, About the Target Value of 110Å. Symbols Represent:Increase ⇑, Decrease ⇓, No Change ⊗.

4.3.7 Effect of Varying Fowler-Nordheim Coefficients, A and B

To appreciate the sensitivity of FETMOS operation to these coefficients, a study has been made, although the value of these coefficients is not usually a consideration in the design of an EEPROM process. In figure 4–22 program and erase threshold voltages are seen to fall dramatically with increasing B. However, figure 4–23 shows threshold voltages to rise with increasing A.

In figures 4–24, 4–25, 4–26 and 4–27, it is seen that a reduction in current density is allied to an marked increase in the peak electric field. Both these effects are associated with an increased opposition to tunnelling through the oxide. Since the peak electric field also varies as A and B vary, reliability is not simply a function of charge density. Fowler-Nordheim coefficients will be effected by asperities at the injecting interface, doping concentration, oxide thickness and the degree of charge trapping. Table 4–7 summarises the effect of varying A and B.

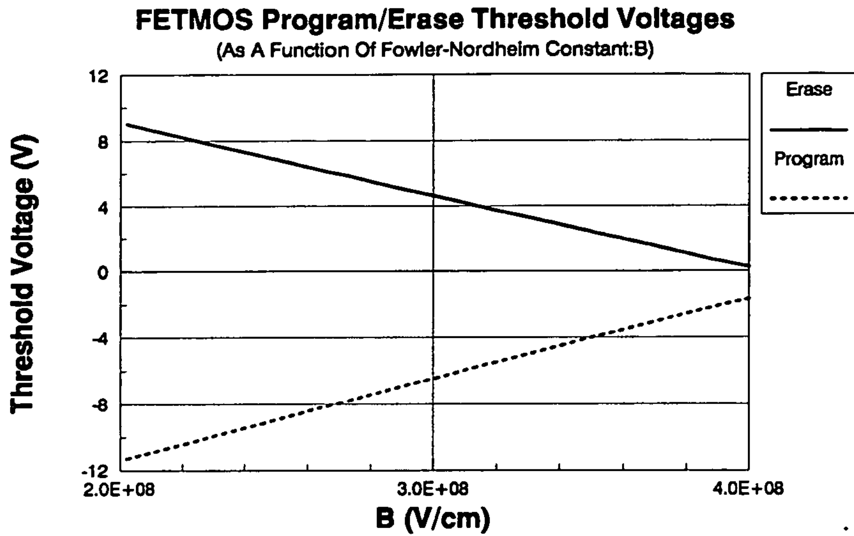

Figure 4–22: FETMOS Threshold Window as a Function of Fowler-Nordheim Coefficient B.

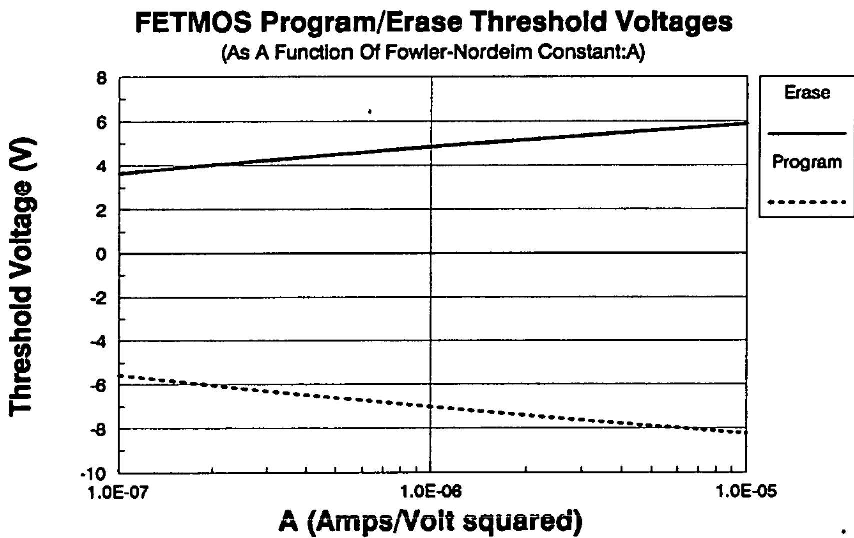

Figure 4–23: FETMOS Threshold Window as a Function of Fowler-Nordheim Coefficient A.

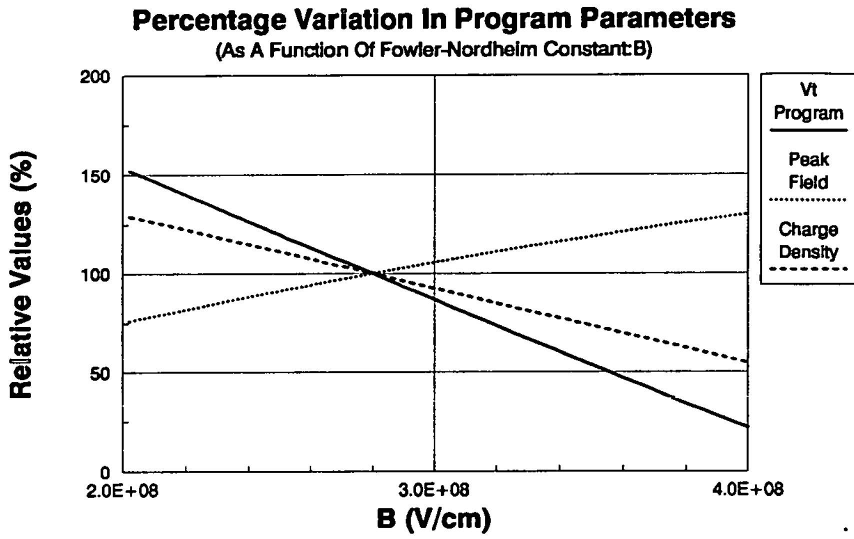

Figure 4–24: Percentage Variation in Program Parameters as a Function of B.

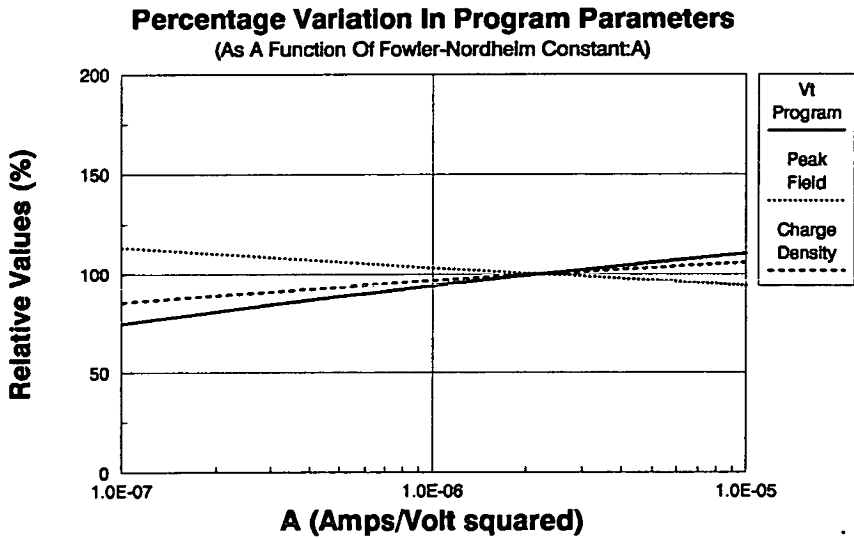

Figure 4–25: Percentage Variation in Program Parameters as a Function of A.

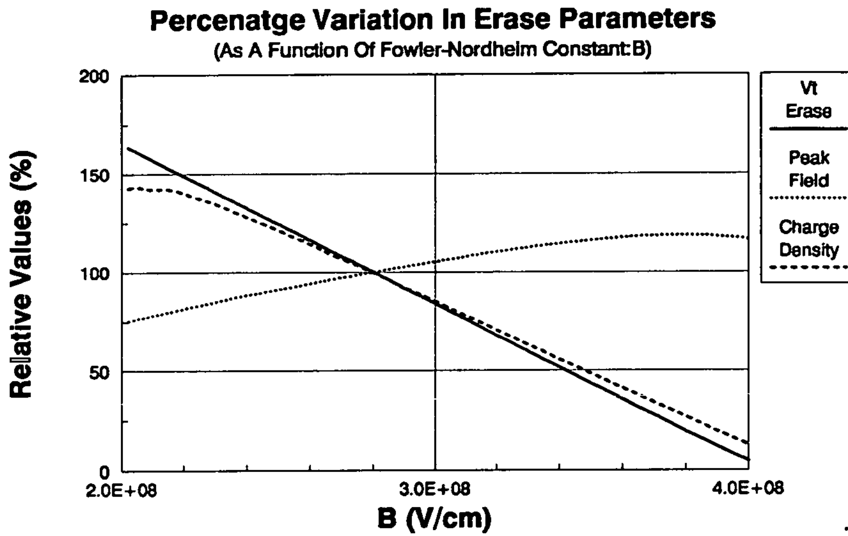

Figure 4–26: Percentage Variation in Erase Parameters as a Function of B.

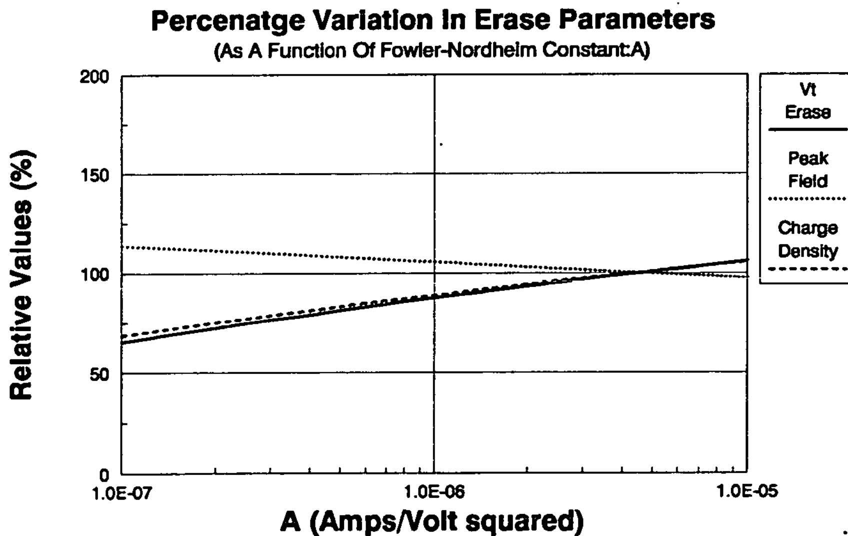

Figure 4–27: Percentage Variation in Erase Parameters as a Function of A.

Parameter	Reduce A	Increase A	Reduce B	Increase B
C_{fg}	$\otimes$	$\otimes$	$\otimes$	$\otimes$
C_{fd}	$\otimes$	$\otimes$	$\otimes$	$\otimes$
C_{fs}	$\otimes$	$\otimes$	$\otimes$	$\otimes$
C_{fc}	$\otimes$	$\otimes$	$\otimes$	$\otimes$
Program Coupling Ratio	$\otimes$	$\otimes$	$\otimes$	$\otimes$
Program Tunnel Area	$\otimes$	$\otimes$	$\otimes$	$\otimes$
V_{tp}	$\Downarrow$	$\Uparrow$	$\Uparrow$	$\Downarrow$
Erase Coupling Ratio	$\otimes$	$\otimes$	$\otimes$	$\otimes$
Erase Tunnel Area	$\otimes$	$\otimes$	$\otimes$	$\otimes$
V_{te}	$\Downarrow$	$\Uparrow$	$\Uparrow$	$\Downarrow$

Table 4–7: Effect of Fowler-Nordheim Coefficient Variation, About Their Measured Values. Symbols Represent:Increase $\Downarrow$, Decrease $\Uparrow$, No Change $\otimes$.

4.4 Conclusion

A model has been developed for the FETMOS cell, which encompasses transient response, threshold window and reliability. A good correlation has been shown between modelled data and experimental results, testifying to the model's accuracy. The effect of basic design parameters upon threshold window has been characterised, thus indicating how processing variations may be used to tailor the threshold window. Equally, the model can be used to predict the effect of sizing down a circuit. In general, any change in coupling ratio has a significant effect on the erase threshold voltage, whereas the program threshold voltage is more susceptible to changes in tunnelling area. It has been seen that the floating gate/drain overlap effects both the program tunnel area and program coupling ratio. However, the two effects act in opposition, and the threshold window is stable as overlap is increased.

Reliability has been modelled in terms of peak electric field and endurance. Parameter variations were seen to have little effect on the peak field. In contrast, endurance had a strong dependence on parameter variations. Program endurance is of particular concern [6], and large improvements can be made in this by increasing the floating gate/drain overlap (with little effect on threshold window). The overlap is therefore is a promising avenue for improvement of FETMOS endurance. Modern VLSI processes require low thermal budgets. Thus, it is proposed that overlap could be increased by increasing the tilt angle of the drain implantation, increasing the drain doping density, or the use of a more diffusive dopant.

Bibliography

[1] R.Bez, D.Cantarelli, and P.Cappelletti. Experimental transient analysis of the tunnel current in EEPROMs. *IEEE Transactions On Electron Devices*, 37(4):1081–1086, 1990.

[2] A.Kolodny, S.T.K.Nieh, B.Eitan, and J.Shappir. Analysis and modelling of floating-gate EEPROM cells. *IEEE Transactions On Electron Devices*, 33(6):835–844, 1986.

[3] Ih-C.Chen, S.E.Holland, and C.Hu. Electrical breakdown in thin gate and tunneling oxides. *IEEE Transactions On Electron Devices*, 32(2):413–422, February 1985.

[4] H.E.Meas, J.Witters, and G.Groeseneken. Trends in non-volatile memory devices and technologies. In *ESSDERC Bologna*, pages 743–754, 1987.

[5] Y.Hokari, T.Baba, and N.Kawamura. Reliability of 6-10nm thermal SiO_2 films showing intrinsic dielectric integrity. *IEEE Transactions On Electron Devices*, 32(11):2485–2491, November 1985.

[6] J.S.Witters, G.Groesenken, and H.E.Maes. Degradation phenomena of tunnel oxide floating gate EEPROM devices. In *IMEC*, pages 167–170, 1987.

[7] C.Kuo, Y.R.Yeargain, and W.J.Downey. An 80ns 32K EEPROM using the FETMOS cell. *IEEE J.Solid State Circuits*, (5):821–827, October 1982.

[8] S.Haddad and M-S.Liang. The nature of charge trapping responsible for thin-oxide breakdown under dynamic field stress. *IEEE Electron Device Letters*, 8(11):524–527, 1987.

[9] D.R.Wolters, A.T.A.Zegers-van, and Duynhoven. On the mechanism of intrinsic breakdown in thin dielectrics. In *Proceedings Of The First International Symposium On Ultra Large Scale Integration Science And Technology*, pages 101–118, May 1987.

[10] M-S.Liang, S.Haddad, W.Cox, and S.Cagnina. Degradation of very thin gate oxide MOS devices under dynamic high field/current stress. In *IEDM*, pages 394–398, 1986.

[11] J.L.Prom, J.Castagne, G.Sarrabayrouse, and A.Munoz-Yague. Influence of preoxidation cleaning on the electrical properties of thin SiO_2 layers. *IEE Proceedings.*

[12] A.Hariri. Evaluate wafer cleaning effectiveness. *Semiconductor International*, pages 74–78, 1989.

[13] S.M.Sze, editor. *VLSI Technology*, chapter 14. McGraw-Hill International Editions, 1988.

[14] R.E.Shiner, N.R.Mielke, and R.Haq. Characterisation and screening of SiO_2 defects in EEPROM structures. In *IEEE/IRPS*, pages 248–256, 1983.

[15] K.Yamabe, K.Taniguchi, and Y.Matsushita. Thickness dependence of dielectric breakdown failure of thermal SiO_2 films. In *IEEE IRPS*, pages 184–190, 1983.

[16] J.Callder. Master's thesis, University Of Edinburgh, 1988.

Chapter 5

Fabrication of EEPROM Structures

Modelling of the EEPROM has shown that an increase in the floating gate/drain overlap, can be used to improve reliability. This result will now be investigated experimentally. A DC tunnel current between the floating gate and drain may be used to emulate programming [1]. Both the fabrication and testing can therefore be simplified, since a complete EEPROM is not required. Simple MOS transistors with a single gate may be used, providing they have a high integrity tunnel oxide. A batch of MOS transistors, with a range of gate/drain overlaps, have been fabricated using the Progressional OffseT (POT) technique [2].

It has been observed by Motorola, that phosphorus doped EEPROMs are more reliable than their arsenic counterparts. To examine this effect two sets of POT transistors have been designed, one in arsenic and one in phosphorus. These have the same doping profiles, eg. equal junction depthes, so that results may easily be compared. Computer simulation has been used to design these profiles.

5.1 The Progressional Offset Technique

The Progressional OffseT (POT) technique has been implemented in a number ways [3] [4] [2]. Briefly, a column of MOS transistors is fabricated, in which gate/drain overlap covers a range of values. The idea can be expressed most succinctly in terms of a diagram, see figure 5–1.

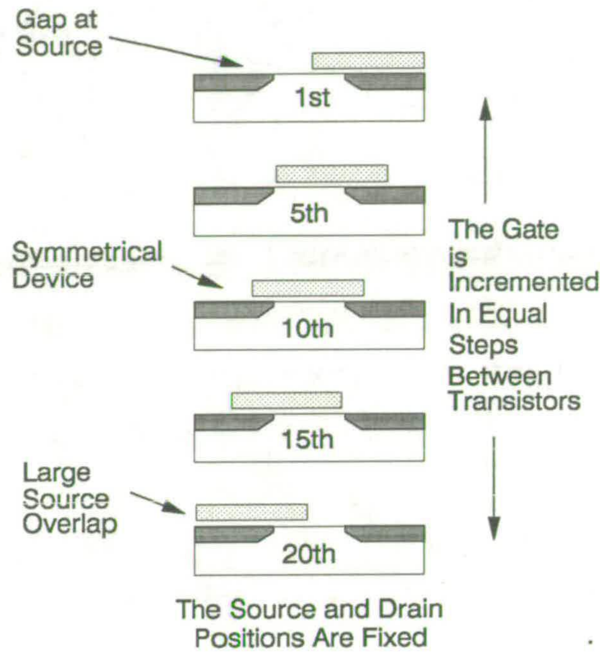

Figure 5–1: Schematic Diagram Illustrating a Column of Progressional Offset Transistors.

All transistors have the same dimensions, but the location of the gate is incremented by an equal step from one device to the next. Transistors at either end of the column have a source gap, or drain gap, whereas transistors in the centre are symmetrical. The gapped and symmetrical transistors may be distinguished, by comparing electrical characteristics, such as subthreshold or substrate current [3] [4]. Even a slight variation in gate/drain overlap can have a significant effect on electric fields within a device [5], so small step sizes are generally preferable.

5.2 Processing

5.2.1 Overview

A new process has been developed to *reliably* produce the POT transistors. This includes several novel features. Polysilicon gates are deposited in two stages [1], as in advanced BiCMOS processes [6], and for this reason it is named the Bi-Poly process. A new oxidation recipe has also been developed, which uses a long-time postanneal to ensure oxide integrity [7]. In addition, many steps are adapted from the Edinburgh Microfabrication Facility's 1.5μ and 6.0μ NMOS processes. Much of the processing, such as wet etching, plasma etching and furnace oxidation, was carried out by the author.

5.2.2 The Bi-Poly Process

This is a non-aligned process, which may be summarised in terms of the lithography stages required. In optical lithography light is projected onto a silicon wafer though a patterned mask. The mask pattern exposes photographically sensitive material, and this defines transistor features. The Bi-Poly process requires 5 photomasks, or reticles.

Photomask1

Active regions are defined: Illustrated in figure 5–2.

---

[1] 300$\mathring{A}$ are initially deposited, while the remaining 6000$\mathring{A}$ are added later.

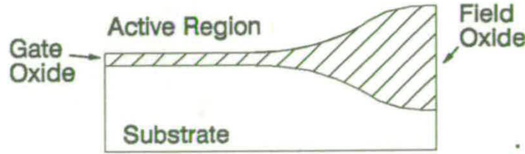

Figure 5–2: Definition of Active Regions.

Photomask2

Definition of *Phantom gates*: Transistor gates are defined in photoresist. These provide a mask during source-drain implantation, see figure 5–3. After implantation the phantom gates are removed.

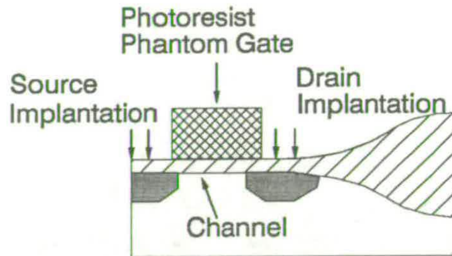

Figure 5–3: Phantom Gate Definition, for Implantation of Drain Regions.

Photomask3

Definition of polysilicon gates: The real polysilicon gates are defined, see figure 5–4.

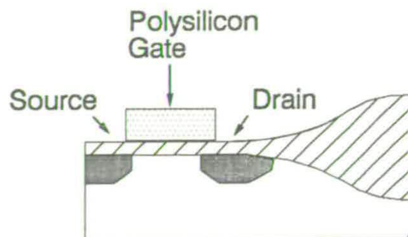

Figure 5–4: Definition of the Polysilicon Gate.

At this stage the gate/drain overlaps are defined - since the gate may be positioned anywhere. The registration accuracy between one mask layer and the next is $\sim 0.2\mu m$. Therefore, the POT column will be skewed to the left or right, as illustrated in figure 5–5.

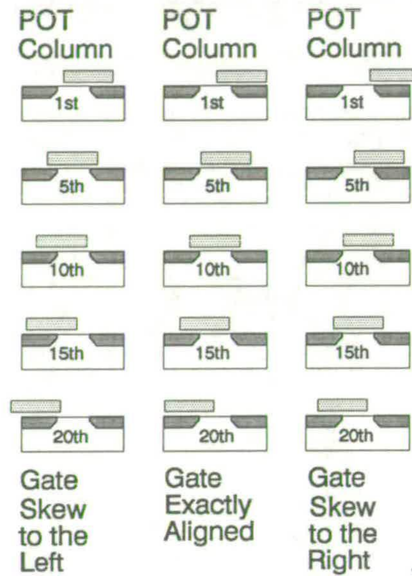

Figure 5–5: Schematic Diagram Illustrating Three POT Columns: Two Skewed and One Symmetrical.

Photomask4

Definition of contact holes: The contacts allow connection from the metal layer to the source,drain and gate. As illustrated in figure 5–6.

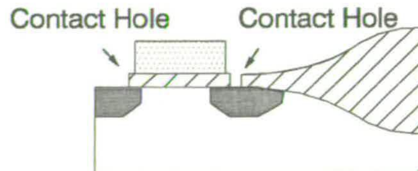

Figure 5–6: Definition of Contact Holes.

Photomask5

Definition of metallisation: The metal layer connects transistors to probe pads, see figure 5–7.

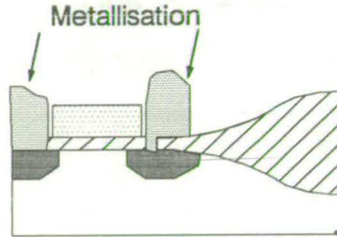

Figure 5–7: Metal Layer.

5.2.3 Growth of High Integrity Thin Oxide Films

The most challenging aspect of the process, is the growth of a high integrity thin oxide film. The oxide recipe developed for this purpose is illustrated in figure 5–8.

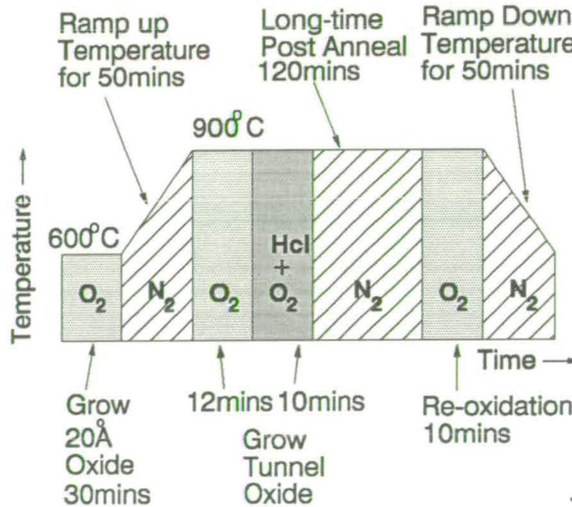

Figure 5–8: Thin Oxidation Process.

Chemical Pre-Clean

To remove contamination before oxidation, a chemical pre-clean is required. The EMF uses a four stage RCA clean:

- $5:1:1,\ H_2O : H_2O_2 : NH_4OH$
  Wafers are boiled at 80^oC in a solution of hydrogen peroxide and ammonia, which removes organic material.

- $4:1,\ HF : H_2O$
  Hydrofluoric acid removes nascent oxide.

- $5:1:1,\ H_2O : H_2O_2 : HCL$
  Wafers are boiled at 80^oC in a solution of hydrogen peroxide and hydrochloric acid, which to remove heavy metals and mobile contaminants (eg. Na^+).

- $4:1,\ HF : H_2O$
  Hydrofluoric acid removes nascent oxide.

Silicon oxidation in the thin regime is highly sensitive to treatments given to the silicon wafers before oxidation [8] [9]. For this reason it is important to use an identical clean for each batch of wafers.

Furnace Loading Conditions

Wafers are loaded at $600°C$ in oxygen, and remain there for 20 minutes to allow the growth of a thin oxide. They are then ramped to $900°C$ in N_2. During the ramp the oxide layer protects the silicon from attack by the N_2 [10].

Hole Trap Generation

Oxide breakdown under TDDB stress has been linked to hole trapping [1], greater reliability may therefore be obtained by limiting the number of hole traps generated during processing. It is believed that hole traps (formed during processing)

result from a deficiency of oxygen in the SiO_2 [11]. Trap generation occurs during high temperature steps, and is believed to be caused by the diffusion of Si from the substrate/gate oxide interface into the SiO_2 [12]. Temperature is the primary driving force for this diffusion, which argues for the use of low temperatures wherever possible. It should also be remembered, that high temperature processing steps subsequent to the gate oxidation will be equally effective in generating hole traps.

Stacking Faults

Oxidation induced stacking faults (OSF) may lead to poorer oxide quality, although the correlation between dielectric strength and OFS is not straightforward [13]. A stacking fault is a dislocation in the silicon crystal lattice, which may become decorated with metallic impurities and may also result in a thinning of the oxide in the immediate vicinity. The metallic impurities provide a path of least resistance to electrical field lines, resulting in a convergence of field lines in that region and a proportionally higher field. High temperatures ($\sim 1000°C$) propagate these stacking faults, and lowering the processing temperature will lessen the problem [14]. Careful wafer handling will also help to avoid surface damage which may "seed" stacking faults.

Thickness Uniformity

Uniformity and reproducibility of oxide thickness are important to the operation of circuits including EEPROMs. This is especially so during the Program/Erase operation, when the thickness determines the size of tunnelling field experienced by the oxide, and therefore the size of the tunnelling current. To simplify comparisons between wafers in the experiments, a uniform and reproducible oxide is also needed. A temperature in the order of $900°C$ will give a slow and *controlled* growth rate.

Mobile Ions

Mobile ion contamination (eg. Na^{+}) has a dual effect on device reliability. It can of itself lead to oxide failure, but will also to cause threshold voltages to vary as it moves within the gate oxide. HCl is the standard treatment, added during oxidation to getter mobile ions [15]. However, attention should be given to the reaction between HCl and bare silicon at $\sim 900°C$. This tends to pit the Si surface, and HCl should only be added following the growth of a thin protective layer of SiO_2

Long-Time Postanneal

Inhomogeneities in the oxide film thickness can lead to low field breakdown events $\leq 1MVcm^{-1}$. These inhomogeneities are most pronounced in the form of pin holes, as illustrated in figure 5–9.

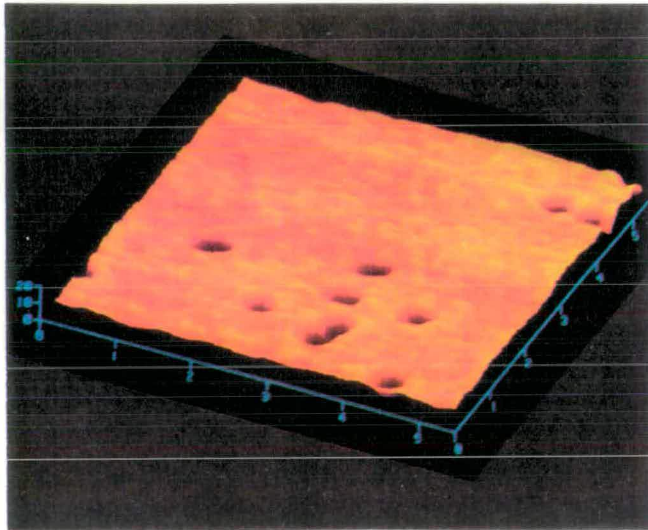

Figure 5–9: Quantum Tunnelling Micrograph of a Thin Oxide, Revealing Pin Holes.

It has been found that a long anneal given subsequent to oxidation significantly reduces the fraction of such events [7]. These long-time postanneals should be given at the oxidation temperature ($900°C$) for several hours. Thermal relaxation of the

oxide, due to viscous flow, has been proposed as the mechanism responsible for improved integrity. In the light of other work it seems reasonable to assume that plastic flow also plays a role in this relaxation [16]. Not only does relaxation give a more homogeneous oxide film, but it also reduces the thermal stress between the oxide and Si substrate [17] [18]. Such stress would otherwise degrade oxide integrity [19]. A long time is required for relaxation to take place, since the viscosity of the oxide at $900°C$ is relatively high.

Hole trap formation is the only detrimental effect of such a long anneal. These traps may be removed through the use of a second oxidation step, "re-oxidation", at $900°C$ for 10 minutes at the end of the anneal [7] [20] [12]. Re-oxidation provides O_2, which reacts with any SiO, which is thought to cause the hole traps.

Wafer Cooling

Withdrawal of the wafers from the furnace results in the formation of a stressed region between the Si substrate and the oxide film, due to the mismatch in the coefficients of their thermal expansivity. To mitigate this effect, wafers may be cooled to $600°C$ before unloading. Wafers withdrawn in such a manner have shown a lower infant mortality rate and a substantially higher dielectric strength [7]. The furnace then sits at 600°C.

Polysilicon Deposition

After oxidation, wafers should be immediately transferred to the polysilicon deposition furnace, where $\sim$ 300$\AA$ of polysilicon is deposited. This protects the oxide layer from contamination during further steps. It will also provide a path to ground for static charge which tends to build up during ion implantation. A thicker polysilicon layer could *not* be used at this stage, as implanted arsenic and phosphorus ions must be able to penetrate through it, to form the source and drain. This step is also used in advanced BiCMOS processes [6].

5.2.4 Complete Process Flow

- Starting Point.

  3 inch, P-type, Czochralski, silicon substrates were used, with a resistivity in the range 14.0 to 20.0 Ω cm and a ⟨100⟩ crystal orientation.

- STEP 1. Initial Clean.

  Wafers should be free from organic films, metals and particulates. New wafers, as used here, usually require no treatment.

- STEP 2. Pad Oxide.

  A 350 $\AA$ oxide is grown at 950^oC.

- STEP 3. Silicon Nitride Deposition.

  A 1000 $\AA$ layer of silicon nitride (Si_3N_4) is deposited at 800°C, by low pressure chemical vapour deposition (LPCVD).

- STEP 4. 1st Photomask.

  Active regions are defined, using optical lithography.

- STEP 5. Silicon Nitride RIE.

  Reactive ion etching (RIE) is used to remove Si_3N_4 from areas unprotected by photoresist.

- STEP 6. Silicon Nitride RIE.

  Repeat RIE with wafers face down to etch Si_3N_4 from wafer backs.

- STEP 7. Channel Stop Implant.

  To increase the threshold voltage of the field oxide regions, boron is implanted. This prevents the creation of a conduction path under the field oxide, between active regions.

  - Boron dose=1×10^{13}atoms cm^{-2}, energy=100 keV.

  Values for dose and energy are based on those used in the Edinburgh Microfabrication Facility's 6.0μm process. This step is illustrated in figure 5–10.

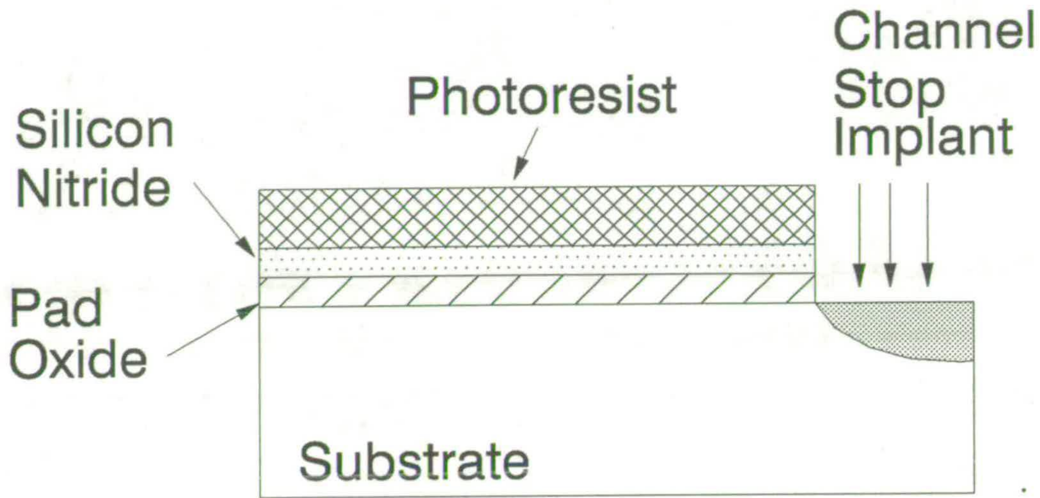

Figure 5–10: Step 7. Implant Boron to Create Channel Stop.

- STEP 8. Photoresist Strip in Oxygen Plasma.

  Exposure to an oxygen plasma removes photoresist.

- STEP 9. Photoresist Strip in Fuming Nitric Acid.

  Any remaining photoresist is removed in fuming nitric acid.

- STEP 10. Field Oxide Growth.

  15 hours 45 minutes at $950°C$ in steam grows a $13000Å$ field oxide. A thick oxide is desired to reduce the capacitance of field oxide regions. This step is illustrated in figure 5–11.

- STEP 11. Photoresist Coat.

  This protects field oxide on the front of the wafers.

- STEP 12. 4:1 HF Dip.

  A 4:1 mixture of ammonium fluoride (NH_3F) and hydrofluoric acid (HF) etches oxide from wafer backs.

- STEP 13. Photoresist Strip in Fuming Nitric Acid.

- STEP 14. 4:1 HF Dip.

  The oxide film on the Si_3N_4 surface is removed. The field oxide will be etched simultaneously, so a short etch time is used.

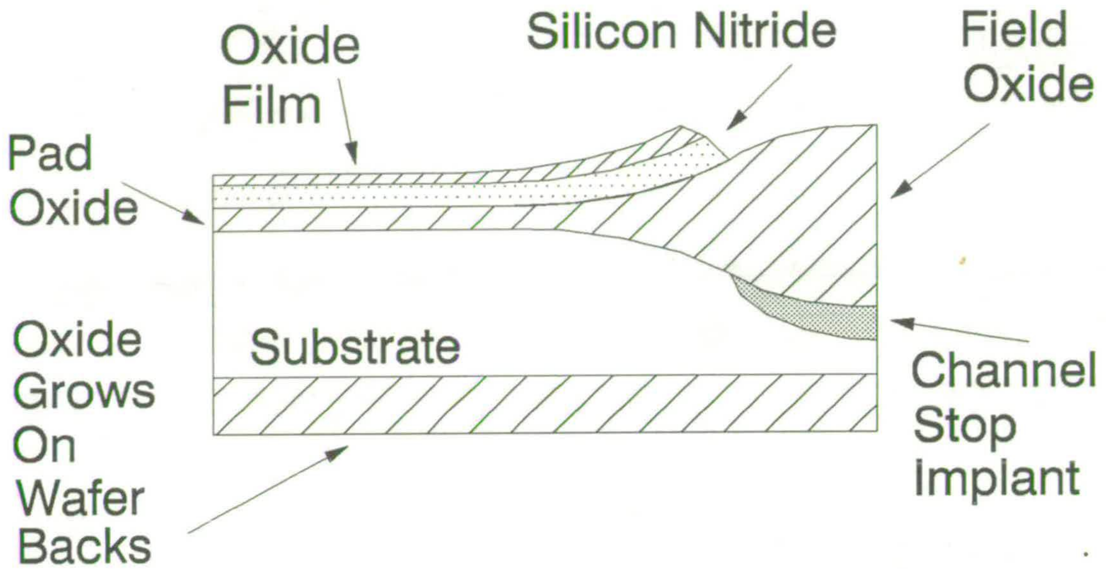

Figure 5–11: Step 10. Field Oxidation.

- STEP 15. Silicon Nitride Wet Etch.

  Ortho-phosphoric acid (H_3PO_4) [21] at 165°C, is used to remove the Si_3N_4.

- STEP 16. Threshold Adjust Implant.

  Boron is implanted through the pad oxide, to raise the P-type dopant concentration at the silicon surface. This gives the transistor channel regions a positive threshold voltage.

  – Boron dose=2×10^{12} atoms cm^{-2}, energy=50 keV.

  The required doses and energies are simulated in section 5.4. Note that these simulations cannot be performed in context, until the full process has been established.

- STEP 17. 4:1 HF Dip.

  The pad oxide is removed.

- STEP 18. Sacrificial Oxide Growth.

  A 300Å sacrificial oxide is grown at $950^{o}C$. This consumes the silicon surface, so lifting away contamination. In addition, it offers the surface protection

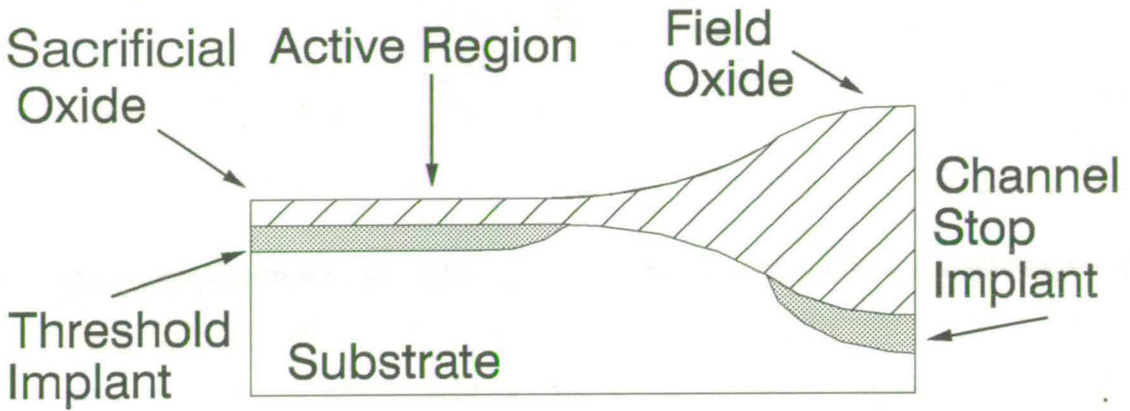

Figure 5–12: Step 18. Growth of Sacrificial Oxide.

from contamination prior to gate oxidation. This step is illustrated in figure 5–12.

- STEP 19. 4:1 HF Dip.

  Remove sacrificial oxide.

- STEP 20. RCA Clean.

  Any remaining contamination is removed from the wafer.

- STEP 21. Gate Oxidation.

  A 110$\AA$ oxide of high integrity is grown at 900^oC, using the recipe described earlier. Note: steps 20, 21, 22 and 23 should be carried out in *immediate* succession, on the same day, to minimise the possibility of oxide contamination.

- STEP 22. 300$\AA$ Polysilicon Deposition .

  Wafers are *immediately* transferred to the polysilicon deposition furnace, where $\sim$300$\AA$ of polysilicon are deposited by LPCVD. This provides protection for the thin oxide film, and acts as a path to earth during ion implantation.

- STEP 23. 2nd Photomask .

  The "phantom gates" are defined in photoresist. These mask transistor

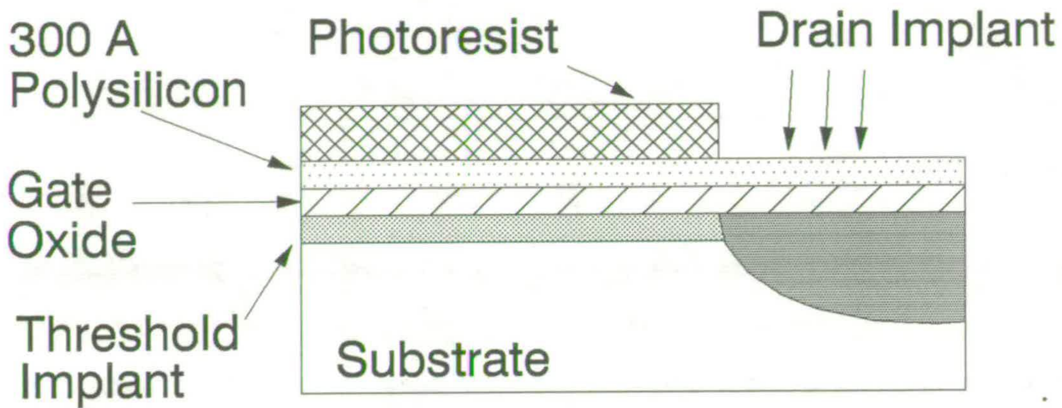

Figure 5–13: Step 24. Implantation of the Source and Drain Regions.

channel regions, during implantation of the source and drain. A high temperature photoresist (OFPR 800) is needed, since energy is dissipated in the photoresist as it absorbs incoming ions.

- STEP 24. Ion Implantation of Source-Drain Regions.

  Source-drain regions are formed by ion implantation. The polysilicon layer deposited in step 23, provides a path to ground for static charge[2]. This protects the gate oxide from static discharge. A thin polysilicon layer is used, so that ions may pass through without needing excessive acceleration energies. Two sets of implantations are required, one for arsenic transistors and one for phosphorus. This step is illustrated in figure 5–13.

  - Split 1. Arsenic dose=2×10^{15} atoms cm^{-2}, energy=160keV.
  - Split 2. Phosphorus dose=2×10^{15} atoms cm^{-2}, energy=75keV.

  Doses and energies are simulated in section 5.4.

- STEP 25. Photoresist Strip in Oxygen Plasma.

---

[2]A single rupture or pin hole in the gate oxide will be sufficient to connect the 300Å polysilicon layer to the earthed substrate. Once this low resistance path to earth is established, all subsequent static charge build up will be conducted through it.

- STEP 26. Photoresist Strip in Fuming Nitric Acid.

- STEP 27. 10% HCL Clean.

  A 10:1 solution of H_2O:HCL removes any sodium contamination from the wafer surface.

- STEP 28. 10% HF Dip.

  A 10:1 solution of H_2O:HF removes nascent oxide *immediately* prior to further polysilicon deposition. A short etch time $\sim$ 1 second is sufficient to remove nascent oxide.

- STEP 29. Polysilicon Deposition.

  LPCVD is used to deposit a further 6000Å of polysilicon, for the transistor gates.

- STEP 30. Ion Implantation to Dope the Polysilicon.

  The polysilicon is doped by ion implantation in two batches, one for arsenic transistors and one for phosphorus.

  - Split 1. Arsenic, dose=2×10^{16} atoms cm^{-2}, energy=50keV.
  - Split 2. Phosphorus, dose=4×10^{14} atoms cm^{-2}, energy=50keV.

  Doses and energies are simulated in section 5.4.

- STEP 31. High Temperature Anneal.

  A high temperature anneal in N_2 causes the source-drain implantations to diffuse laterally and vertically. To lessen any stress induced in the wafers, the temperature is ramped up to its anneal value, over 30 minutes, before the 60 minute anneal begins. Temperature is also ramped down after the anneal. Two anneals are given, one for each species of dopant.

  - Split 1. Arsenic, 60 minutes at 1072°C.
  - Split 2. Phosphorus, 60 minutes at 984°C.

  Times and temperatures are simulated in section 5.4.

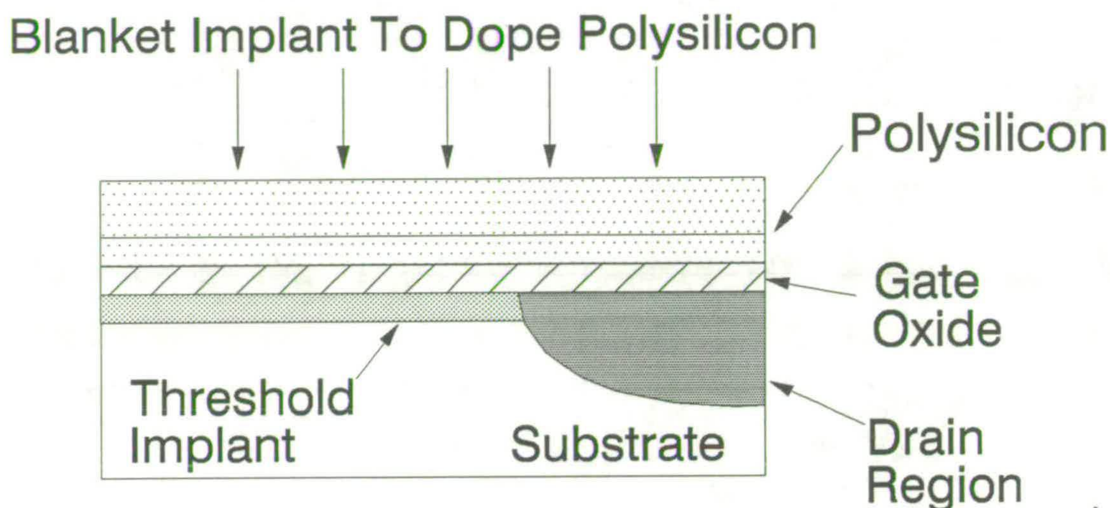

Figure 5–14: Step 30. Ion Implantation to Dope Polysilicon.

- STEP 32. Photoresist Coat.

  This protects the front of the wafers.

- STEP 33. Polysilicon Wet Etch.

  Polysilicon is removed from the wafer backs.

- STEP 34. Photoresist Strip in Fuming Nitric Acid.

- STEP 35. 3rd Photomask .

  The polysilicon gates are defined.

- STEP 36. Polysilicon RIE.

  RIE is used to etch vertically into the polysilicon so forming the gates. Care is needed to stop etching before the gate oxide is removed. Since plasma etching is more rapid at the wafer edges, dice at the wafer perimeter may be over etched, while those at the center may be under etched. This step is illustrated in figure 5–15.

- STEP 37. Photoresist Strip in Oxygen Plasma.

- STEP 38. Photoresist Strip in Fuming Nitric Acid.

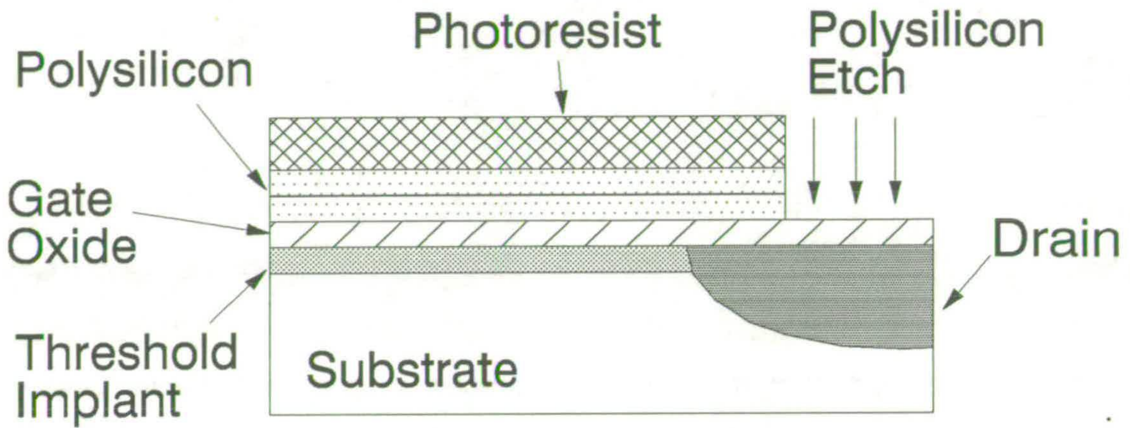

Figure 5–15: Step 36. Polysilicon RIE Etch.

- STEP 39. Interlevel Oxidation.

  350Å of oxide are grown at 950°C in steam and 5% HCL. This is equivalent to the interlevel oxide grown in an EEPROM process. Associated with this oxidation is a degree of "bird's beaking" under the gate, resulting in the so called Graded Gate Oxide (GGO) structure [22]. In addition, this step will remove any moisture or mobile ions from the wafer surface.

- STEP 40. Silicon Nitride Deposition.

  Wafers are immediately transferred to the nitride furnace, before any moisture can collect. A 2000Å layer of Si_3N_4 is deposited at 800°C, which will faithfully trace the contours of the surface. This will act as passivation, to protect the underlying devices, by providing a barrier to moisture and to the diffusion of mobile ions [21]. One feature which made Si_3N_4 particularly suitable, was the low temperature, 800°C. Contrast this to phosphosilicate glass which requires a reflow at 1050°C [23], and would upset the doping profiles tailored during the high temperature anneal.

- STEP 41. Oxidation.

  An oxide film is grown on the nitride surface, which takes 5 minutes at $950^{o}C$ in steam. This is used to promote adherence between the nitride layer and the overlying metallisation.

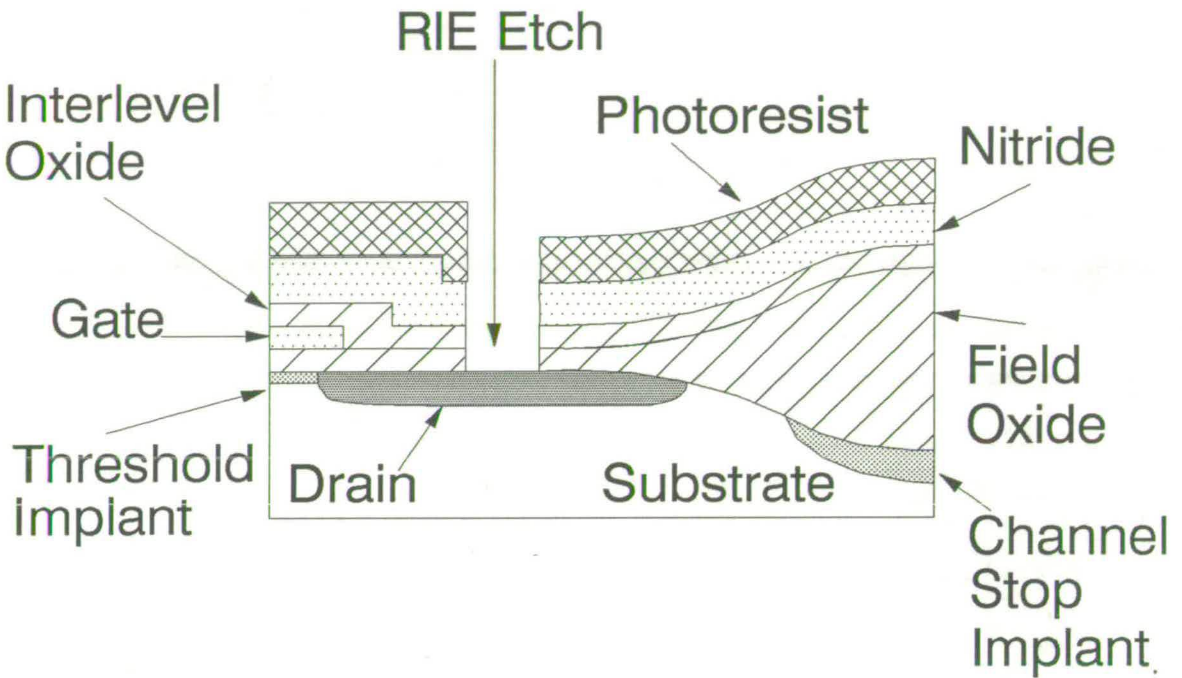

Figure 5–16: Step 43. Silicon Nitride RIE Etch.

- STEP 42. 4th Photomask.

  Contact holes are defined, for connection between the source, drain, gate and the metal layer.

- STEP 43. Silicon Nitride RIE.

  Silicon Nitride is removed from the contact holes. This etch also removes oxide from the contact holes, as illustrated in figure 5–16.

- STEP 44. Silicon Nitride RIE.

  Repeat with wafers face down to etch Si_3N_4 from backs.

- STEP 45. Photoresist Strip in Oxygen Plasma.

- STEP 46. Photoresist Strip in Fuming Nitric Acid.

- STEP 47. 25:1 HF Dip.

  A 25:1,NH_4:HF solution is used to clear contact holes of nascent oxide.

- STEP 48. Aluminium Deposition.

  A $1\mu m$ thick aluminium layer is deposited using a sputterer. Due to the large size of the contact holes ($4\mu m \times 4\mu m$), step coverage presents no problem.

- STEP 49. 5th Photomask.

  The metal interconnection pattern and probe pads are defined.

- STEP 50. Aluminium RIE.

  RIE is used to etch vertically into the aluminium, so forming the probe pads and connections to the transistors.

- STEP 51. Aluminium Wet Etch.

  Any remaining aluminium is removed with a wet etch solution containing orthophosphoric acid.

- STEP 52. Photoresist Coat.

  This protects the wafer front surface.

- STEP 53 . 4:1 HF Dip.

  Oxide is removed from the wafer backs, *immediately* before aluminium deposition.

- STEP 54. Aluminium Deposition.

  To give good contact to the substrate, aluminium is sputtered onto the wafer backs.

- STEP 55. Photoresist Strip in Fuming Nitric Acid.

- STEP 56 Sinter.

  A 20 minute sinter is given in forming gas at 430°C, to promote good contact between the aluminium and silicon. This is the final step of the process, the finished transistor is illustrated in figure 5–17

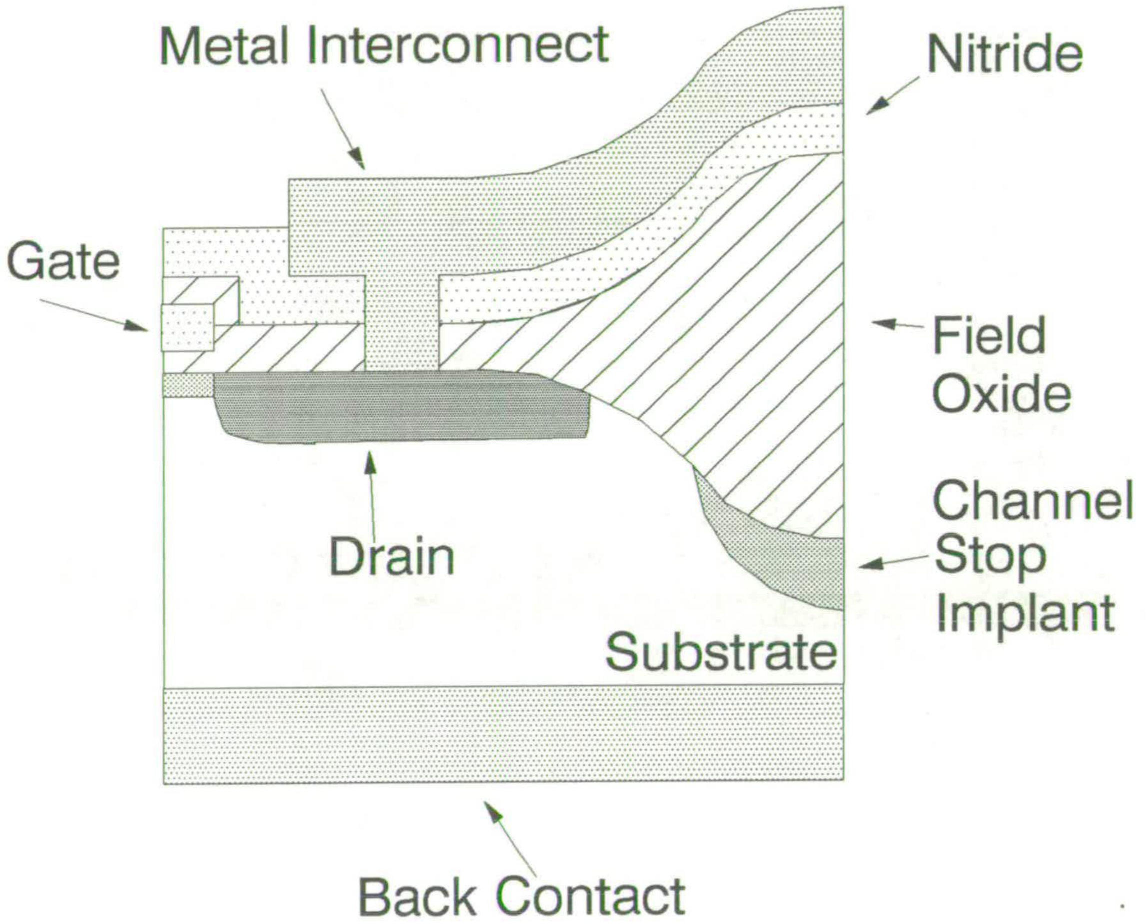

Figure 5–17: Step 56. Section Through a Completed Transistor.

5.3 Circuit Design

5.3.1 Reticle Production

The circuit was designed using the CAESAR, a software package for computer aided design. A reticle set was then made using an electron beam pattern generator, whose resolution was $0.167\mu m$. On a silicon wafer this would give a resolution of $0.0167\mu m$, since dimensions are reduced by a factor of 10 during optical lithography. To remain well within the pattern generators operating limits, $0.05\mu m$ was used as the minimum dimension during design work [3]. For practical purposes, this limited the increment possible in a POT array to $0.05\mu m$.

A design area of 5mm by 5mm was used, and due to the high cost of reticle production, optimum use was made of the available space. For *this* reason, a large number of transistor arrays and test structures were included. Figure 5–18 gives a schematic view of the design, while figure 5–19 is a plot of the design itself.

5.3.2 Transistor Design

Each POT column contains 20 transistors, which are closely grouped together. A single array covers a distance of only $960\mu m$. During fabrication, processing parameters such as photoresist thickness and photographic illumination vary across the wafer [4]. Close grouping helps to minimise the variation in processing parameters from one transistor to another. The dimensions of the POT transistors are summarised in table 5–1.

---

[3]CAESAR can only create dimensions in units of $0.02\mu m$. To overcome this, the design was drawn $10\times$ oversize, then reduced during its loading into the pattern generator.

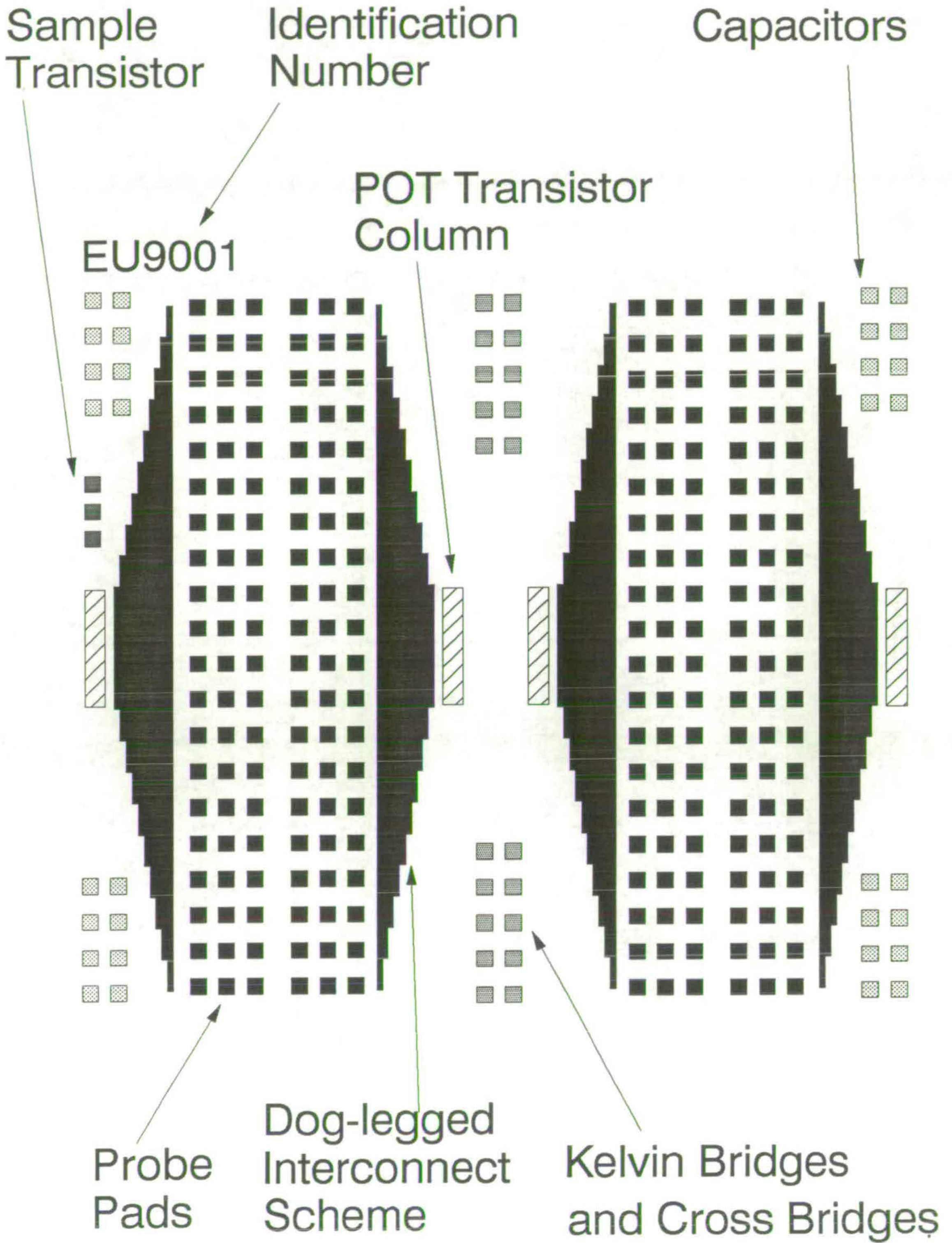

Figure 5–18: Schematic Diagram of Reticle Design

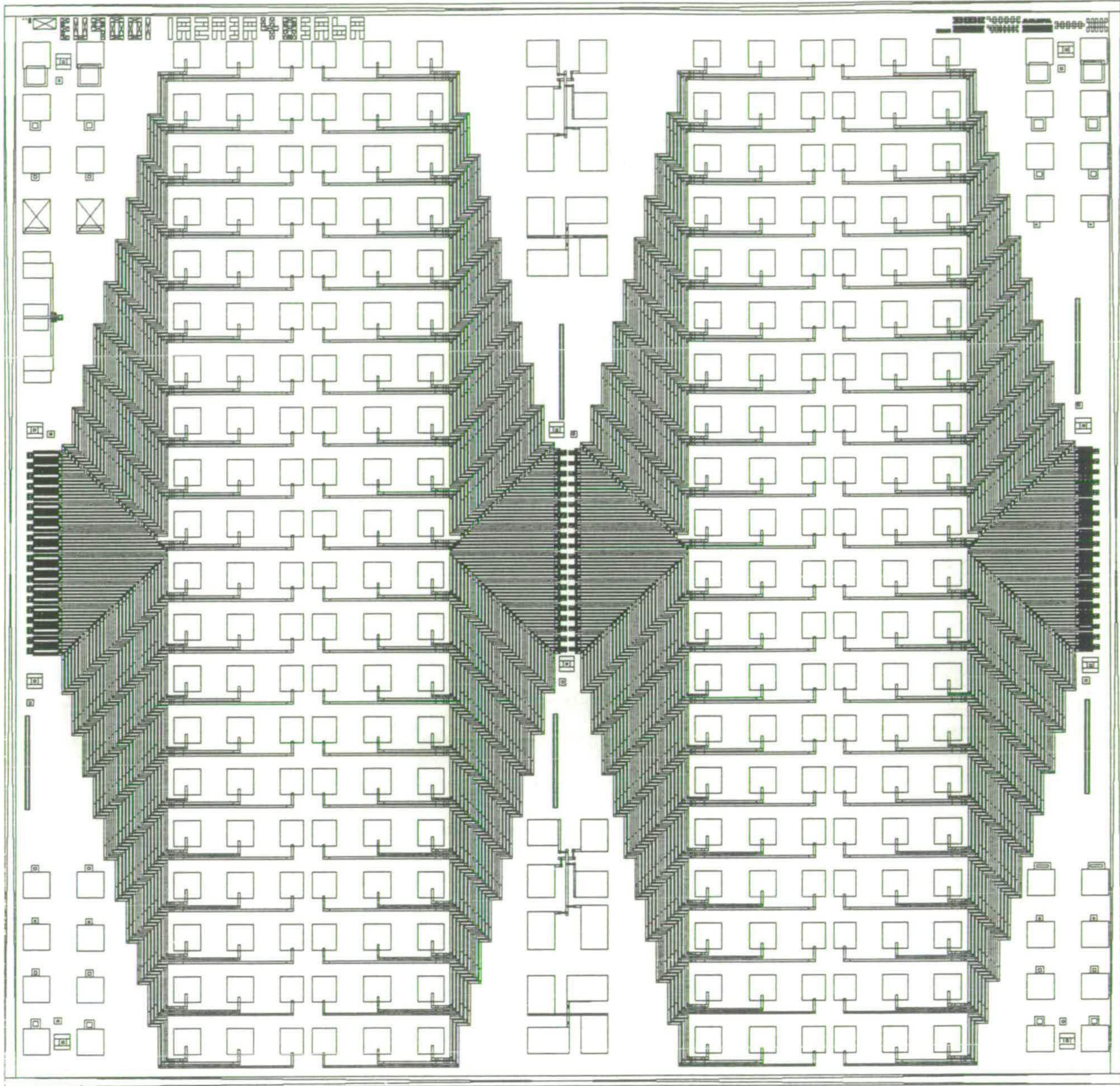

Figure 5–19: Plot of the Reticle Design

Array	Length of Phantom Gate	Length of Polysilicon Gate	Width of the Transistor	Lateral Drain Diffusion	POT Step Size
A	$3\mu m$	$2.6\mu m$	$5\mu m$	$0.45\mu m$	$0.05\mu m$
B	$6\mu m$	$5.6\mu m$	$50\mu m$	$0.45\mu m$	$0.05\mu m$

Table 5–1: Design Parameters for Transistor Arrays

- Array A. This contains small geometry transistors, $2.6\mu m$ long by $5.0\mu m$ wide, which are directly equivalent to the EEPROM. With their small gate areas array A devices should be relatively insusceptible to oxide defects.

- Array B. These transistors are $5.6\mu m$ long by $50.0\mu m$ wide, which will increase tunnel currents by an order of magnitude during testing. These devices were included, should the tunnel currents in array A devices prove difficult to monitor. However, array B devices will be more susceptible to tunnel oxide defects.

Diffusion of the Source-Drain Regions

Following ion implantation a high temperature anneal is given, which causes the the source-drain regions to diffuse laterally. One would like to test reliability over a wide range of overlap values, and a large lateral diffusion is therefore preferred [4]. A lateral diffusion of $0.45\mu m$ was chosen, as compared to $\simeq 0.3\mu m$ gate/drain overlap in an EEPROM. This gave a large overlap, while still retaining a realistic profile.

[4] Oxide outside the laterally diffused regions has been damaged during ion implantation, and may not give meaningful results.

Processing Friendly POT

In principle a POT array should have gapped devices at the beginning and end, and fully overlapped devices at the centre [3]. However, processing can be unpredictable and may give a smaller *or* larger source-drain diffusion than designed. By drawing the polysilicon gate $0.4\mu m$ shorter than the phantom gate, the design will tolerate a variation of $\pm 0.25\mu m$. Thus, the leeway for process variation is optimised for *either* circumstance, this is illustrated in figure 5–20.

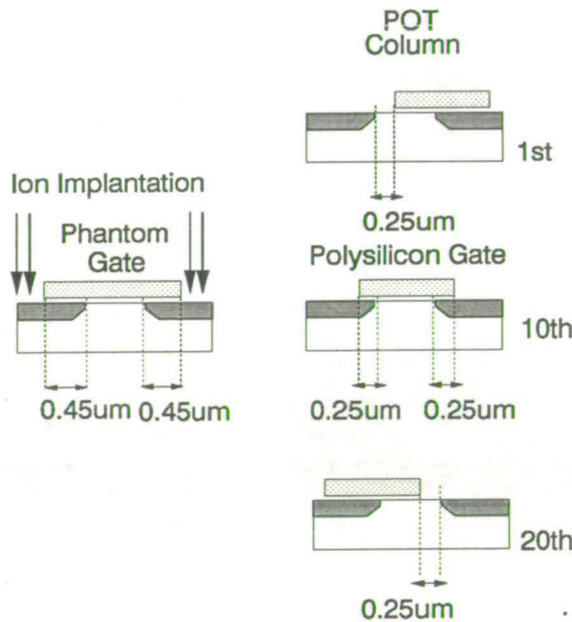

Figure 5–20: Processing Friendly POT.

5.3.3 Interconnection Scheme

A "dog legged" scheme was used for metal interconnections between transistors and probe pads, which allowed more space on the reticle for other structures. Metallisation and contact holes were designed in accordance with the Edinburgh Microfabrication Facility's $6\mu m$ design rules, and standard EMF pad sizes of $120\mu m$ by $120\mu m$ were used. These conservative design rules ensure that defects in metallisation are minimised.

5.3.4 Test Structures

Sample Transistor

To quickly identify working dice during testing, a *sample* transistor was included. This was an ordinary, fully overlapped transistor, whose good operation indicates a successful process. In order to ensure a fully overlapped transistor, the real polysilicon gate was longer than the phantom photoresist gate.

Kelvin Bridges And Cross Bridge Structures

Standard test structures were reproduced directly from the Edinburgh Microfabrication Facility's test strip [24]. These were added as a precaution, to help in analysing problems, should any transistors display aberrant characteristics.

1. Kelvin Bridges are designed to monitor metal-polysilicon and metal-drain contact resistances.

2. Cross Bridge Structures are designed to monitor polysilicon and drain sheet resistances.

5.4 Simulation

Once the required doping profiles have been determined, the implantation parameters and thermal budget to meet this specification must be calculated. In general, the commitment of a semiconductor process to silicon is a costly and time consuming affair. However, with careful simulation, the process may be tailored beforehand, to give the desired results. Although the full gamut of process steps can be simulated with today's programs, simulations offer the most help when deciding:

1. Thermal Budget.

   Allocation of anneal temperatures, anneal times, etc.

2. Implantation Energies.

3. Implantation Doses.

In contrast, parameters such as gate oxide thickness are very sensitive to small fluctuations in processing conditions. These include the temperature within a furnace [8], which is never uniform, traces of moisture in the oxidising ambient [9] and the pre-oxidation clean [25]. Here, a simulator would need to be finely calibrated to a given fabrication facility, before reliable results could be produced. Even with the aid of software tools, the designer must use his intuition during the design process.

Technology Modelling Associates (TMA) software was used throughout this project. They offer two packages for process simulation:

1. SUPREM-3. For one dimensional simulations.

2. TSUPREM-4. For two dimensional simulations.

One dimensional simulations give *faster* results than two dimensional simulation. This is because, the size of a 1D simulation is by its nature much smaller than for 2D. SUPREM-3 can be used to calculate *electrical parameters*, such as threshold voltage and sheet resistance. For these reasons, SUPREM3 was used in the majority of simulations. Four sets of parameters must be decided for each batch:

1. Drain implantation, dose and energy.

2. Polysilicon gate implantation, dose and energy.

3. Anneal temperature and time.

4. Threshold adjust implantation,dose and energy.

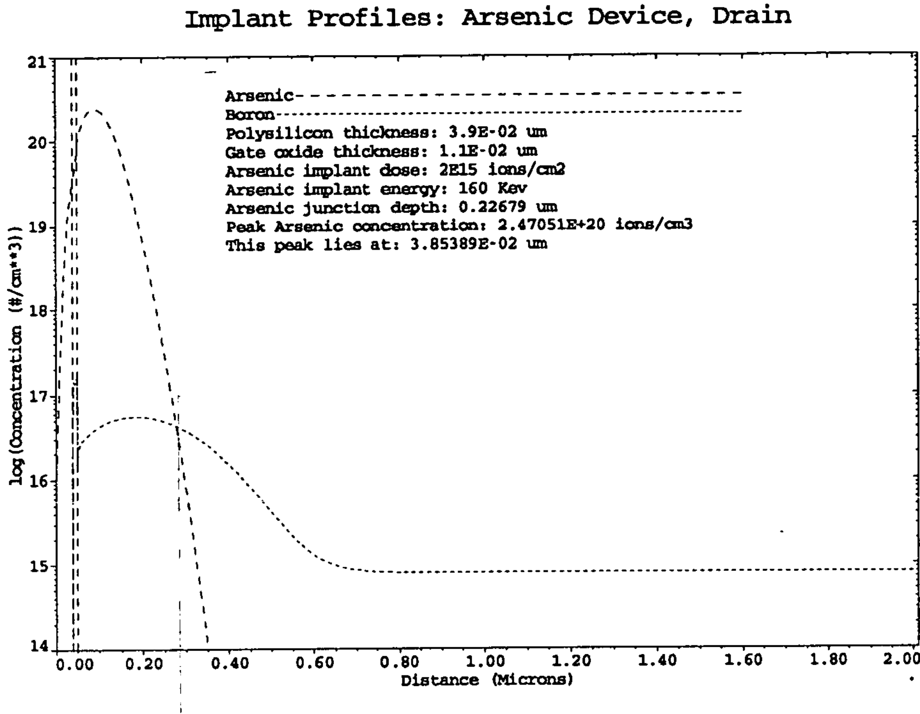

Figure 5–21: Implantation to Form Arsenic Drain Region.

5.4.1 1 Dimensional Simulation of Arsenic Transistors

Drain Implantation Energy

The drain implantation passes through a 500Å thick sandwich, of SiO_2 and polysilicon, before reaching the substrate. A high energy ($160keV$) was therefore required. This was near the limit of the ion implanter's capability, and gave a junction depth of $\simeq 0.225\mu m$, as illustrated in figure 5–21.

To a first approximation, diffusion may be assumed to be isotropic [26] [27]. Thus, the verticle junction depth calculated by SUPREM-3, can be used to estimate the degree of lateral diffusion. If the drain profile diffuses $0.45\mu m$ vertically, it will also diffuse $\simeq 0.45\mu m$ laterally. Since the required lateral diffusion is $0.45\mu m$, the required junction depth will be give by equation 5.1:

$$Junction\ depth \simeq 0.225\mu m + 0.45\mu m = 0.675\mu m \tag{5.1}$$

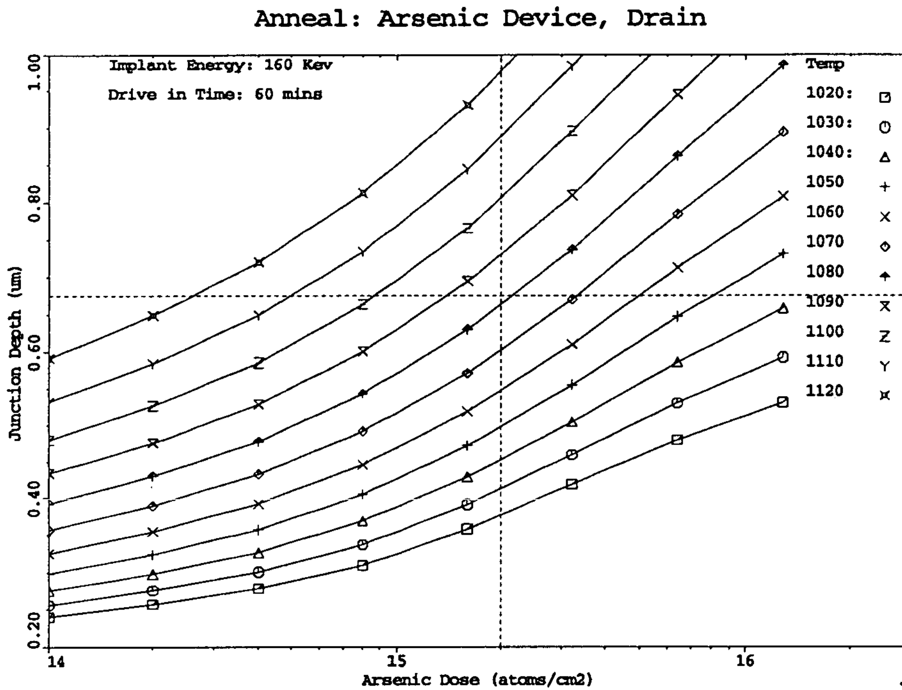

Figure 5–22: Junction Depth as a Function of Anneal Temperature and Implantation Dose.

Drain Implantation Dose and Anneal Conditions

An anneal time of 60 minutes was chosen, leaving the anneal temperature and implant dose to be decided. The phantom gate is composed of high temperature photoresist, which must absorb energy from the incoming ions, during implantation. To reduce the power dissipated in this resist, and so reduce the possibility of it charring, a *low dose* is preferred. A low dose will also help in emulating the profiles of a standard EEPROM, which has a high drain resistance [28]. However, a *low temperature* has the advantage of maintaining oxide quality [11]. Figure 5–22 illustrates how implantation dose and anneal temperature are related to junction depth. A dose of $2.0 \times 10^{15} atoms\ cm^{-2}$ was chosen, requiring an anneal temperature of 1080^oC. This gives a junction depth of $0.675\mu m$, which is equivalent to a lateral diffusion of $\simeq 0.45\mu m$.

Gate Implantation

The polysilicon gate is implanted with arsenic to increase its conductivity, and an energy of 50keV was used. Figure 5–23 indicates that a dose of $2 \times 10^{16} atoms\ cm^{-2}$, gives a sheet resistance of $\sim 50\Omega/\square$.

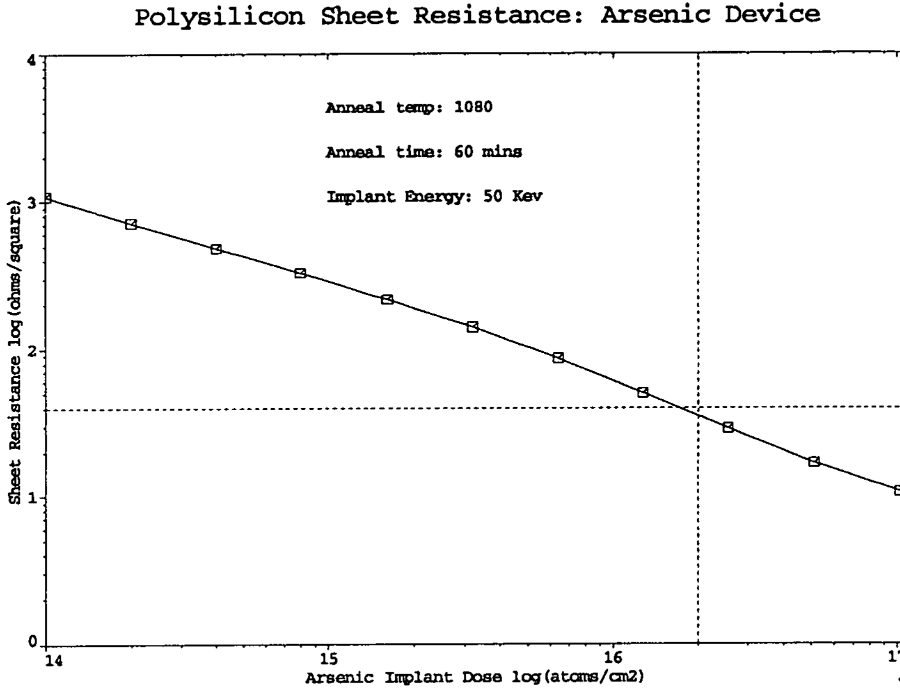

Figure 5–23: Polysilicon Sheet Resistance as a Function of Implant Dose.

Drain Profile

In the finished anneal recipe, the temperature was ramped up to the required value, over 30 minutes. It was then ramped down for 30 mins at the end. These ramps are introduced to improve the oxide quality [7]. Figure 5–24 gives the resulting one dimensional drain doping profile. The anneal conditions required to give $0.675\mu m$ junction depth, was calculated automatically by SUPREM-3, giving:

1. 30 min ramp, $922^oC \Rightarrow 1072^oC$
2. 60 min anneal, 1072^oC

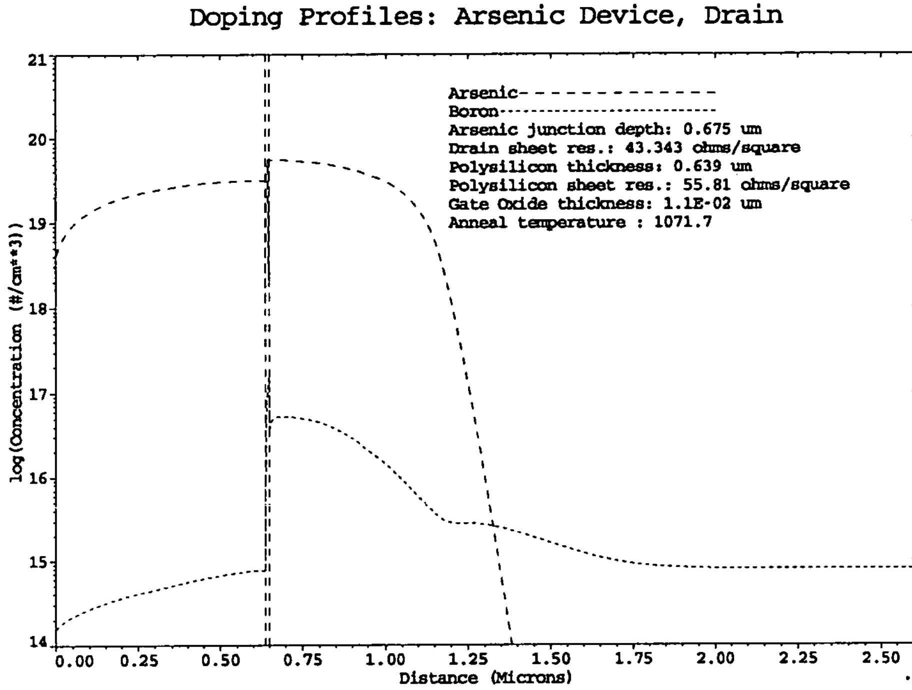

Figure 5–24: Doping Profiles of the Drain Region.

3. 30 min ramp, $1072^{\circ}C \Rightarrow 922^{\circ}C$

The program for this simulation is given in appendix D.

Threshold Adjust Implantation

Threshold implant dose and energy were based on the Edinburgh Microfabrication Facility's $6\mu m$ process. The channel region doping profile, and threshold voltage of the transistor are given in figures 5–25 and 5–26 respectively. It is seen that arsenic diffusion through the gate oxide is negligible, and the threshold voltage is positive.

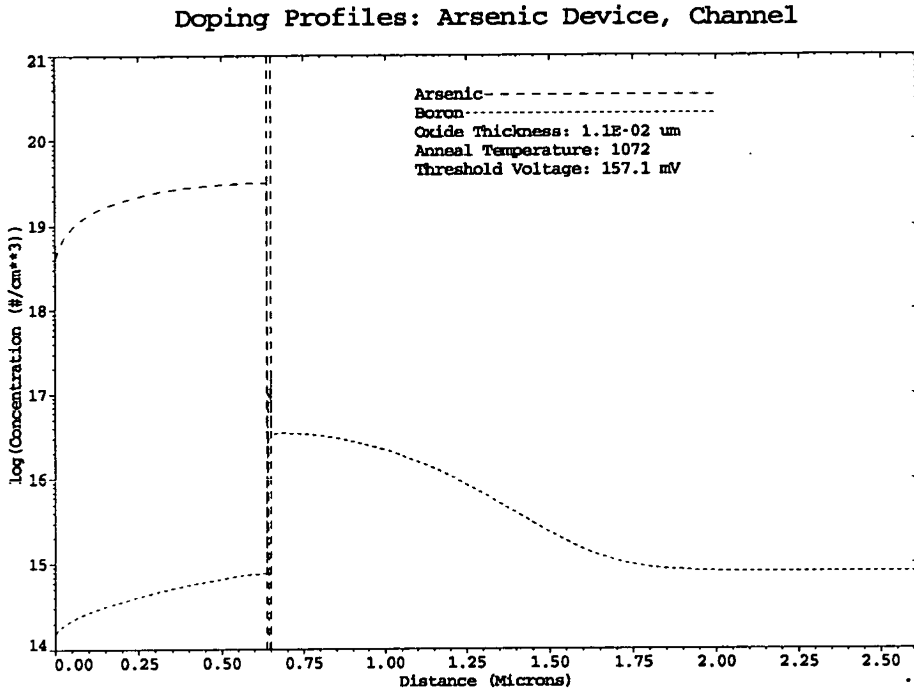

Figure 5–25: Doping Profiles of the Channel Region.

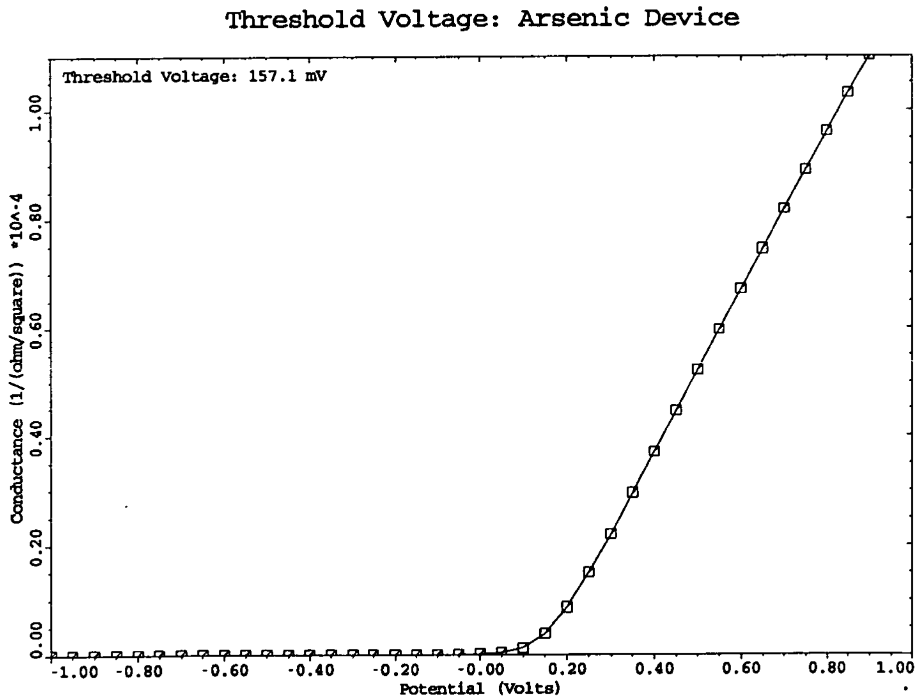

Figure 5–26: Threshold Voltage Plot for the Arsenic Transistor.

5.4.2 1 Dimensional Simulation of the Phosphorus Transistors

Drain Implantation

The phosphorus transistors should have the same doping profiles as the arsenic transistors. For this reason, the same implantation dose is used for each, $2 \times 10^{15} atoms\ cm^{-2}$. SUPREM-3 is then used to automatically calculate the required energy. An energy of 75Kev gave an junction depth of $0.225\mu m$, see figure 5–27.

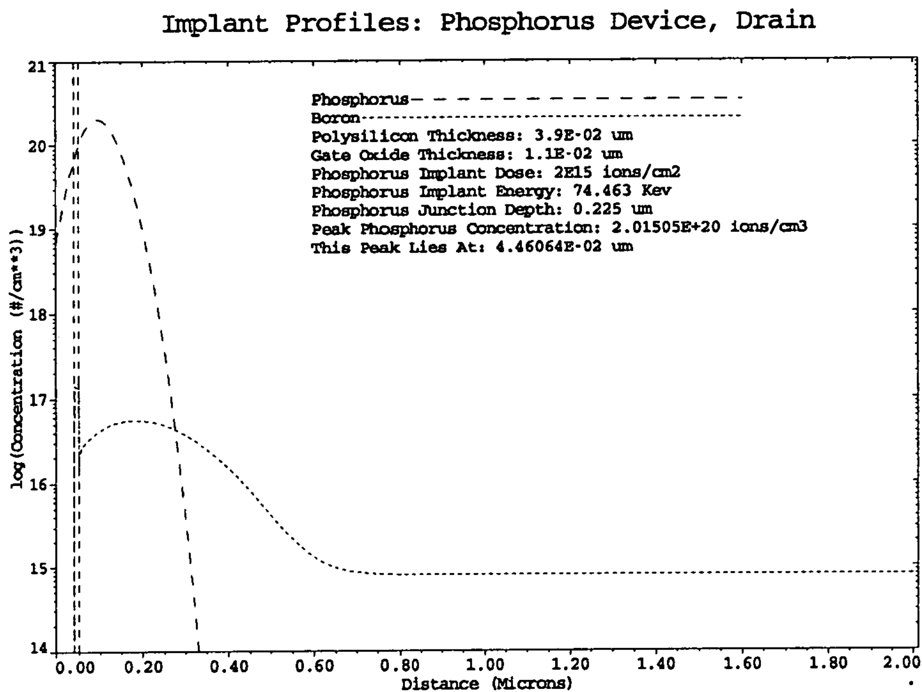

Figure 5–27: Implantation Profile of Phosphorus Transistor.

Gate Implantation

For phosphorus and arsenic devices to be equivalent, each should each have the same polysilicon sheet resistance. Again, an energy of 50keV was used. A dose of $4 \times 10^{15} atoms\ cm^{-2}$ gives a sheet resistance of $\sim 50\Omega/\Box$, as illustrated in figure 5–28.

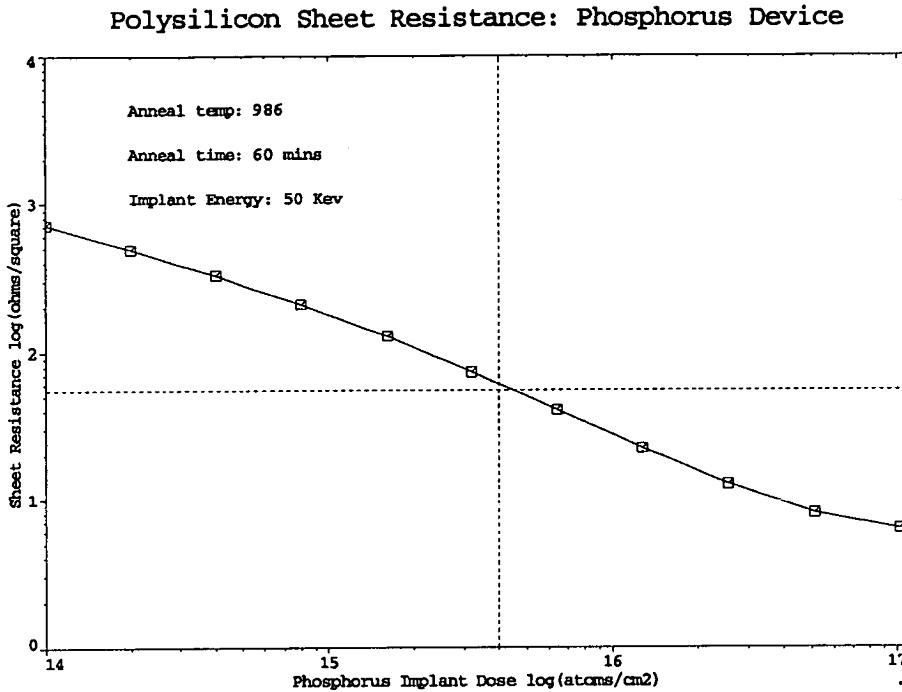

Figure 5–28: Polysilicon Sheet Resistance, as a Function of Implant Dose.

Anneal Conditions

Figure 5–29 shows the profiles for the phosphorus drain region. The anneal conditions required to give $0.675\mu m$ junction depth, were calculated automatically by SUPREM-3, giving:

1. 30 min ramp, $836^oC \Rightarrow 986^oC$
2. 60 min anneal, 986^oC
3. 30 min ramp, $986^oC \Rightarrow 836^oC$

Threshold Adjust Implantation

The same threshold adjust implant was given for each set of transistors. The doping profiles for the phosphorus device channel region is given in figure 5–30, indicating that phosphorus diffusion though the gate oxide is insignificant. Figure 5-31 indicates that these have a threshold voltage $\sim 160mV$.

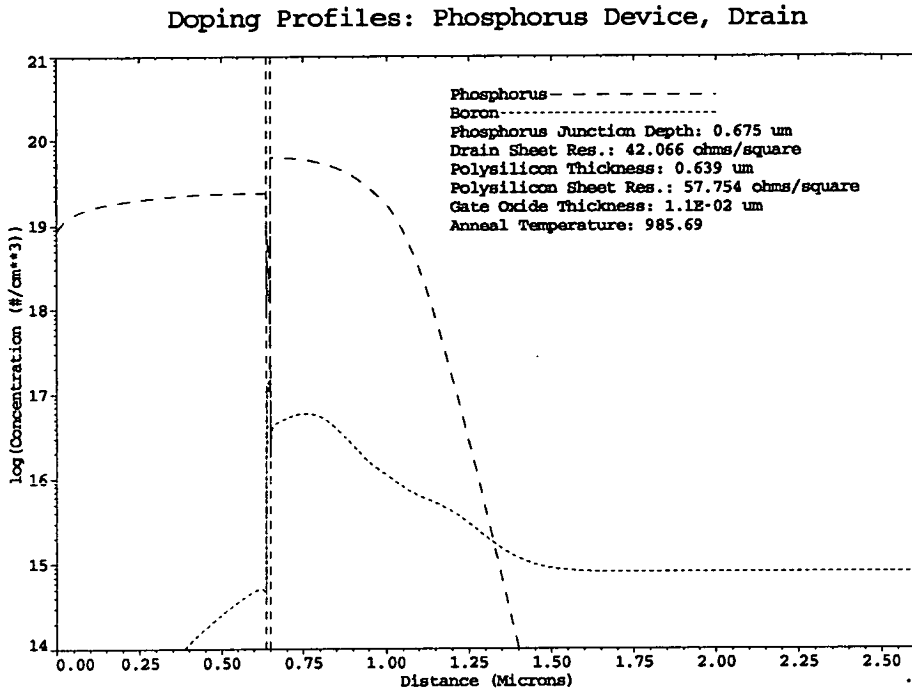

Figure 5–29: Drain Profiles for Phosphorus Transistor.

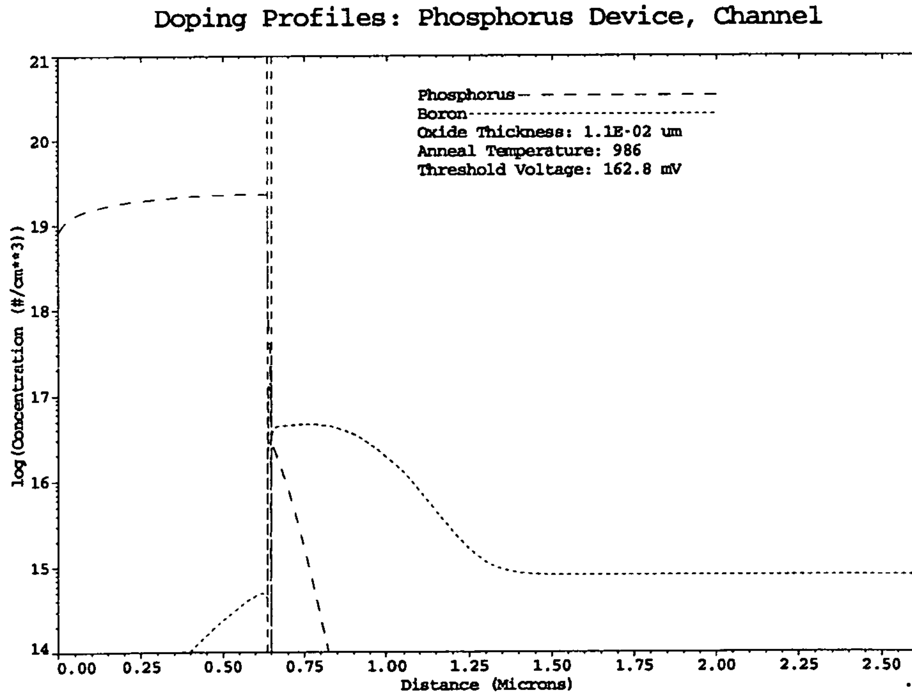

Figure 5–30: Doping Profile of Channel Region.

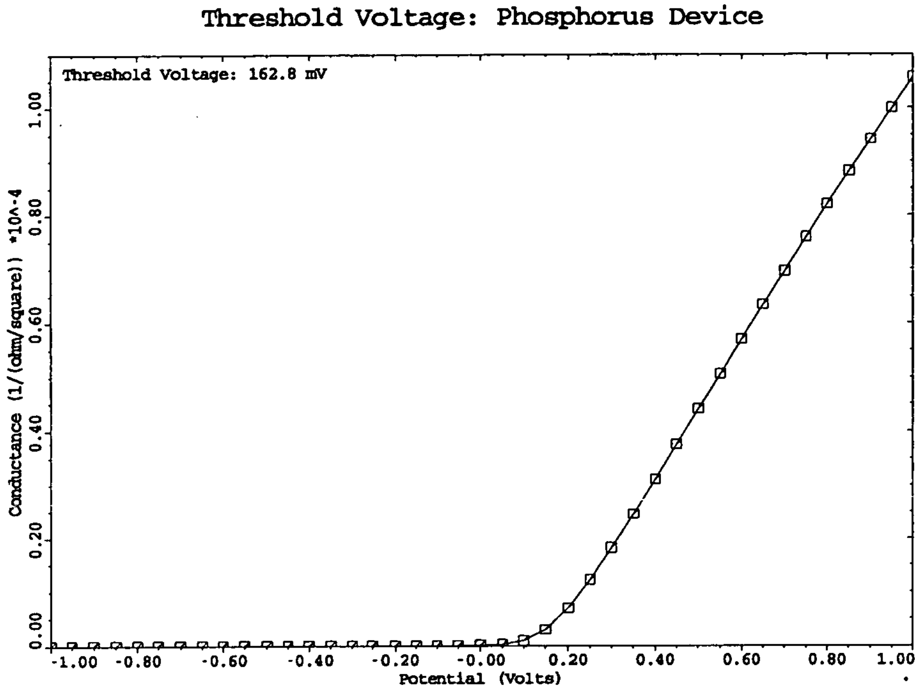

Figure 5–31: Threshold Voltage Plot of Channel Region.

5.4.3 2 Dimensional Simulation of Transistors

The gate/drain overlaps were checked using TSUPREM-4. Although this cannot extract electrical parameters, and is time consuming, it can simulate in two dimensions. Figure 5-32 gives a two dimensional plot of the drain region. Figure 5-33 illustrates the doping concentrations, along a line taken horizontally though the channel and drain. This indicates a gate/drain overlap of $\simeq 0.46\mu m$, the program used to simulate this is given in appendix D.

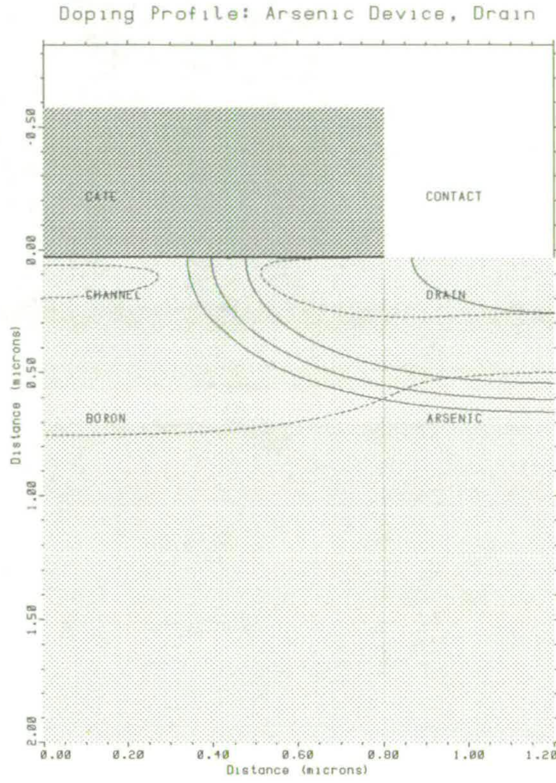

Figure 5–32: Drain Region of the Arsenic Transistor.

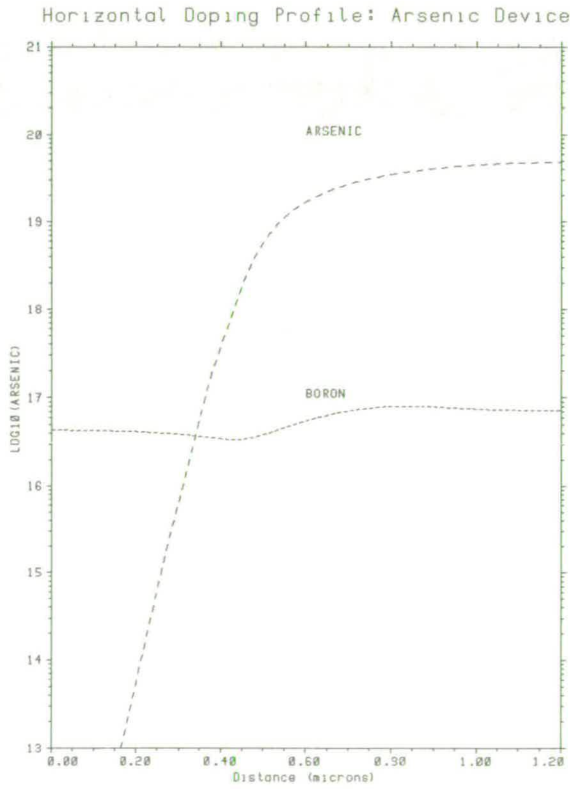

Figure 5–33: Horizontal Doping Profile.

5.5 Conclusion

A process has been designed, to produce an array of progressional offset transistors. Special emphasis is placed on the producing and maintaining a good integrity oxide, as in standard EEPROM devices. These may now be used to assess the improvement in programming reliability, caused by an increase in the gate/drain overlap. A contrast between the reliability of phosphorus and arsenic transistors may also be made.

Bibliography

[1] Ih-C.Chen, S.E.Holland, and C.Hu. Electrical breakdown in thin gate and tunneling oxides. *IEEE Transactions On Electron Devices*, 32(2):413–422, February 1985.

[2] R.C.Smith and A.J.Walton. The design, fabrication and measurement of assymetrical LDD transistors. 1992.

[3] J.Serack. *Edge Effects In Silicon IGFETs.* PhD thesis, University Of Edinbugh, 1988.

[4] H.Neves. *Hot Carrier Effects In IGFETs.* PhD thesis, University Of Edinbugh, 1991.

[5] T.Y.Chan, A.T.Wu, P.K.Ko, and C.Hu. Effects of gate-to-drain/source overlap on MOSFET characteristics. *IEEE Electron Devices Letters*, 8(7):326–328, 1987.

[6] A.Iranmanesh, M.Biswal, and B.Bastani. Total system solution with advanced BiCMOS. *Solid State Technology*, 35(7):37–40, 1992.

[7] S.S.Cohen. Electrical properties of post-annealed thin SiO_2 films. *J.Electrochem.soc Solid State Science And Technology*, 130(4):929–932, April 1983.

[8] H.Z.Massoud, J.D.Plummer, and E.A.Irene. Thermal oxidation of silicon in dry oxygen growth-rate enhancement in the thin regime. *J.Electrochem.soc Solid State Science And Technology*, 132(11):2685–2693, November 1985.

[9] E.A.Irene and R.Ghez. Silicon oxidation studies: The role of H_2O. *J.Electrochem.soc Solid State Science And Technology*, 124(11):1757–1761, November 1977.

[10] J.Ruzyllo. Effects of preoxidation ambient in very thin thermal oxide on silicon. *J.Electrochem.Soc Solid State Science And Technology*, 133(8):1677–1682, 1986.

[11] S.Holland and C.Hu. Correlation between breakdown and process-induced positive charge trapping in thin thermal SiO_2. *J.Electrochem.soc Solid State Science And Technology*, 133(8):1705–1712, August 1986.

[12] K.Hofmann, D.R.Young, and G.W.Rubloff. Hole trapping in SiO_2 films annealed in low-pressure oxygen atmosphere. *Journal Of Applied Physics*, 62(3):925–930, 1987.

[13] H.Shirai, K.Kanya, and A.Yamaguchi. Effect of oxide-induced stacking faults on dielectric breakdown characteristics of thermal silicon dioxide. *J. Appl. Phys.*, 66(11):5651–5653, December 1989.

[14] S.Chichibu, T.Harada, and S.Matsumoto. Effect of dopant concentration on oxidation -induced stacking faults in boron-doped CZ silicon. *Japanese Journal Of Applied Physics*, 27(8):1543–1545, 1988.

[15] C.Hashimoto, S.Muramoto, N.Shiono, and O.Nakajima. A method of forming thin and highly reliable gate oxides. *Journal Of The Electrochemical Society*, 127(1):129–135, 1980.

[16] C.S.Rafferty, L.Borucki, and R.W.Dutton. Plastic flow during thermal oxidation of silicon. *Applied Physics Letters*, 54(16):1516–1518, 1989.

[17] S.R.Stiffler. Oxidation-induced substrate strain in advanced silicon integrated-circuit fabrication. *Journal Of Applied Physics*, 68(1):351–355, 1990.

[18] P.Murray and G.F.Carey. Determination of interfacial stress during thermal oxidation of silicon. *Journal Of Applied Physics*, 65(9):3667–3670, May 1990.

[19] T.B.Hook and T-P.Ma. Electron trapping during high field tunneling injection in metal-oxide-silicon capacitors: The effect of gate induced strain. *Journal Of Applied Physics*, 62(3):931–938, 1987.

[20] K.Hofmann, G.W.Rubloff, M.Liehr, and D.R.Young. High temperature reaction and defect chemistry of the Si/SiO_2 interface. In *INFOS*, pages 25–30, 1987.

[21] S.M.Sze, editor. *VLSI Technology*, chapter 6. McGraw-Hill International Editions, 1988.

[22] P.K.Ko, S.Tam, and C.Hu. Enhancement of hot electron currents in graded-gate-oxide(GGO)-MOSFETs. In *IEDM*, pages 88–91, 1984.

[23] W.P.Ruska. *Microelectronic Processing*, page 278. McGraw-Hill International Editions, 1988.

[24] W.R.Gammie. *The 5 Micron Edinburgh Microfabrication Facility Test Strip.* 1987.

[25] G.Gould and E.A.Irene. The influence of silicon surface cleaning procedures on silicon oxidation. *J.Electrochem.soc Solid State Science And Technology*, 124(11):1031–1033, April 1987.

[26] W.P.Ruska. *Microelectronic Processing*, page 81. McGraw-Hill International Editions, 1988.

[27] S.M.Sze, editor. *VLSI Technology*, page 320. McGraw-Hill International Editions, 1988.

[28] C.Kuo, Y.R.Yeargain, and W.J.Downey. An 80ns 32K EEPROM using the FETMOS cell. *IEEE J.Solid State Circuits*, (5):821–827, October 1982.

Chapter 6

Analysis of EEPROM Structures

The novel test structures described in chapter 5 may be used to investigate EEPROM reliability. Avenues for exploration are two fold:

1. Results from phosphorus and arsenic doped devices should be compared, in order to evaluate the effect of chemistry upon reliability.

2. The effect of floating gate/drain overlap upon reliability should be studied. A contrast may then be made between model predictions, produced in chapter 4, and experimental data. Thus, any phenomena which were not detected in the model may be high-lighted.

Of principle interest is the end-of-life wearout, hence devices of good integrity are required. A discussion of transistor characteristics and quality, will therefore be useful.

6.1 Process Quality

IC quality is not to be confused with IC reliability. Quality describes how closely an IC conforms to its specifications, directly after completion of the process, and is a manufacturing concern. Whereas, reliability describes how closely an IC continues to conform to its specifications over years of use [1].

6.1.1 Wafer Yield

In all 8 wafers were processed, each containing 88 dice, from which there was a degree of yield loss. A *sample* MOS transistor was included on every die. If this displayed good characteristics then the die as a whole was assumed to be good. To test this the drain was held at 2.0V, the gate ramped from 0.0V to 2.0V, and the source and substrate were earthed. A good working device was one which gave:

- Drain current $< 1 \times 10^{-10} A$, for $0V$ gate voltage.
- Drain current $> 1 \times 10^{-7} A$, for $2V$ gate voltage.

The test was automated using a KLA Automatic Wafer Prober and a Hewlett-Packard 4062B Semiconductor Parametric Test System. This allowed every sample transistor to be measured efficiently. A map of working dies was produced for each wafer, of which the highest yielding is illustrated in figure 6–1.

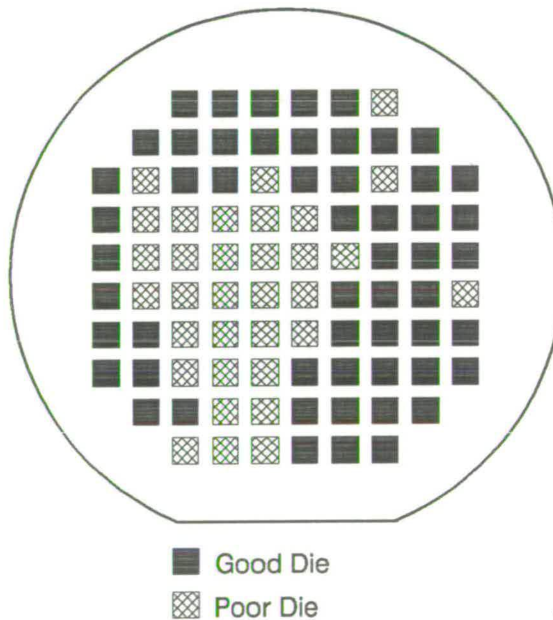

Figure 6–1: Map of Working Dies on a Wafer.

Devices around the perimeter of the wafer work well, while those at the center had a short circuit between the source and drain. This indicates that some

polysilicon has been left behind, during gate etching. Inhomogeneity of plasma etching would account for this result, since plasma etching acts more quickly at the perimeter of a wafer than it does at the center [2]. This type of yield loss was a common feature of the batch. In figure 6–1 55 dice were good, giving a yield of $\simeq 60\%$.

6.1.2 Gate Oxide Thickness

A bare wafer was included in the batch during gate oxidation, in order to measure gate oxide thickness. Figure 6–2 gives a map of oxide thickness, measured using a GRQ Instruments E-Probe 200, automatic ellipsometer. The oxide has an average thickness of $119\AA$, which compares favourably with the target value of $110\AA$, and a good uniformity.

6.1.3 Threshold Voltage

During reliability testing, a positive voltage is applied to the n-type drain, while the gate is earthed. A current should then flow between the drain and gate, to emulate EEPROM programming [3]. However, *no* current should flow between the channel and gate. A positive threshold voltage is therefore required, so that the p-type channel does not enter inversion. The threshold voltage may be defined as the gate voltage at which a transistor begins to conduct. A more physically meaningful criterion is given by assuming the SPICE standard transistor model for threshold voltage determination [4]. This method was adopted to measure threshold voltage on each wafer. The drain was held at $0.1V$, the gate was ramped from $0V$ to $1V$, while the source and substrate were earthed. Results gave similar threshold voltages for both arsenic and phosphorus devices, $\simeq 0.6V$. It was expected that threshold voltages should be similar, since all wafers received the same threshold adjust implant. Figure 6–3 gives a typical threshold plot, for a transistor in the centre of a POT array.

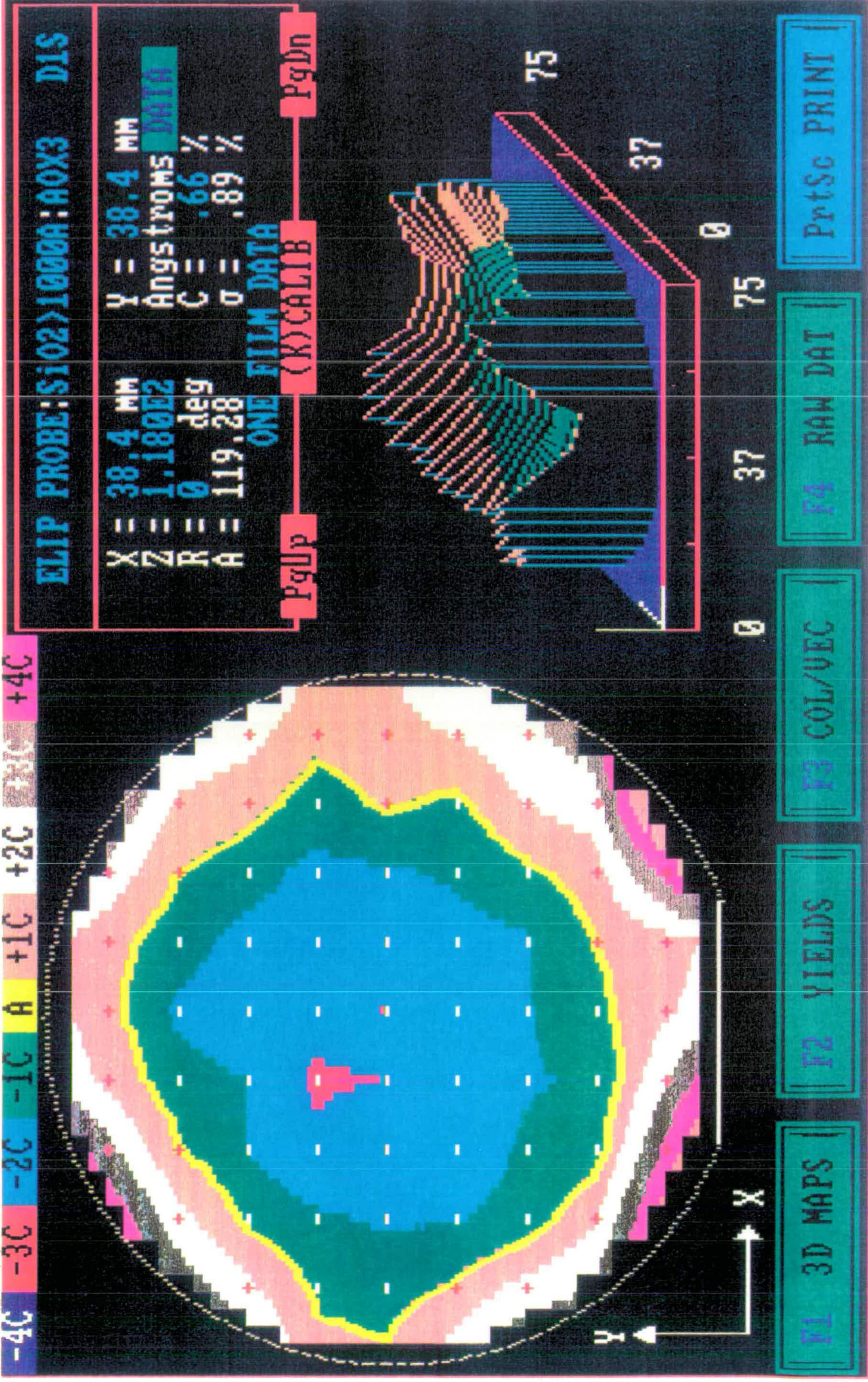

Figure 6–2: Oxide Thickness Across a Wafer, Measured Using Automatic Ellipsometery.

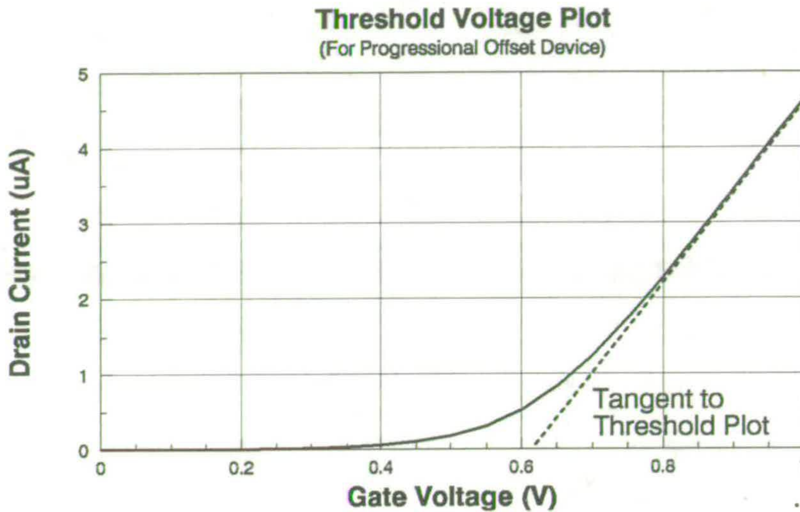

Figure 6–3: Threshold Characteristic for a POT Transistor, 5μm Wide.

6.1.4 Drain Characteristics

Drain characteristics helped verify the correct operation of the transistors. For these the drain was ramped from 0V to 4V, the gate voltage was incremented in steps of 1.0V, while the source and substrate were earthed. Phosphorus and arsenic devices each gave similar characteristics, which was to be expected as their geometries were equal. A typical characteristic is given in figure 6–4, taken from a device in the centre of a POT array.

6.2 Assessment of POT Array Symmetry

To recap, a POT array contains 20 transistors, which have a range of gate/drain overlaps. Ideally the $10th$ device in an array should be symmetrical, with equal gate/drain and gate/source overlaps. However, inaccuracies are introduced during optical lithography. These result in a skew arrays, as illustrated in figure 6–5. Locating the symmetrical device is the first step in analysing a POT array.

This problem has been given detailed attention by previous authors [5] [6] [7]. Asymmetrical transistors, with unequal gate/drain and gate/source overlaps,

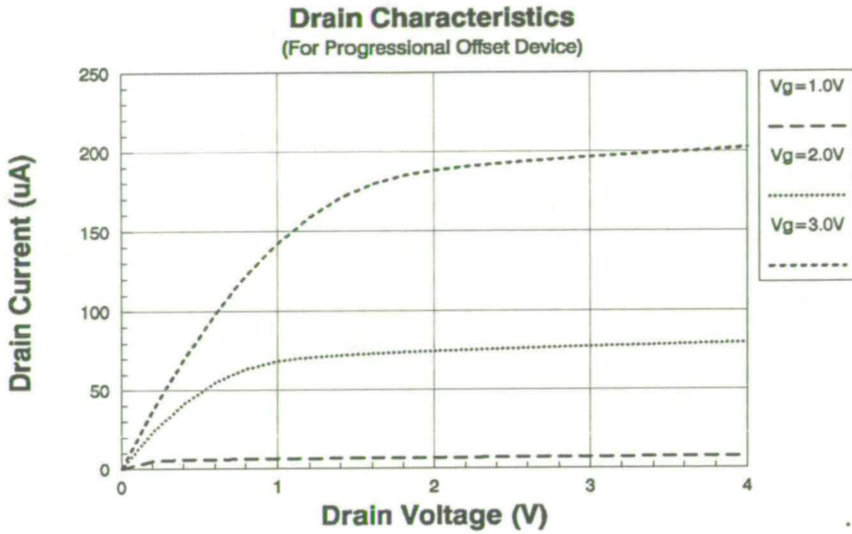

Figure 6–4: Drain Characteristic for a POT Transistor, 5μm Wide.

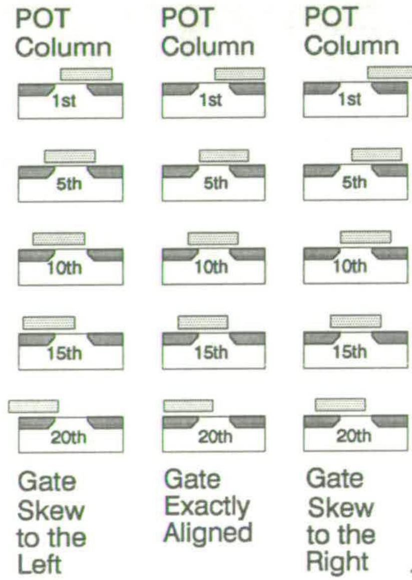

Figure 6–5: Schematic Diagram, Illustrating the Gate/Drain Overlap in Three POT Columns.

have different electrical characteristics when the source and drain are exchanged. The difference is especially pronounced when a gap appears between the gate and drain, or source. The subthreshold current of a transistor may be used to indicate whether the device has drain end gap, or source end gap. This technique provides a particularly sensitive test [6].

In the subthreshold regime, the channel is in weak inversion, and the drain current is dominated by diffusion [8] [9]. For a drain voltage $V_d > \frac{3KT}{q}$, current is independent of V_d, and rises rises exponentially with gate voltage V_g [10]. This mode of operation is equivalent to a bipolar transistor, where the source, gate and drain are equivalent to the emitter base and collector respectively [8].

In a POT transistor, a gap at the source end will cause a significant drop in the subthreshold current. This is due to an increase in the series resistance, and a reduction in the "emitter efficiency". However, a gap at the drain end has a less marked effect on current, since there is no change in "emitter efficiency". These phenomena have been verified using computer simulation [6]. Thus, in a gapped device, swapping the drain and source terminals will cause a change in subthreshold current.

Figures 6–6 and 6–7 give subthreshold characteristics for two POT devices, one with a gap, one without. The device without a gap, shows no change in current when source and drain are reversed. Whereas, the gapped device with a shows a large difference when terminals are reversed.

The difference between forward and reverse currents needs to be quantified, so that devices in an array can be compared. For this subthreshold currents were measured, with 0.5V on the drain, 0.4V on the gate, and other terminals earthed. Measurements were then taken with the source and drain swapped to give I_1 and I_2 respectively. Equation 6.1 was used to calculate I_{diff}, which is proportional to the gap size.

$$I_{diff} = |Log_{10}(I_1) - Log_{10}(I_2)| \tag{6.1}$$

Figure 6–8 gives the results for three POT arrays.

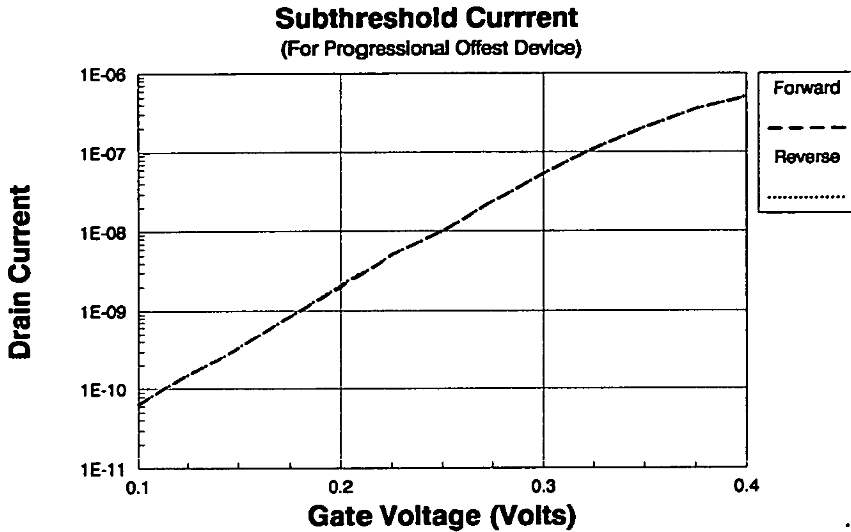

Figure 6–6: Subthreshold Characteristic for a POT Transistor $5\mu m$ Wide, Which has *No* Gap. Drain Voltage= $0.5V$, Gate Voltage is Ramped from $= 0.1V$ to $= 0.4V$, Source and Substrate Voltage= $0V$.

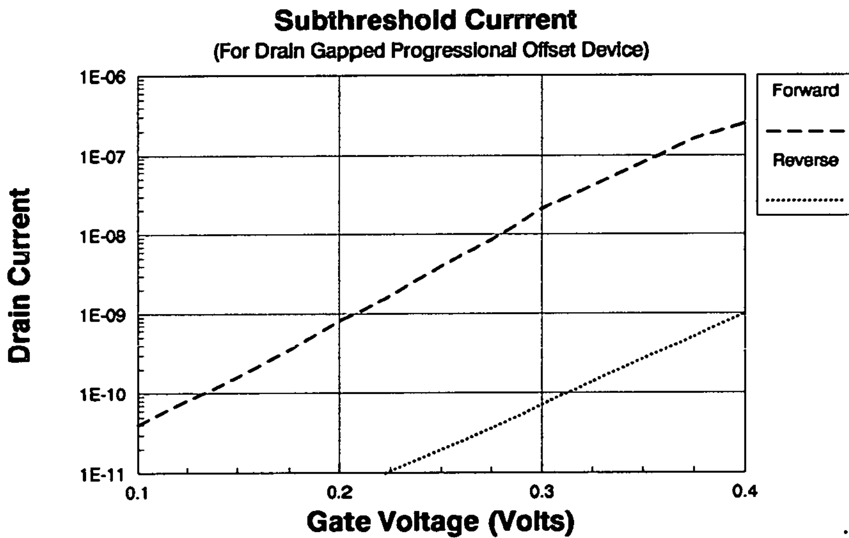

Figure 6–7: Subthreshold Characteristic for a POT Transistor $5\mu m$ Wide, Which *Has* a Gap. Drain Voltage= $0.5V$, Gate Voltage is Ramped from $= 0.1V$ to $= 0.4V$, Source and Substrate Voltage= $0V$.

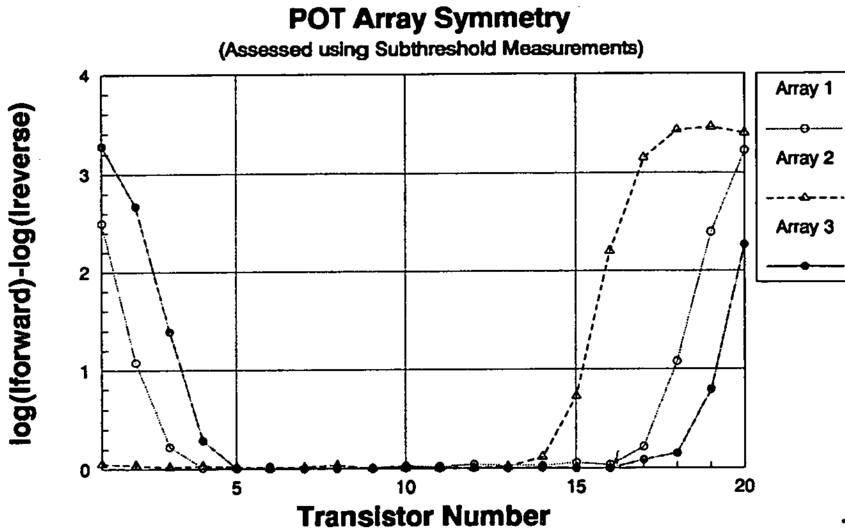

Figure 6–8: Symmetry Analysis of Three POT Arrays, Using Comparison of Subthreshold Voltage.

Array 1 is symmetrical about the $10th$ transistor, Ty_{10}. Inspection reveals that the $1st$ and $19th$ transistors, Ty_1 and Ty_{19}, have the same value of I_{diff}. Thus, one may move along the array by 9 transistors from Ty_1 or Ty_{19}, to arrive at Ty_{10}. Arrays 2 and 3 are both asymmetric. However, one may still locate the symmetrical device. It is assumed that a device with $I_{diff} > 1.7$ is 9 steps away from the symmetrical device, and one with $I_{diff} > 2.8$ is 10 steps away from the symmetrical device. An algorithm was written to analyse a POT array and return the symmetrical device, based on this set of criteria, see appendix E. This was included in a test program written for the KLA Automatic Wafer Prober and HP 4062 Semiconductor Parameter Test System. Infact, lithography is unlikely to produce an error of exactly $0.05\mu m$. Hence, there may be no exactly symmetrical device in an array. Even so, it is necessary to group transistors into discrete sets, for statistical analysis. The above criteria were used to find the closest transistor to symmetricality. Once the symmetrical device has been located, the gate/drain overlaps of all other devices can be found, knowing:

1. Gate/drain overlap of symmetrical device is designed to be $0.25\mu m$, see chapter 5.

2. Step size in the POT array is $0.05\mu m$

6.3 Reliability Analysis

6.3.1 Test Methodology

Of interest is the EEPROM programming operation. Constant voltage stressing between the drain and gate may be used to emulate this [3], with the substrate earthed and the source *floating*. This produces current flow between the gate and drain, which stresses the oxide. Current stress leads to charge trapping and eventual oxide rupture [3]. The charge to breakdown, Q_{BD}, then indicates the reliability of the device [11]. Ten volts was applied to the drain, the gate and substrate were earthed, and the source was allowed to float. This produced $8.4MVcm^{-1}$ across the $119\mathring{A}$ gate oxide.

Two types of POT transistor were designed, one $5\mu m$ wide, and one $50\mu m$. Reliability tests revealed that the $50\mu m$ wide devices had a substantial infant mortality rate. The $5\mu m$ devices were therefore chosen as the most suitable for reliability analysis. For a gate drain/overlap of $0.3\mu m$, $8.4MVcm^{-1}$ gave a current of $\simeq 0.2nA$, and time to breakdown of $\simeq$ 3 minutes. Q_{BD} was calculated by integrating gate current as a function of time. It should be noted that the channel was in depletion during testing, hence only the gate/drain region was stressed.

As discussed in chapter 5, a symmetrical POT device will have a gate/drain overlap of $\simeq 0.25\mu m$. A range of gate/drain overlaps may be evaluated by testing transistors on either side of the symmetrical device. In order to double the population of data points for statistical analysis, both the gate/drain and the gate/source regions were stressed. This was possible because the channel region was in depletion during testing, so the source and drain were isolated from one another. A program was written to perform these tests using a KLA automatic wafer prober and an HP 4062 semiconductor parameter test system, see appendix E. Figure 6–9 illustrates the test set-up.

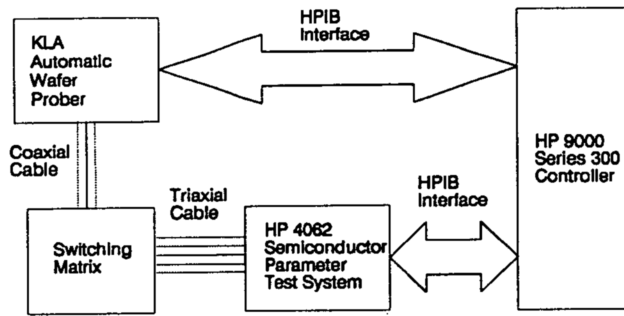

Figure 6–9: Block Scheme of Test Set-Up.

6.3.2 Reliability of Progressional Offset Transistors

In all 37 arsenic die and 83 phosphorus die were tested, for which both the gate/drain, and the source/drain regions, were stressed. In analysing results from POT arrays, it is important to appreciate that they exhibit the same features as any other population of oxide capacitors. In such a population, time to breakdown shows a lognormal distribution. This is true for both constant voltage and constant current stressing [12]. Unfortunately, because the POT transistors had different tunnel areas, time to breakdown did not offer a means of comparing reliability. Instead charge to breakdown was used, which should also show a lognormal distribution. Figures 6–10 and 6–11 gives the charge to breakdown for 20 arsenic and 20 phosphorus dies, as a function of gate/drain overlap. These display a wide range of values, indicative of such a lognormal distribution.

For each overlap value there is a percentage of infant mortality, which was defined as a device which ruptured before $t = 5$ seconds. In devices with a small overlap, the gate *only* extends over high integrity oxide. However, in devices with large overlaps of $\geq 0.5\mu m$, the gate extends over oxide which was damaged during implantation. This oxide has a reduced dielectric strength, which was unable to support the stressing field of $8.4MVcm^{-1}$. Hence, devices with an overlap $\geq 0.5\mu m$ show high infant mortality.

Figures 6–12 and 6–13 present the same Q_{BD} data on a logarithmic axis. Here, the spread in the data become(s) more uniform.

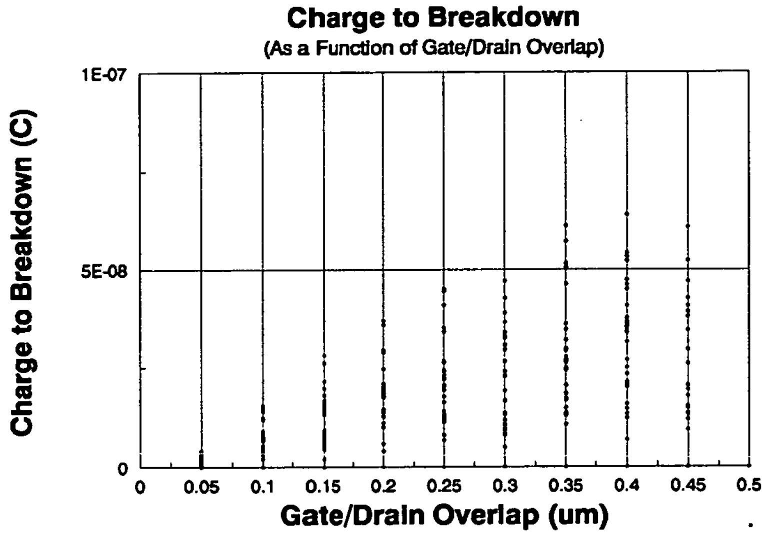

Figure 6–10: Charge to Breakdown of Arsenic Transistors, as a Function of Gate/Drain Overlap. These Gate/Drain Overlap Values Were Simulated in Chapter 5.

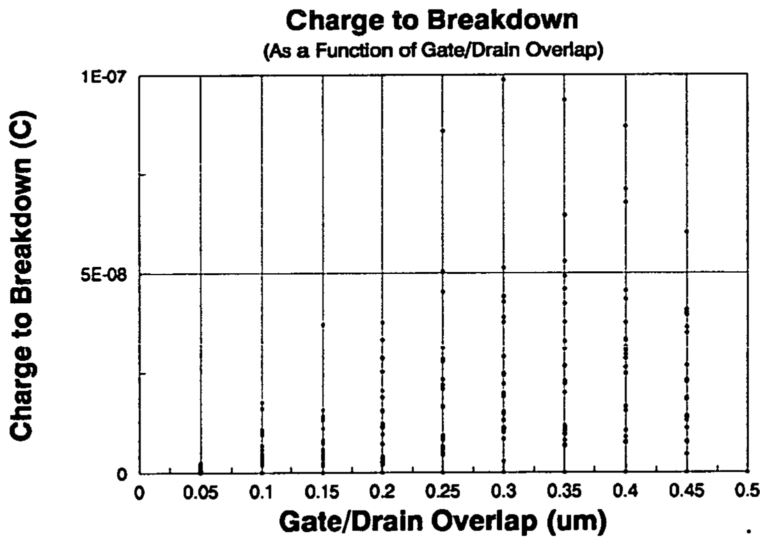

Figure 6–11: Charge to Breakdown of Phosphorus Transistors, as a Function of Gate/Drain Overlap. These Gate/Drain Overlap Values Were Simulated in Chapter 5.

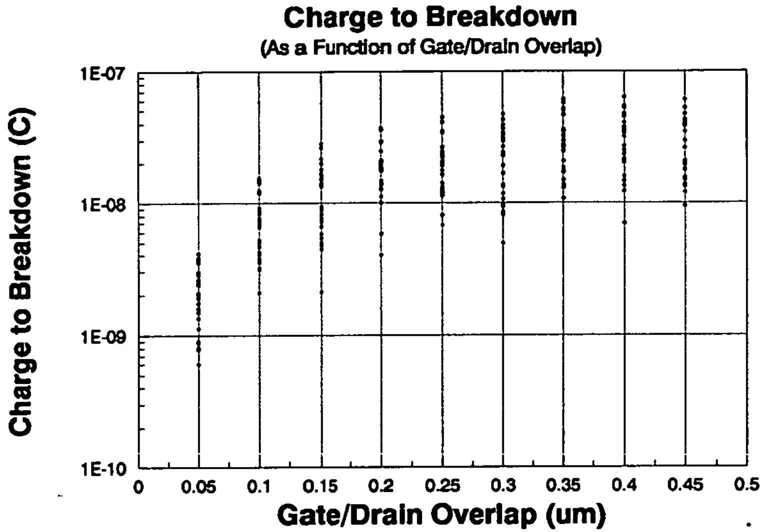

Figure 6–12: Charge to Breakdown of Arsenic Transistors, as a Function of Gate/Drain Overlap. These Gate/Drain Overlap Values Were Simulated in Chapter 5.

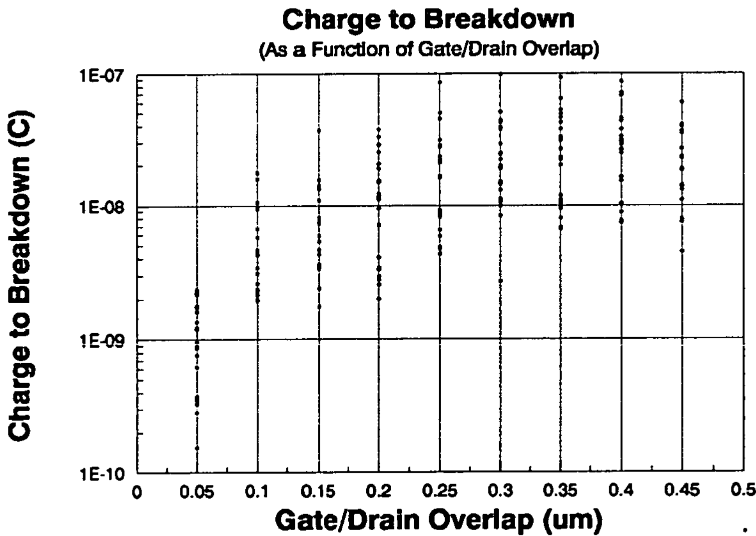

Figure 6–13: Charge to Breakdown of Phosphorus Transistors, as a Function of Gate/Drain Overlap. These Gate/Drain Overlap Values Were Simulated in Chapter 5.

To verify the lognormal nature of the results, data may be plotted on a lognormal graph [13]. Figures 6-14 and 6-15 give lognormal plots of Q_{BD}, for a gate/drain overlap $0.3\mu m$. The linearity of graphs, confirms the lognormal distribution of the results. Data for other overlap values also displayed this linearity.

6.3.3 Average Charge to Breakdown

The average value of Q_{BD} for each overlap was calculated using equation 6.2, assuming a lognormal distribution.

$$Q_{AV} = \frac{\sum_{i=1}^{n} (log_{10} Q_{BD})}{n} \tag{6.2}$$

Where:

- Q_{AV} =Average charge to breakdown.
- i=Integer.
- n=Number of data points, *excluding* infant mortalities.

Figure 6–16 illustrates average values of Q_{BD}, for arsenic and phosphorus devices, as a function of gate/drain overlap. The average values were calculated using all data points, other than infant mortalities. To these graphs, straight lines were fitted by linear regression. For arsenic devices with $0.3\mu m$ overlap, $Q_{AV} = 2.5561 \times 10^{-8} C$. Whereas, for phosphorus devices with $0.3\mu m$ overlap, $Q_{AV} = 2.6954 \times 10^{-8} C$. Now, the tunnel area is given by:

$$0.3\mu m \times 5.0\mu m = 1.5 \times 10^{-8} cm^{-2}$$

Thus the charge densities to breakdown may be calculated:

- For arsenic, charge density to breakdown $= 1.704 C cm^{-2}$.
- For phosphorus, charge density to breakdown $= 1.797 C cm^{-2}$.

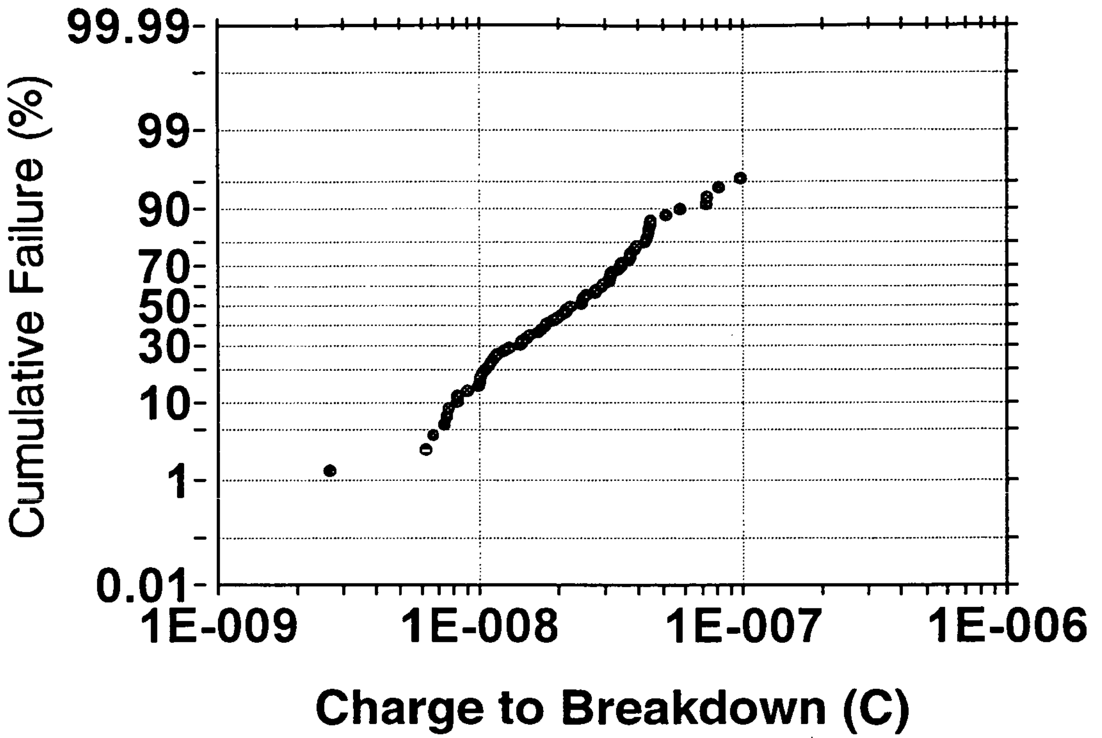

Figure 6–14: Lognormal Distribution of Q_{BD}, for Arsenic Devices.

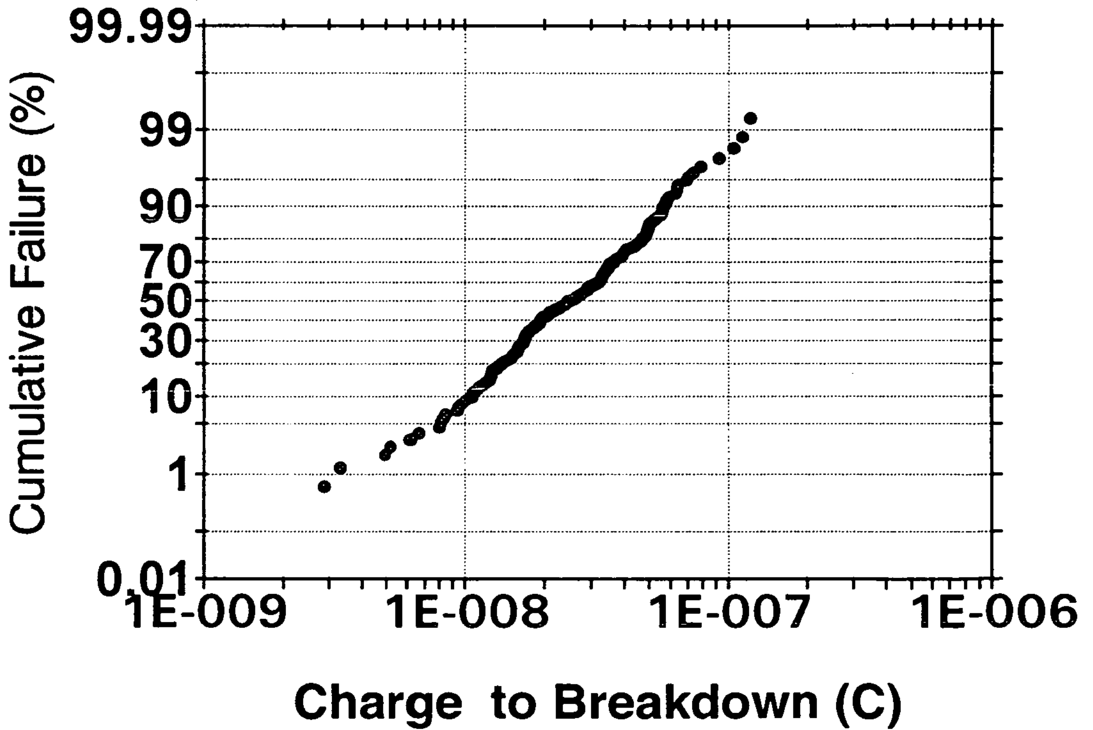

Figure 6–15: Lognormal Distribution of Q_{BD}, for Phosphorus Devices.

These values are relatively low, suggesting that the integrity was compromised during fabrication. A thin polysilicon layer was deposited directly after gate oxidation, to provide protection. Even so, it is conjectured that the oxide experienced mechanical stress during steps involving the phantom gate, see chapter 5. The thin polysilicon layer was given a 1 second etch in a 10% solution of hydrofluoric acid, prior to deposition of further polysilicon. However, it is possible that pin holes in the polysilicon film, allowed hydrofluoric acid to impair the underlying oxide. A degree of contamination may also have been residual, even after the extensive cleaning sequence given to each wafer.

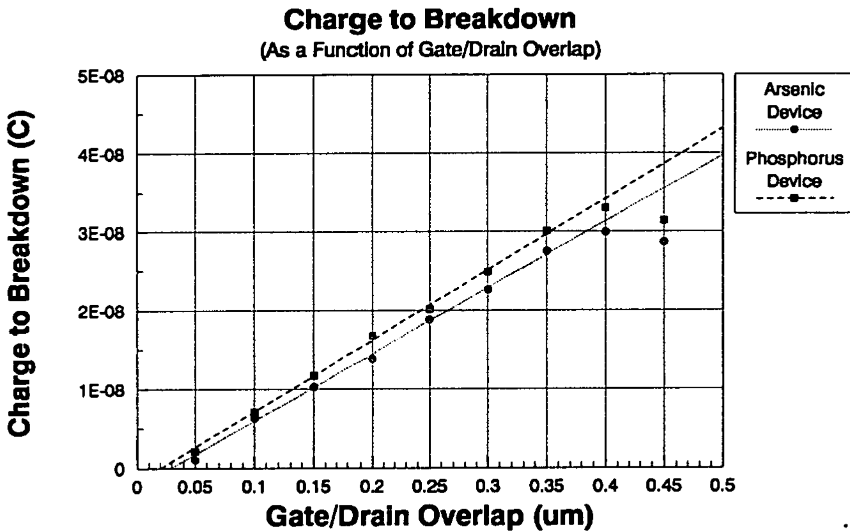

Figure 6–16: Average Q_{BD} as a Function of Gate/Drain Overlap. These Gate/Drain Overlap Values Were Simulated in Chapter 5.

6.3.4 Calculation of Lateral Diffusion

Process simulation indicated a lateral diffusion of $\simeq 0.45\mu m$. However, the experimental values may be derived from figure 6–16. As gate/drain overlap is reduced, so charge to breakdown falls linearly. In the limit, when gate/drain overlap is zero,

then charge flow is zero [1]. Strictly speaking such a device would not breakdown. However, $Q_{AV} = 0.0$ does indicate the point where gate/drain overlap is zero.
In figure 6–16 the arsenic line intercepts the horizontal axis at $0.03\mu m$, hence:
Lateral diffusion of arsenic devices $= 0.45 - 0.03 = 0.42\mu m$.
In figure 6–16 the phosphorus line intercepts the horizontal axis at $0.02\mu m$, hence:
Lateral diffusion of phosphorus devices $= 0.45 - 0.02 = 0.43\mu m$.
Figure 6–17 illustrates Q_{AV}, as a function of these derived overlap values.

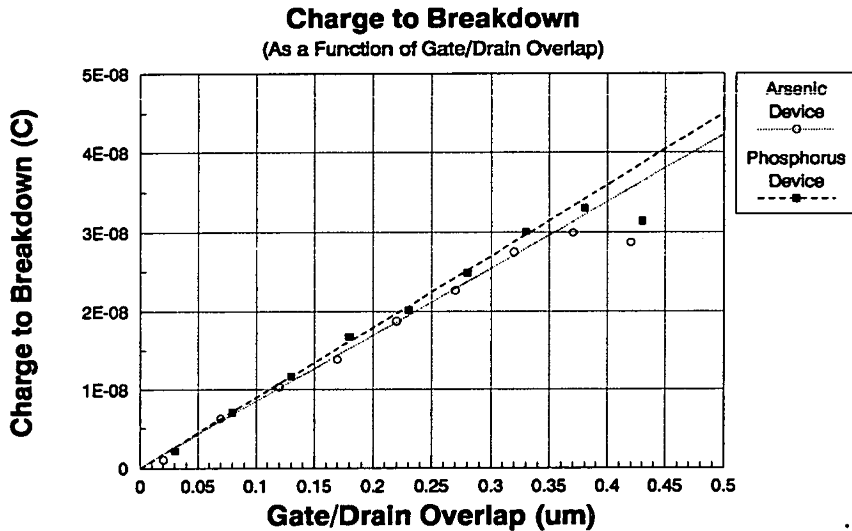

Figure 6–17: Average Q_{AV} as a Function of Derived Gate/Drain Overlap.

The arsenic devices here, are less reliable than phosphorus ones. For an overlap of $0.3\mu m$, arsenic devices have $\simeq$ 7% lower Q_{AV}, than phosphorus devices. However, this is believed to be due to the higher thermal budget experienced by the arsenic batch in processing. High thermal budgets are known to lower oxide integrity [14].

The results were not perfectly linear, but showed a dip in Q_{AV}, at large overlaps. During photolithography a mask will be aligned to the previous layer, to an accuracy of within $0.2\mu m$. In a non-aligned process, some error is therefore added

[1] Infact there is a very small charge flow, but this is dwarfed by the 3pA noise in the system.

to the value of the gate/drain overlap [6]. The data has been organised into a number of discrete sets, for $0.02\mu m$ overlap, $0.07\mu m$ overlap ...etc. However, the $0.42\mu m$ set will include data from transistors with overlaps in the range $0.395\mu m$ to $0.445\mu m$. Thus, some arsenic devices with derived overlaps of $0.42\mu m$ (and phosphorus devices of $0.43\mu m$) will have gates which overlap implant damaged oxide. It is tunnelling though this oxide which increases infant mortality, and reduces Q_{AV}.

6.3.5 Prediction of EEPROM Endurance from Experimental Results

Programming endurance will be considered. This is the major reliability issue for the FETMOS, since current densities and electric fields are highest, as described in chapter 4. A similar methodology will be used to investigate endurance, as was used in chapter 4. Thus program endurance of the FETMOS may be described by equation 6.3:

$$Q_f = Q_p \times N_{cycles} \tag{6.3}$$

Where:

- N_{cycles} = The number of program/erase cycles the FETMOS can withstand, before the programmed threshold window closes. This defines the program endurance of the device and is typically $\sim 10^5$ cycles [15].

- Q_p = The charge which passes through the floating gate/drain overlap region, during each program operation. An average midlife value was calculated in chapter 4, this was $\simeq 5.6 \times 10^{-13}C$.

- $Q_f = 5.6 \times 10^{-13}C \times 10^5 = 5.6 \times 10^{-8}C$
  This is the total charge to pass through the floating gate/drain overlap region, during programming, before failure. This value depends upon oxide integrity, and a typical value has been given.

The POT transistors gave Q_{BD} as a function of overlap. For endurance analysis, Q_{BD} will be assumed to be equivalent to Q_f. Some approximations would be needed to calculate the absolute endurance, since:

- FETMOS programming is a dynamic operation. At the end of each operation the trapped charge relaxes, and some will be "detrapped" [16]. Thus, degradation would occur more slowly under dynamic stress [16].

- Stressing in the FETMOS is bi-directional, since the field is reversed during erasing. This has been found to further enhance detrapping [16].

- EEPROMs are programmed at a higher field strength than were used in stressing POT transistors. Higher fields would be expected accelerate the degradation rate [11].

- Trap up is the principal failure mode in the FETMOS, whereas Q_{BD} gives the charge to dielectric breakdown. However, related charge trapping mechanisms are responsible for each failure mode [3].

Infact, absolute values for the degradation rate need not be calculated. It is the relative improvement in endurance which is of interest. It is seen in equation 6.3 that Q_f is *directly* proportional to endurance, N_{cycles}. Thus, if Q_f doubles N_{cycles} must double, to maintain the equality of equation 6.3. Percentage variations in endurance will be calculated using equation 6.4:

$$R_{endurance} = \left(\frac{Q_{BD}}{Q_{BD(for\ 0.3\mu m)}} \right) \times 100 \tag{6.4}$$

Where:

- $R_{endurance}$ = Relative endurance of a FETMOS, compared to one with the standard floating gate/drain overlap, of $0.3\mu m$.

- $Q_{BD(for\ 0.3\mu m)}$ = Charge to breakdown for a POT transistor, with a gate/drain overlap of $0.3\mu m$.

- Q_{BD} = Charge to breakdown for a POT transistor, in which the gate/drain overlap is variable.

At an overlap of $0.3\mu m$, $R_{endurance}$ = 100%, this is the endurance of a standard FETMOS. At an overlap of $0.4\mu m$, $R_{endurance}$ = 130%, so the endurance has increased by 30%. Figure 6–18 gives endurance as a function of overlap, calculated from arsenic and phosphorus POT devices.

In chapter 4, the charge density passing through the EEPROM during programming was modelled. Endurance was said to be the reciprocal of this. Hence, modelled results for endurance can also be included in figure 6–18. These are in agreement with experimental results, proving overall consistency within the thesis.

6.4 Conclusion

An experiment has been conducted to measure EEPROM reliability, as a function of floating gate/drain overlap and doping species. It has been seen that, doping species has little effect on reliability, but that overlap has an important role to play. Endurance was shown to be proportional to floating gate/drain overlap.

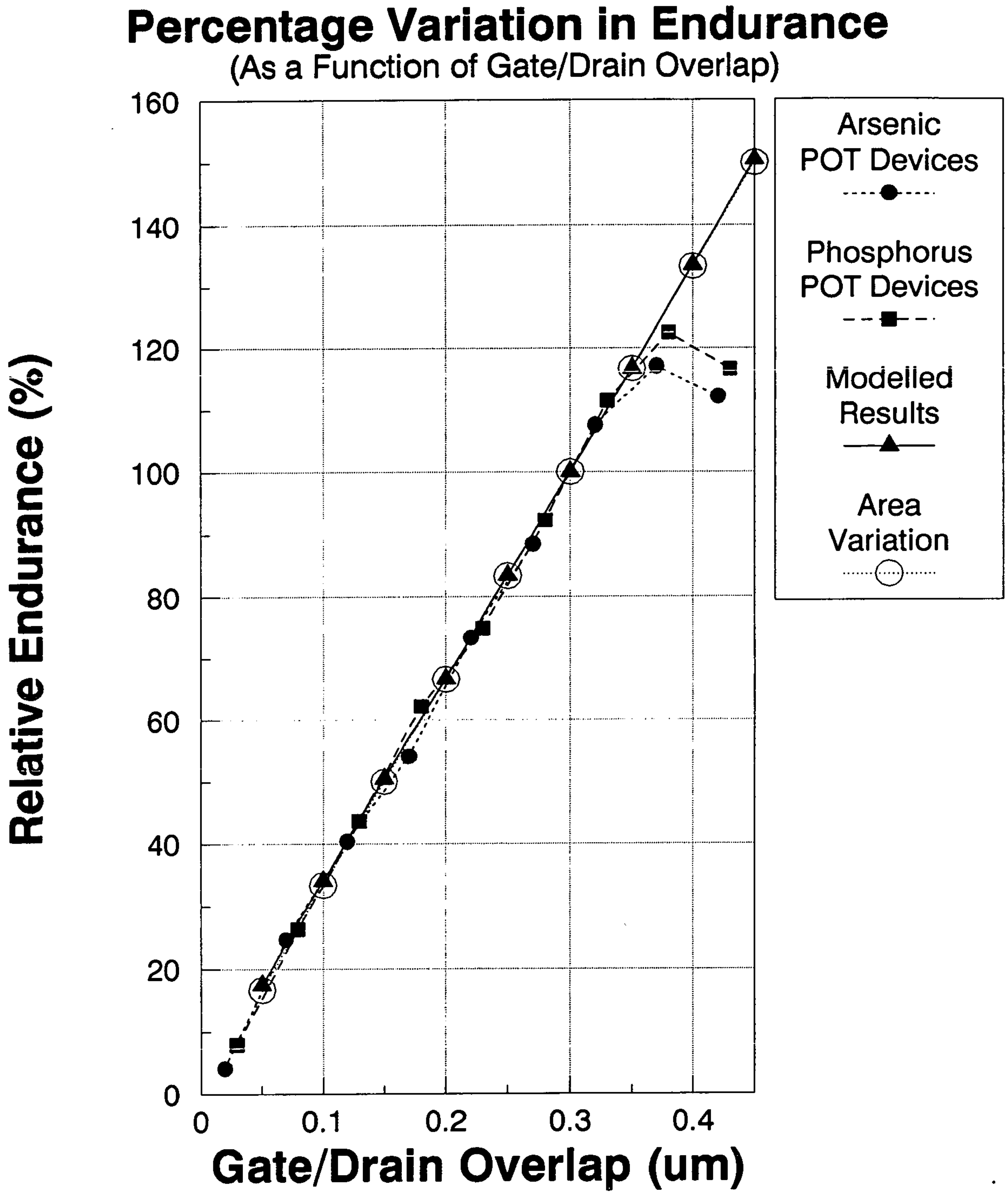

Figure 6–18: Relative Endurance of a FETMOS, Compared to one with the Standard Floating Gate/Drain Overlap of $0.3\mu m$.

Bibliography

[1] R.D.Pashley and S.K.Lai. Flash memories: The best of two worlds. *IEEE Spectrum*, 26(12):30–33, December 1989.

[2] S.M.Sze, editor. *VLSI Technology*, chapter 5. McGraw-Hill International Editions, 1988.

[3] Ih-C.Chen, S.E.Holland, and C.Hu. Electrical breakdown in thin gate and tunneling oxides. *IEEE Transactions On Electron Devices*, 32(2):413–422, February 1985.

[4] A.J.Walton and A.Gribben. A review of parametric testing. In *SEMICON Birmingham*, pages 39–63, 1987.

[5] R.C.Smith and A.J.Walton. The design, fabrication and measurement of assymetrical LDD transistors. 1992.

[6] J.Serack. *Edge Effects In Silicon IGFETs*. PhD thesis, University Of Edinbugh, 1988.

[7] H.Neves. *Hot Carrier Effects In IGFETs*. PhD thesis, University Of Edinbugh, 1991.

[8] S.M.Sze. *Physics of Semiconductor Devices*, chapter 8. John Wiley and Sons, 1981.

[9] Y.P.Tsividis. *Operation and Modelling of the MOS Transistor*, chapter 4. McGraw-Hill International Editions, 1988.

[10] A.Bar-Lev. *Semiconductors and Electronic Devices 2nd Edition*, chapter 12. Prentice/Hall international, 1984.

[11] J.C.Lee, I-C.Chen, and C.Hu. Modelling and characterisation of gate oxide reliability. *IEEE Transactions On Electron Devices*, 35(12):2268–2277, 1988.

[12] C-F.Chen, C-Y.Wu, and M-K.Lee. The dielectric reliability of intrinsic thin SiO_2 films thermally grown on a heavily doped si substrate- characterization and modeling. *IEEE Transactions On Electron Devices*, 34(7):1540–1552, 1987.

[13] S.M.Sze, editor. *VLSI Technology*, chapter 14. McGraw-Hill International Editions, 1988.

[14] S.Holland and C.Hu. Correlation between breakdown and process-induced positive charge trapping in thin thermal SiO_2. *J.Electrochem.soc Solid State Science And Technology*, 133(8):1705–1712, August 1986.

[15] C.Kuo, Y.R.Yeargain, and W.J.Downey. An 80ns 32K EEPROM using the FETMOS cell. *IEEE J.Solid State Circuits*, (5):821–827, October 1982.

[16] M-S.Liang, S.Haddad, W.Cox, and S.Cagnina. Degradation of very thin gate oxide MOS devices under dynamic high field/current stress. In *IEDM*, pages 394–398, 1986.

Chapter 7

Conclusion

Reliability problems are a key concern in EEPROM design, due to the extreme voltage and current stressing conditions under which they operate. These problems also limit the operating speed of the EEPROM, since an increase in speed reduces the reliability. In this thesis reliability has been investigated both experimentally, and through modelling. Thus, a consistent picture of EEPROM reliability has emerged.

It has been seen that, an increase in the floating gate/drain overlap will improve reliability substantially. Modelling has also shown that the threshold window remains stable as floating gate/drain overlap is increased. Increasing the tilt angle during drain implantation, would be the most suitable method for improving overlap, since the thermal budget of the process may then be conserved. An increase in the doping density could also be used, to similar effect. Chemistry has not been seen to play a major role in EEPROM reliability, although phosphorus should be used, to extend floating gate/drain overlap. These results will now be expanded upon.

7.1 Modelling

7.1.1 Discussion of Results

A new model for the FETMOS has been developed, and has been verified against experimental measurements. With this it is possible to predict the variations in threshold window, caused by a change in processing conditions. Until now such variations had been investigated using split-lots, in which wafers were fabricated using different processing parameters. These wafers were destined for sale, so the variations in process parameters had to be finely judged. The model will allow the most promising process parameters to be varied, and will ensure that the split-lots *are* saleable.

A new methodology has also been developed to model EEPROM endurance. This is elegant, has a sound footing on established principles, and has not been used before, to the authors knowledge. Here again, results of endurance modelling have been verified against experimental data. Program reliability is of principal concern, since programming fields and current densities are highest. It has been shown that a 30% improvement in reliability can be expected, for a $0.1\mu m$ increase in floating gate/drain overlap. This is significant, since modelling has also shown that the threshold window remains open, as overlap increases. Thus, larger overlap has a wholly beneficial effect.

7.1.2 Areas for Further Modelling

Modelling FETMOS Program/Erase Time

The model developed in chapter 3 includes an RC time constant, which enables dynamic program/erase operations to be investigated. Hence, the model could be used to investigate program/erase time, as a function of parameter variations. Instead of calculating the threshold window after $10ms$, the time required for the program/erase threshold voltages to reach $\pm 5V$, would be investigated. Parameter

variations which increase speed could be found, while the endurance was monitored to ensure devices remained sufficiently reliable. This approach is complimentary to that adopted in chapter 4, and would help provide a fuller picture of FETMOS operation.

Modelling of Flash EEPROMs

Given the growing popularity of flash EEPROMs, and their low endurance [1], modelling flash EEPROM reliability would be a useful area of research.

All flash EEPROMs use Fowler-Nordheim tunnelling for programming [1], where electrons are removed from the floating gate [1]. In this thesis a model has been developed to investigate EEPROM reliability, which includes equations to describe Fowler-Nordheim tunnelling. This model may be applied directly to many flash designs. Simple flash structures, such as the Intel and Seeq devices, are equivalent to the the FETMOS, and could be described by a similar capacitive network. However, more complex structures, such as the Toshiba device, use three gates [1]. Care would be needed in the analysis of this capacitor network, before it could be reduced to the form used in the model.

The majority of flash EEPROMs use channel hot electron injection, for erasing. The hot electron current density J_{he}, during erase, can be written as [2]:

$$J_{he} = -q \int_{\epsilon_B}^{+\infty} v_{\perp}(\epsilon) f(\epsilon) g(\epsilon) d\epsilon \tag{7.1}$$

where:

- q = the charge on an electron.
- ϵ = the electron energy.

---

[1] For consistency within this thesis, program describes the condition with electrons removed from the floating gate. However, many flash EEPROM manufacturers consider program to denote stored electrons.

- ϵ_B = the height of the Si/SiO_2 barrier.

- $v_\perp(\epsilon)$ = the energy dependent electron velocity normal to the interface.

- $f(\epsilon)$ = the electron energy distribution, which is non-Maxwellian in the high electric field regime.

- $g(\epsilon)$ = the density of allowable electron states.

Two dimensional device simulators such as TMA MEDICI and HFIELDS [2] may be used solve these equations, in a finite mesh analysis. MEDICI allows the inclusion of a floating gate in the structure, and the extraction of all relevant data, such as threshold voltage and charge on the floating gate. A transient analysis could be made, after which the charge density passing though the oxide could be extracted. Thus a reliability analysis could be conducted using the methodology developed in chapter 4. Finite mesh analysis is is costly in terms of computing time, and care would needed to collect results in an efficient manner.

Modelling at Circuit Level

EEPROMs are being included in many nascent technologies, such as artificial intelligence, self adaptive systems and neural networks [3] [4]. In these applications the EEPROM is no longer confined to a regimented memory array, but plays a more active role in circuit operation. In such a system, EEPROM reliability become difficult to quantify, since the EEPROM operating environment is more complex. Thus, reliability should not only be simulated at device level, but also at circuit level.

The Berkeley Reliability Tool (BERT) contains models for reliability phenomena such as electromigration and hot electron degradation [5]. This is linked to SPICE [6], and may be used to locate reliability hot spots in a new circuit. The model for time dependent dielectric is based on DC stress data [7]. Thus, it would not necessarily be applicable to the EEPROM, for which dynamic, high

field, trapping/detrapping mechanisms would be important [8] [9]. It would therefore be interesting, to include the EEPROM reliability methodology developed in chapter 4, within BERT.

This would be a particularly useful aid in the design of neural networks. Here analogue weights are stored on the EEPROM. However, the threshold window of a device narrows with cycling, which necessitates the use of a feedback based programming scheme [4]. BERT could thus be used to provide a better understanding of these circuits.

7.2 Experimental

7.2.1 Discussion of Results

A new methodology has been developed for investigating EEPROM reliability, using the progressional offset technique. With this, an array of transistors was produced, which had a spectrum of gate drain overlaps. It was shown that reliability of the EEPROM is directly proportional to the floating gate/drain overlap, this result was consistent with model predictions.

Degradation phenomena are believed to take place in the bulk of the oxide [10] [9], as opposed to the interfaces. These phenomena are not well understood, and the effect of chemical species, which diffused into the oxide from the gate and drain, was investigated. Infact, chemistry was seen to play only a minor role in reliability, since arsenic devices were only $\simeq 7\%$ less reliable than phosphorus ones. It is likely that this reduction in reliability was caused by the higher temperatures, experienced by arsenic devices during fabrication [11].

7.2.2 Areas for Further Experimental Investigation

Methods for Increasing Floating Gate/Drain Overlap

The two most promising methods for increasing floating gate/drain overlap, would be an increase in the drain implant angle and an increase in the implant dose. T-MA TSUPREM-4 allows two dimensional process simulation, and could be used to carry out an extensive investigation of these parameters. Suitable combinations of implant angle and doping density, could then be tested experimentally. Commercial EEPROM structures would offer a good test ground for such an experiment. Definitive values for the improvement in reliability, could thus be obtained.

Improvement in Dielectric Integrity

Despite extreme care during fabrication, the oxides produced for the POT experiment had a relatively low integrity. This highlights the concern for improvement of dielectric reliability. Moves to improve dielectric integrity are being made in several directions. The inclusion of nitrogen in oxide films is an option which currently receiving much attention [12] [13] [14]. Nitrided oxides are generally formed by exposing an oxide thin film to an NO_2 ambient, either in a furnace environment at $\sim 900^oC$ [13], or in a rapid thermal processing (RTP) chamber at $\sim 1050^oC$ [15]. Furnace nitridation is similar to the long-time postanneal described in chapter 5 [16], in that each process uses a reoxidation step to remove charge trapping sites [14]. Nitrided oxides exhibit improved hot-carrier reliability, lower charge trapping rates and higher charge to breakdown [12]. Such qualities would be of benefit to EEPROM technologies [15], and to any future POT experiment.

Preoxidation cleaning can also effect oxide induced stacking faults and dielectric properties [17]. It has been seen that incorporation of fluorine into the S_i/S_iO_2 interface, dramatically improves the oxide's resistance to hot-electron damage and its dielectric strength [18] [19]. Fluorine may be introduced by immersion in an solution of hydrofluoric acid, prior to oxidation [19], or by ion implantation. This may provide another path for dielectric improvement.

Three Dimensional EEPROM Structures

So far, EEPROM designers have only thought in two dimensions. This is in contrast to the more mature DRAM technologies, which use three dimensional trenches to form storage capacitors [20]. The author believes that a three-dimensional EEPROM structure would have very significant advantages. Such an EEPROM would have a high floating gate, rising in one or more vertically ridges above the channel region of the device. A large capacitance could then be derived from the vertical walls of the floating gate. Care would be needed, to ensure that capacitive coupling between control gates of different EEPROM, was minimised. Plasma enhanced chemical vapour deposition could then be used to planarise the surface, while maintaining a low thermal budget [21] [22]. The advantages would be two fold:

1. Integration density would be increased, due to the reduced cell area.

2. A three dimensional structure could be used to increase coupling ratios. High oxide fields could then be generated across tunnel oxides, using lower operating voltages. Thus, EEPROMs could be embedded more easily into logic circuitry.

A number of other avenues have been investigated for DRAM improvement, but could be applied to the EEPROM. The floating gate capacitance may also be increased, by the inclusion of a film with a high dielectric constant, such as TaO_x or Si_3N_4, in a sandwich with the interlevel oxide [23]. Another possibility would be to "roughen" the top surface of the floating gate, thus increasing its effective area [23]. Photoresist particles would be used as masks, during plasma etching of the surface. The deposition of silicon in hemispherical grains, in a layer covering the floating gate, is another possibility. This is a low temperature process producing hemispherical grains with a diameter $\sim 0.1\mu m$ [23]. This has provided a capacitance increase of $\sim 30\%$ in tests using the DRAM [23].

7.3 Concluding Remarks

This thesis has produced many interesting results, some of which are currently the subject of a patent application with Motorola. A paper has been accepted for presentation at the IEEE sponsored International Conference on Microelectronic Test Structures, to be held in California, 1994. A summary of the paper is provided in appendix A. Finally, it is planned to continue this research in association with Motorola.

Bibliography

[1] R.D.Pashley and S.K.Lai. Flash memories: The best of two worlds. *IEEE Spectrum*, 26(12):30–33, December 1989.

[2] S.Keeney, R.Bez, and D.Cantarelli. Complete transient simulation of flash eeprom devices. *IEEE Transactions On Electron Devices*, 39(12):2750–2757, 1992.

[3] H.E.Meas, J.Witters, and G.Groeseneken. Trends in non-volatile memory devices and technologies. In *ESSDERC Bologna*, pages 743–754, 1987.

[4] C-K.Sin, A.Kramer, and V.Hu. EEPROM as an analog storage device, with particular application in neural networks. *IEEE Transactions On Electron Devices*, 39(6):1410–1419, 1992.

[5] C.Hu. IC reliability simulation. *IEEE Journal Of Solid State Circuits*, 27(3):241–246, March 1992.

[6] A.J.Walton and A.Gribben. A review of parametric testing. In *SEMICON Birmingham*, pages 39–63, 1987.

[7] J.C.Lee, I-C.Chen, and C.Hu. Modelling and characterisation of gate oxide reliability. *IEEE Transactions On Electron Devices*, 35(12):2268–2277, 1988.

[8] M-S.Liang, S.Haddad, W.Cox, and S.Cagnina. Degradation of very thin gate oxide MOS devices under dynamic high field/current stress. In *IEDM*, pages 394–398, 1986.

[9] S.Haddad and M-S.Liang. The nature of charge trapping responsible for thin-oxide breakdown under dynamic field stress. *IEEE Electron Device Letters*, 8(11):524–527, 1987.

[10] Ih-C.Chen, S.E.Holland, and C.Hu. Electrical breakdown in thin gate and tunneling oxides. *IEEE Transactions On Electron Devices*, 32(2):413–422, February 1985.

[11] S.Holland and C.Hu. Correlation between breakdown and process-induced positive charge trapping in thin thermal SiO_2. *J.Electrochem.soc Solid State Science And Technology*, 133(8):1705–1712, August 1986.

[12] G.W.Yoon, A.B.Joshi, J.Kim, and D-L.Kwong. MOS characteristics of NH_3 nitrided N_2O-grown oxides.

[13] Z.Liu, H-j.Wan, and P.K.Ko. Improvement of charge trapping characteristics of N_2O-annealed and reoxidised N_2O-annealed thin oxides. *IEEE Electron Device Letters*, 13(10):519–521, 1992.

[14] B.S.Doyle. p-channel hot-carrier optimisation of RNO gate dielectrics through the reoxidation step. *Electron Device Letters*, 14(4):161–163, 1993.

[15] G.W.Yoon, A.B.Joshi, J.Kim, and D-L.Kwong. High-field-induced leakage in ultrathin N_2O oxides. *Electron Device Letters*, 14(5):231–233, May 1993.

[16] S.S.Cohen. Electrical properties of post-annealed thin SiO_2 films. *J.Electrochem.soc Solid State Science And Technology*, 130(4):929–932, April 1983.

[17] A.Hariri. Evaluate wafer cleaning effectiveness. *Semiconductor International*, pages 74–78, 1989.

[18] L.Vishnubhotla, T-P.Ma, H.H.Tseng, and P.J.Tobin. Effects of avalanche hole injection in fluorinated S_iO_2 MOS capacitors. *Electron Device Letters*, 14(4):196–198, 1993.

[19] J.L.Prom, J.Castagne, G.Sarrabayrouse, and A.Munoz-Yague. Influence of preoxidation cleaning on the electrical properties of thin SiO_2 layers. *IEE Proceedings.*

[20] S.M.Sze, editor. *VLSI Technology*, chapter 11. McGraw-Hill International Editions, 1988.

[21] P.N.Kember. Plasma deposition of insulators for semiconductor applications. *Plasma Technology News*, (4), 1990.

[22] Article Enquiry No. 004. ECR planarization.

[23] H.Watanabe, A.Sakai, T.Tatsumi, and T.Niio. Hemispherical grain silicon for high density DRAMs. *Solid State Technology*, 35(7):29–33, 1992.

Appendix A

Summary of Paper

A summary of the paper accepted for presentation, at the 1994 IEEE International Conference on Microelectronic Test Structures, is given below.

Experimental Investigation of EEPROM Reliability Issues

A. J. Chester and A. J. Walton,
Edinburgh Microfabrication Facility,
Department of Electrical Engineering,
University of Edinburgh,
Edinburgh, EH9 3JL, UK.

P. Tuohy,
Motorola Ltd.,
MOS Memory and Microprocessor Division,
Kelvin Industrial Estate,
East Kilbride, G75 OTG, UK.

Abstract

A set of novel EEPROM test structures have been designed and fabricated, in which gate/drain overlap is incremented in a number of well defined steps. These

have been used to emulate EEPROM programming conditions, and measure endurance. The structures have enabled EEPROM endurance to be investigated as a function of drain doping species (As and P), for a spectrum of floating gate/drain overlaps.

Introduction

Demand for Electrically Erasable Programmable Read Only Memories (EEPROM-s) is growing inexorably, in applications such as lap-top PCs and microcontrollers. These devices work under stressful operating conditions, which lead to current induced failure modes, and limits their useful life. Customers normally require a fast memory. However, the price paid for increased speed, is reduced reliability. Reliability is therefore a vital consideration, when introducing EEPROM technology to a system. In this paper a test structure has been designed to evaluate EEPROM reliability, as a function of gate/drain overlap and doping species (As or P).

The Floating Gate Electron Tunnelling MOS (FETMOS) used by Motorola, has been chosen for study. It has been observed that FETMOS devices fabricated with a phosphorus drain, are more reliable than equivalent arsenic devices. It is therefore of great interest to clarify whether this results from the differing programming areas, or whether it relates directly to the chemistry. To identify the important reliability factors, a set of test devices with both arsenic and phosphorus drains have been fabricated. These each have the same degree of gate/drain overlap, so that meaningful comparisons can be made.

Design and Fabrication of EEPROM Test Structures

MOS transistors with a thin gate oxide, can be used as test structures for comparing the reliability of EEPROMs, with arsenic and phosphorus drains. A column of test transistors, in which the gate/drain overlap is incremented, can be produced using the Progressional Offset (POT) Technique, as shown in figure A–1. This uses a non-aligned process, in which the source-drain regions, and the polysilicon gate, are defined at separate steps. Each POT column contains 20 devices, with

$0.05\mu m$ difference between the position of each gate. Within a column, one of the devices will be symmetrical, while those about it will be skewed to the left or right. The location of the symmetrical device may be determined electrically. Although the batches with phosphorus and arsenic drains use different thermal budgets, the POT technique allows performance comparisons to be made for the two different species, since devices with equal degrees of overlap are available.

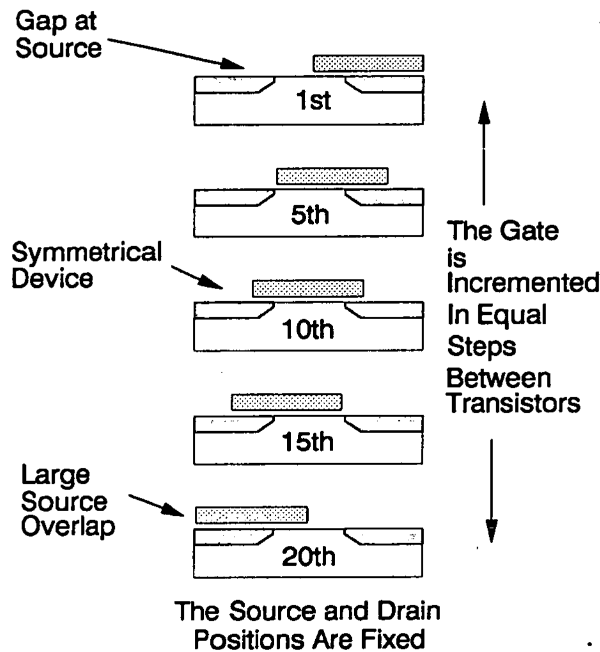

Figure A–1: Schematic Diagram Illustrating a Column of Progressional Offset Transistors.

Detecting Symmetry of a POT Column

The symmetry of a device is detected by comparing its electrical characteristics in the forward and reverse directions, where a reverse bias device has its source and drain swapped. Both current drive and subthreshold voltage have been used to conduct such tests. The subthreshold test has been well characterised, and was used for this study.

Endurance of POT Column

Ten volts was applied to the drain, with the gate and substrate grounded, and the source floating. This produces a field of $8.4MVcm^{-1}$ between the gate and drain, which causes Fowler-Nordheim tunnelling through the oxide. This measurement was performed using a Hewlett-Packard 4062 Semiconductor Parametric Test System and a KLA automatic probe station, with gate current integrated as a function of time, to obtain charge to breakdown, Q_{BD}. Figure A–2 gives the results obtained for Q_{BD}, as a function of gate/drain overlap. Figure A–3 shows the average Q_{BD} (assuming a lognormal distribution) for phosphorus and arsenic devices, as a function of gate/drain overlap. In each figure, the gate drain overlap values are those predicted by process simulation, using TMA TSUPREM4.

There is some uncertainty over the overlap predicted by process simulation. However the experimental gate/drain overlap may be calculated from figure A–3. As gate/drain overlap is reduced, so charge to breakdown falls linearly. In the limit, when gate/drain overlap is zero, then charge flow is zero [1]. Strictly speaking such a device would not breakdown. However, $Q_{AV} = 0.0$ does indicate the point where gate/drain overlap is zero.
In figure A–3 the arsenic line intercepts the horizontal axis at $0.03\mu m$, hence:
Lateral diffusion of arsenic devices = $0.45 - 0.03 = 0.42\mu m$.
The phosphorus line intercepts the horizontal axis at $0.02\mu m$, hence:
Lateral diffusion of phosphorus devices = $0.45 - 0.02 = 0.43\mu m$.
Figure A–4 illustrates Q_{AV}, as a function of the derived overlap values.

Conclusion

It is seen that the reliability of phosphorus and arsenic devices are comparable. Chemistry appears to have a less significant role in the reliability of the devices, than

[1]Infact there is a very small charge flow, but this is dwarfed by the 3pA noise in the system.

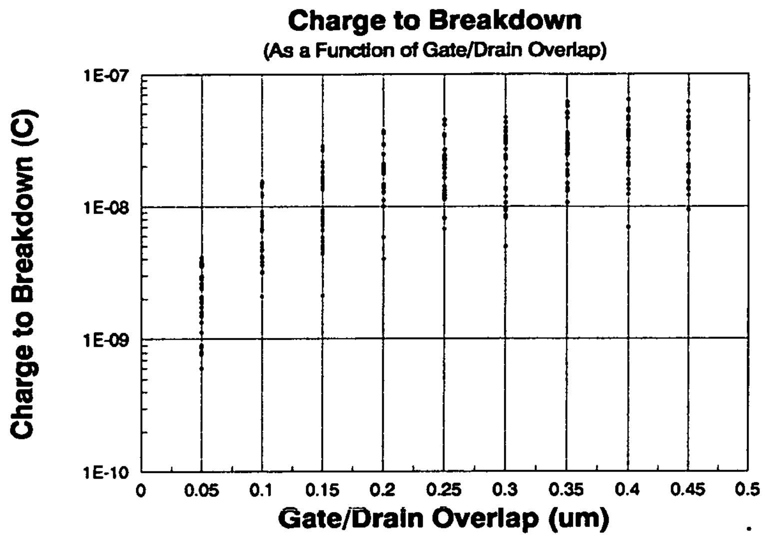

Figure A–2: Charge to Breakdown of Arsenic Transistors, as a Function of Gate/Drain Overlap.

the effect of gate/drain overlap. It is conjectured that the reliability of arsenic devices may be reduced, due to the increased thermal budget they experienced, since this is known to have a detrimental effect on reliability.

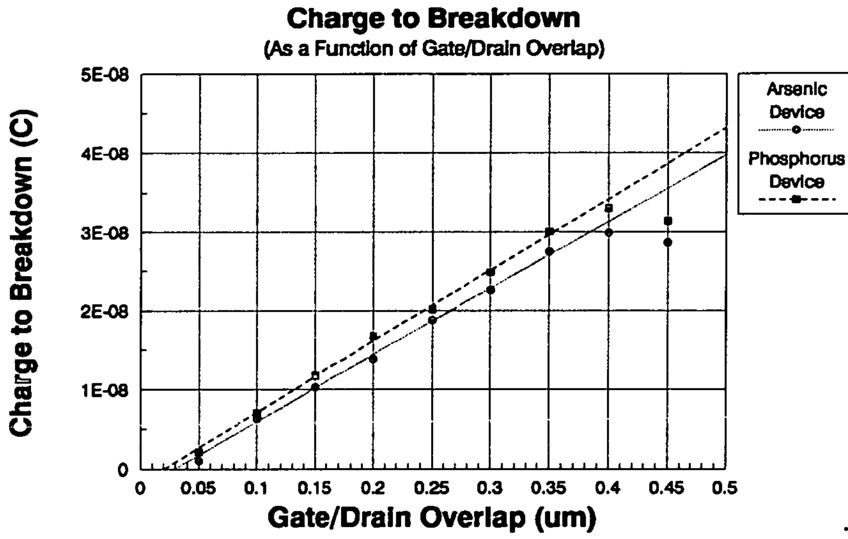

Figure A–3: Average Charge to Breakdown, as a Function of Gate/Drain Overlap.

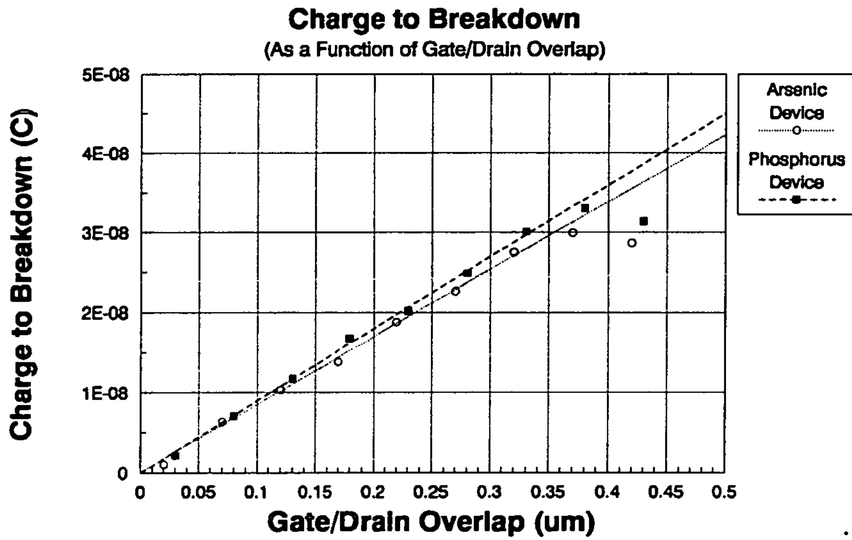

Figure A–4: Average Q_{BD} as a Function of Derived Gate/Drain Overlap.

Appendix B

Program in C to Model the EEPROM

The following is a program in C. This uses the Runge-Kutta algorithm to solve differential equations describing the EEPROM.

```
#include <math.h>
/* ew.c
   SOLVES DIFFERENTIAL EQUATION USING 4th
   ORDER RUNGE-KUTTA METHOD  For Program/Erase  */
double FNDiff_equatione(double X,double Y);
double FNDiff_equationp(double X,double Y);
/*............................................................ */
int     i=1;
int     j=1;
/*............................................................ */
double Nos=1000;
double T0=1.0E-6;
double Tfin=1.0e-2;
double Tau=1.0E-4;
/*............................................................ */
double Eee=3.9;
double Eoo=8.85E-12;
/*............................................................ */
double Weff,Leff,Lg,Latdif,Od,Aers,Bers,Aprog,Bprog,Xo,Xint,Depl;
double Ei,Pp,Pe,Rp,Re,Afg,Cfc,Cfd,Cfs,Cfg,Ct;
double Jintp,Jprog,Vtnat,Vpe,Vtp,Vte,Vtei,Vtpi;
double Tr,Er,Olderror,Newerror,Expttau,Vtei1,Vtpi1;
double Eepeak,Eppeak,Qde,Qdp,Expbee,Expbep,EB,EBprg,Vary;
double Eepeakcent,Eppeakcent,Qdecent,Qdpcent,Vtecent,Vtpcent;
/* ---------------------------------------------------------- */

FILE *data;
main(argc,argv)
int argc;
char *argv[];
{
        if ((data=fopen(argv[1],"r"))==NULL)
        {
                printf("fopen failed\n");
                exit(0);
        }
fscanf (data,"%lf %lf %lf %lf %lf %lf %lf %lf %lf %lf %lf %lf ",
&Aers,&Bers,&Aprog,&Bprog,&Vtnat,&Xo,&Xint,&Weff,&Lg,&Latdif,&Afg,&Depl);
/* printf ("\n%.5g %.5g %.5g %.5g %.5g %.5g %.5g %.5g %.5g %.5g %.5g %.5g",
Aers,Bers,Aprog,Bprog,Vtnat,Xo,Xint,Weff,Lg,Latdif,Afg,Depl); */
```

```
Vary=0.0;
for (j=1;j<=100;j++)
{
Vary=Vary+0.03E-6;
Weff=Vary;

Leff=Lg-(2*Latdif);
Cfg=(Afg*Eee*Eoo)/Xint;
Cfd=(Weff*Latdif*Eee*Eoo)/Xo;
Cfs=(Weff*Latdif*Eee*Eoo)/Xo;
Cfc=(Weff*Leff*Eee*Eoo)/Xo;
Ct=(Cfg+Cfd+Cfs+Cfc);
Pp=Latdif*Weff;
Pe=Lg*Weff;

Rp=Cfd/(Cfg+Cfd+Cfs+Cfc);
Re=Cfg/(Cfg+Cfd+Cfs+Cfc);
Od=Latdif-Depl;
/* printf ("\n%.5g %.5g %.5g %.5g %.5g %.5g %.5g",Cfg,Cfd,Cfs,Cfc,Pp,Pe,Leff); */

{
     Vtei1=-7.5;
     Ei=((Cfg/Ct)*(Vtnat-Vtei1))/Xo;
     Vpe=18;
     Solve_differene();
 Vtpi=(Vpe*(1.0-exp(-Tr/Tau)))-((Xo*Er*Ct)/Cfg)+Vtnat;
     Ei=((-Cfg/Ct)*(Vtnat-Vtpi))/Xo;
     Vpe=18;
     Solve_differenp();
 Vtp=(Vpe*(1.0-exp(-Tr/Tau))*((Cfd-Ct)/Cfg))+((Xo*Er*Ct)/Cfg)+Vtnat;
     Newerror=(Vtp+7.5)*(Vtp+7.5);
     Olderror=Olderror+Newerror;


     Vtpi1=5.8;
     Ei=((-Cfg/Ct)*(Vtnat-Vtpi1))/Xo;
     Vpe=18;
     Solve_differenp();
 Vtei=(Vpe*(1.0-exp(-Tr/Tau))*((Cfd-Ct)/Cfg))+((Xo*Er*Ct)/Cfg)+Vtnat;
     Ei=((Cfg/Ct)*(Vtnat-Vtei))/Xo;
     Vpe=18;
     Solve_differene();
 Vte=(Vpe*(1.0-exp(-Tr/Tau)))-((Xo*Er*Ct)/Cfg)+Vtnat;
     Newerror=(Vte-5.8)*(Vte-5.8);


     Vtpcent=(Vtp/-7.4481)*100;
     Qdpcent=(Qdp/1.4405)*100;
     Eppeakcent=(Eppeak/1.3867E9)*100;

     Vtecent=(Vte/5.5326)*100;
     Qdecent=(Qde/0.15223)*100;
     Eepeakcent=(Eepeak/1.2296E9)*100;

     printf ("\n%.5g %.5g %.5g %.5g %.5g %.5g %.5g %.5g %.5g %.5g %.5g %.5g %.5g %.5g %.5g",
Vary,Vte,Vtp,Eepeak,Eppeak,Qde,Qdp,Vtecent,Vtpcent,Eepeakcent,Eppeakcent,Qdecent,Qdpcent,Vtei,Vtpi);
}
}
}

/* ##################################################### */
Solve_differene()
{
        double K1,K2,K3,K4,Sth,T,E,Tcalc,Ecalc,X,Y;

        Eepeak=0.0;
        Qde=0.0;
        Sth=(Tfin-T0)/Nos;
        Tcalc=T0;  /* !The 1st values of E and T to be processed (using */
        Ecalc=Ei;  /* !Runge-Kutta) are the initial conditions. */

                for (i=1;i<=Nos;i=i+1)
                {
```

```
                        T=Tcalc;
                        E=Ecalc;
/* !Given two coordinates for E and T, the next adjacent coordinates are
!calculated using the Runge-Kutta method. */
                        X=T;
                        Y=E;
                        K1=Sth*FNDiff_equatione(X,Y);
                        X=T+(Sth/2);
                        Y=E+(K1/2);
                        K2=Sth*FNDiff_equatione(X,Y);
                        X=T+(Sth/2);
                        Y=E+(K2/2);
                        K3=Sth*FNDiff_equatione(X,Y);
                        X=T+Sth;
                        Y=E+K3;
                        K4=Sth*FNDiff_equatione(X,Y);
                        Tcalc=T+Sth;
                        Ecalc=E+((K1+(K2*2)+(K3*2)+K4)/6);


if (Eepeak>Ecalc) goto Eepeaksame;
Eepeak=Ecalc;
Eepeaksame:

if (Ecalc<1.0E-200) goto Qdesame;
EB=Bers/Ecalc;
if (EB>300)  Expbee=0;
else Expbee=exp(-EB);
Qde=Qde+(Aers*Ecalc*Ecalc*Expbee*Sth);
Qdesame:
        }
        Tr=Tcalc;
        Er=Ecalc;
}

/* -------------------------------------------------------------------*/
Solve_differenp()
{
        double K1,K2,K3,K4,Sth,T,E,Tcalc,Ecalc,X,Y;

        Eppeak=0.0;
        Qdp=0.0;
        Sth=(Tfin-T0)/Nos;
        Tcalc=T0;  /* !The 1st values of E and T to be processed (using */
        Ecalc=Ei;  /* !Runge-Kutta) are the initial conditions. */

                for (i=1;i<=Nos;i=i+1)
                {

                        T=Tcalc;
                        E=Ecalc;
/* !Given two coordinates for E and T, the next adjacent coordinates are
!calculated using the Runge-Kutta method. */
                        X=T;
                        Y=E;
                        K1=Sth*FNDiff_equationp(X,Y);
                        X=T+(Sth/2);
                        Y=E+(K1/2);
                        K2=Sth*FNDiff_equationp(X,Y);
                        X=T+(Sth/2);
                        Y=E+(K2/2);
                        K3=Sth*FNDiff_equationp(X,Y);
                        X=T+Sth;
                        Y=E+K3;
                        K4=Sth*FNDiff_equationp(X,Y);
                        Tcalc=T+Sth;
                        Ecalc=E+((K1+(K2*2)+(K3*2)+K4)/6);

if (Eppeak>Ecalc) goto Eppeaksame;
Eppeak=Ecalc;
Eppeaksame:
```

```
if (Ecalc<1.0E-200) goto Qdpsame;
EBprg=Bprog/Ecalc;
if (EBprg>300)  Expbep=0;
else Expbep=exp(-EBprg);
Qdp=Qdp+(Aprog*Ecalc*Ecalc*Expbep*Sth);
Qdpsame:
                }
        Tr=Tcalc;
        Er=Ecalc;
}


/* ------------------------------------------------------------------ */
double FNDiff_equatione(X,Y)
double X,Y;
{
        double Dedt,Dedt1,Dedt2,EE,TT;
        double Dedt3,Expbeint,EBint;
/* !Y=EE      X=TT
!To prvent an underflow, (-B/E) and (-T/Tau) are calculated individually.
!If the magnitude of (-B/E) is greater than 300, EXP(-B/E) will not be
!calculated, but is taken to be 0 - and likewise for EXP(-T/Tau). */
EE=Y;
TT=X;
Expttau=exp(-TT/Tau);
if (EE<1.0E-200) goto notunnelperse;

EB=Bers/EE;
if (EB>300)  Expbee=0;
else Expbee=exp(-EB);
Dedt1=(Vpe*Cfg*Expttau)/(Ct*Xo*Tau);
Dedt2=((Pe*Aers*EE*EE*Expbee)/(Xo*Ct));
Dedt=Dedt1-Dedt2;
goto returners;

notunnelperse:
Dedt=(Vpe*Cfg*Expttau)/(Ct*Xo*Tau);
returners:
return(Dedt);
}

/* ------------------------------------------------------------------ */
double FNDiff_equationp(X,Y)
double X,Y;
{
        double Dedt,Dedt1,Dedt2,EE,TT;
        double Dedt3;
/* !Y=EE      X=TT
!To prvent an underflow, (-B/E) and (-T/Tau) are calculated individually.
!If the magnitude of (-B/E) is greater than 300, EXP(-B/E) will not be
!calculated, but is taken to be 0 - and likewise for EXP(-T/Tau). */
EE=Y;
TT=X;
Expttau=exp(-TT/Tau);
if (EE<1.0E-200) goto notunnelatall;

EBprg=Bprog/EE;
if (EBprg>300)  Expbep=0;
else Expbep=exp(-EBprg);
Dedt1=(Vpe*(1.0-(Cfd/Ct))*Expttau)/(Xo*Tau);
Dedt2=((Pp*Aprog*EE*EE*Expbep)/(Xo*Ct));
Dedt=Dedt1-Dedt2;
goto returnprg;

notunnelatall:
Dedt=(Vpe*(1.0-(Cfd/Ct))*Expttau)/(Xo*Tau);
returnprg:
return(Dedt);
}
```

Appendix C

Program to Measure Threshold Voltage

The following is a program in HP Basic, written to control an HP 4145 Semiconductor Parameter Analyser. This will program and erase an EEPROM, then measure the resulting threshold voltage.

```
10      !T10
20      OPTION BASE 1
30      DIM S(201),A(201)
40      Tim=3
50      Voltage=18.0
60      PRINTER IS 701
70      PRINT "!!!!!!!!!!!!!!!!!!!!!!!!!!!!!"
80      PRINT "                          T10"
90      PRINTER IS CRT
100     !---------------
110     GOSUB Vtna3
120     Idx=Idvtn(Pt)
130     GOSUB Lists
140     !-----------------
150     GOSUB Prg3
160     GOSUB Ers3
170     GOSUB Prg3
180     GOSUB Ers3
190     GOSUB Prg3
200     GOSUB Ers3
210     GOSUB Prg3
220     GOSUB Ers3
240     Voltage=14.0
250     GOSUB Prg3
260     GOSUB Ers3
270     !---------------
```

```
280   Voltage=13.5
290   FOR Jj=1 TO 9
300   Voltage=Voltage+.5
310   !---------------
320   GOSUB Prg3
330   GOSUB Ers3
340   Vtp=0
350   Vte=0
360   FOR Ii=1 TO 2
370   !---------------
380   GOSUB Prg3
390   GOSUB Vtpa3
400   Idx=Idvtp(Pt)
410   Vtp=Vtp+Vt
420   GOSUB Lists
430   WAIT .1
440   !---------------
450   GOSUB Ers3
460   GOSUB Vtea3
470   Idx=Idvte(Pt)
480   Vte=Vte+Vt
490   GOSUB Lists
500   WAIT .1
510   !---------------
520   NEXT Ii
540   NEXT Jj
550   GOTO Endd
560   !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
570 Vtna3:  !
580   !VTNA3 10.6.92
590   DIM Idvtn(201)
600   PRINT "SET UP 4145"
610   !CHANNEL DEFINITION
620   OUTPUT 717;"DE CH1,'VCGVTN','ICGVTN',1,1" !SMU1 V VAR1
630   OUTPUT 717;"DE CH2,'VDVTN','IDVTN',1,3" !SMU2 V CONST
640   OUTPUT 717;"DE CH3,'VSRVTN','ISRVTN',3,3" !SMU3 COM CONST
650   OUTPUT 717;"DE CH4,'VSUB','ISUB',3,3" !SMU4 COM CONST
660   OUTPUT 717;"VS1;VS2;VM1;VM2" !OTHERS NOT USED
670   OUTPUT 717;"IT2 CA1 DR0 BC" !CAL ON, BUFFER CLEAR, MED
680   !SOURCE SETUP
690   OUTPUT 717;"SS VR1,-6.0,6.0,0.06,1.0E-3"  !SMU1
700   OUTPUT 717;"SS VC2,0.1,10.0E-3"           !SMU2
710   !MEASUREMENT AND DISPLAY
720   OUTPUT 717;"SS SM DM1 XN 'VCGVTN',1,-6.0,6.0"    !X-AXIS
730   OUTPUT 717;"SS SM DM1 YA 'IDVTN',1,0.0,50.0E-6"  !Y-AXIS
740   PRINT "MAKE MEASUREMENT"
750   OUTPUT 717;"MD ME1"
760 Label1:   Finn=SPOLL(717)
770           IF BIT(Finn,0)=0 THEN Label1
780           BEEP 700,.1
790   PRINT "FETCH DATA"
800   OUTPUT 717;"DO 'IDVTN'"
810   ENTER 717;Idvtn(*)
820   PRINT "CALCULATION"
```

```
830   !CALCULATE SLOPES
840   FOR J=1 TO 200
850     J2=J+1
860     S(J)=(Idvtn(J2)-Idvtn(J))/.06 !.06=VOLTAGE STEP
870   NEXT J
880   !CALCUATE AVERAGES OF SLOPES AND SIFT OUT LARGEST OF THESE
890   A(1)=(S(1)+S(2)+S(3))/3.0
900   Amax=A(1)
910   Pt=2.0
920   FOR J=2 TO 196
930     Pv=J-1
940     Nx=J+2
950     A(J)=A(Pv)+((S(Nx)-S(Pv))/3.0) !REMOVE 1ST SLOPE ADD NEXT
960     !SIFT OUT LARGEST SLOPE
970     IF A(J)>Amax THEN
980       Amax=A(J)
990       Pt=J+1.0 !CENTER POINT = 3RD ONE ALONG IN GROUP OF 5
1000    END IF
1010  NEXT J
1020  !USE Y=MX+C TO CALCULATE Vt, EXTRAPOLATE TANGENT TO Idvtn=0
1030  Vd=-6.0+(.06*Pt)
1040  Vt=(((Amax*Vd)-Idvtn(Pt))/Amax)-.05 !Vt=Vgs-Vds/2
1050  PRINT "Vt=";Vt," Idvtn=";Idvtn(Pt);" Slope=";Amax;" Pt=";Pt
1060  WAIT 2
1070  RETURN
1080  !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
1090 Prg3: !
1100  !PRG3 10.6.92
1110  PRINT "PROGRAM    "
1120  !CHANNEL DEFINITION
1130  OUTPUT 717;"IT1 CA1 DR0 BC" !CAL ON, BUFFER CLEAR, SHORT
1140  OUTPUT 717;"DE CH1,'VCGPRG','ICGPRG',3,3" !SMU1 COM CONST
1150  OUTPUT 717;"DE CH2,'VDPRG','IDPRG',1,3" !SMU2 V CONST
1160  OUTPUT 717;"DE CH4,'VSUB','ISUB',3,3" !SMU4 COM CONST
1170  OUTPUT 717;"CH3;VS1;VS2;VM1;VM2" !OTHERS NOT USED
1180  !SOURCE SETUP
1190  OUTPUT 717;"SS VC2,";Voltage;",1.0E-3"  !SMU2
1200  !MEASUREMENT AND DISPLAY
1210  OUTPUT 717;"SM WT 0.0"
1220  OUTPUT 717;"SM IN 0.01"
1230  OUTPUT 717;"SM NR ";Tim
1240  OUTPUT 717;"DM2 LI 'IDPRG'"
1250  OUTPUT 717;"MD ME1"
1260 Labela1:  Finn=SPOLL(717)
1270          IF BIT(Finn,0)=0 THEN Labela1
1280  RETURN
1290  !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
1300 Vtpa3: !
1310  !VTPA3 10.6.92
1320  DIM Idvtp(201)
1330  PRINT "SET UP 4145"
1340  !CHANNEL DEFINITION
1350  OUTPUT 717;"DE CH1,'VCGVTP','ICGVTP',1,1" !SMU1 V VAR1
1360  OUTPUT 717;"DE CH2,'VDVTP','IDVTP',1,3" !SMU2 V CONST
1370  OUTPUT 717;"DE CH3,'VSRVTP','ISRVTP',3,3" !SMU3 COM CONST
```

```
1380  OUTPUT 717;"DE CH4,'VSUB','ISUB',3,3" !SMU4 COM CONST
1390  OUTPUT 717;"DE VS1;VS2;VM1;VM2" !OTHERS NOT USED
1400  OUTPUT 717;"IT2 CA1 DR0 BC" !CAL ON, BUFFER CLEAR, MED
1410  !SOURCE SETUP
1420  OUTPUT 717;"SS VR1,0.0,-10.0,-0.05,1.0E-3"  !SMU1
1430  OUTPUT 717;"SS VC2,0.1,1.0E-2"              !SMU2
1440  !MEASUREMENT AND DISPLAY
1450  OUTPUT 717;"SS SM DM1 XN 'VCGVTP',1,0.0,-10.0"    !X-AXIS
1460  OUTPUT 717;"SS SM DM1 YA 'IDVTP',1,0.0,50.0E-6"  !Y-AXIS
1470  PRINT "MAKE MEASUREMENT"
1480  OUTPUT 717;"MD ME1"
1490 Label2:  Finp=SPOLL(717)
1500          IF BIT(Finp,0)=0 THEN Label2
1510          BEEP 700,.1
1520  PRINT "FETCH DATA"
1530  OUTPUT 717;"DO 'IDVTP'"
1540  ENTER 717;Idvtp(*)
1550  PRINT "CALCULATION"
1560  !CALCULATE SLOPES
1570  FOR J=1 TO 200
1580    J2=J+1
1590    S(J)=(Idvtp(J2)-Idvtp(J))/(-.05) !-.05=VOLTAGE STEP
1600  NEXT J
1610  !CALCUATE AVERAGES OF SLOPES AND SIFT OUT LARGEST OF THESE
1620  A(1)=(S(1)+S(2)+S(3))/3.0
1630  Amax=A(1)
1640  Pt=2.0
1650  FOR J=2 TO 196
1660    Pv=J-1
1670    Nx=J+2
1680    A(J)=A(Pv)+((S(Nx)-S(Pv))/3.0) !REMOVE 1ST SLOPE ADD NEXT
1690    !SIFT OUT LARGEST SLOPE
1700    IF A(J)>Amax THEN
1710      Amax=A(J)
1720      Pt=J+1.0 !CENTER POINT = 2RD ONE ALONG IN GROUP OF 3
1730    END IF
1740  NEXT J
1750  !USE Y=MX+C TO CALCULATE Vt, EXTRAPOLATE TANGENT TO I=0
1760  Vd=-.05*Pt
1770  Vt=(((Amax*Vd)-Idvtp(Pt))/Amax)-.05 !Vt=Vgs-Vds/2
1780  PRINT "Vt=";Vt," Idvtp=";Idvtp(Pt);" Slope=";Amax;" Pt=";Pt
1790  RETURN
1800  !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
1810 Ers3: !
1820  !ERS3 10.6.92
1830  PRINT "ERASE       "
1840  !CHANNEL DEFINITION
1850  OUTPUT 717;"IT1 CA1 DR0 BC" !CAL ON, BUFFER CLEAR, SHORT
1860  OUTPUT 717;"DE CH1,'VCGERS','ICGERS',1,3" !SMU1 V CONST
1870  OUTPUT 717;"DE CH2,'VDERS','IDERS',3,3" !SMU2 COM CONST
1880  OUTPUT 717;"DE CH3,'VSRERS','ISRERS',3,3" !SMU3 COM CONST
1890  OUTPUT 717;"DE CH4,'VSUB','ISUB',3,3" !SMU4 COM CONST
1900  OUTPUT 717;"VS1;VS2;VM1;VM2" !OTHERS NOT USED
1910  !SOURCE SETUP
```

```
1920  OUTPUT 717;"SS VC1,";Voltage;",1.0E-3"  !SMU1
1930  !MEASUREMENT AND DISPLAY
1940  OUTPUT 717;"SM WT 0.0"
1950  OUTPUT 717;"SM IN 0.01"
1960  OUTPUT 717;"SM NR ";Tim
1970  OUTPUT 717;"DM2 LI 'ICGERS'"
1980  OUTPUT 717;"MD ME1"
1990 Labela2:   Finn=SPOLL(717)
2000            IF BIT(Finn,0)=0 THEN Labela2
2010  RETURN
2020  !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
2030 Vtea3: !
2040  !VTEA3 10.6.92
2050  DIM Idvte(201)
2060  PRINT "SET UP 4145"
2070  !CHANNEL DEFINITION
2080  OUTPUT 717;"DE CH1,'VCGVTE','ICGVTE',1,1" !SMU1 V VAR1
2090  OUTPUT 717;"DE CH2,'VDVTE','IDVTE',1,3" !SMU2 V CONST
2100  OUTPUT 717;"DE CH3,'VSRVTE','ISRVTE',3,3" !SMU3 COM CONST
2110  OUTPUT 717;"DE CH4,'VSUB','ISUB',3,3" !SMU4 COM CONST
2120  OUTPUT 717;"VS1;VS2;VM1;VM2" !OTHERS NOT USED
2130  OUTPUT 717;"IT2 CA1 DR0 BC" !CAL ON, BUFFER CLEAR, MED
2140  !SOURCE SETUP
2150  OUTPUT 717;"SS VR1,0.0,10.0,0.05,1.0E-3"  !SMU1
2160  OUTPUT 717;"SS VC2,0.1,1.0E-2"            !SMU2
2170  !MEASUREMENT AND DISPLAY
2180  OUTPUT 717;"SS SM DM1 XN 'VCGVTE',1,0.0,10.0"    !X-AXIS
2190  OUTPUT 717;"SS SM DM1 YA 'IDVTE',1,0.0,50.0E-6"  !Y-AXIS
2200  PRINT "MAKE MEASUREMENT"
2210  OUTPUT 717;"MD ME1"
2220 Label3:  Fine=SPOLL(717)
2230          IF BIT(Fine,0)=0 THEN Label3
2240          BEEP 700,.1
2250  PRINT "FETCH DATA"
2260  OUTPUT 717;"DO 'IDVTE'"
2270  ENTER 717;Idvte(*)
2280  PRINT "CALCULATION"
2290  !CALCULATE SLOPES
2300  FOR J=1 TO 200
2310    J2=J+1
2320    S(J)=(Idvte(J2)-Idvte(J))/(.05) !.05=VOLTAGE STEP
2330  NEXT J
2340  !CALCUATE AVERAGES OF SLOPES AND SIFT OUT LARGEST OF THESE
2350  A(1)=(S(1)+S(2)+S(3))/3.0
2360  Amax=A(1)
2370  Pt=2.0
2380  FOR J=2 TO 196
2390    Pv=J-1
2400    Nx=J+2
2410    A(J)=A(Pv)+((S(Nx)-S(Pv))/3.0) !REMOVE 1ST SLOPE ADD NEXT
2420    !SIFT OUT LARGEST SLOPE
2430    IF A(J)>Amax THEN
2440      Amax=A(J)
2450      Pt=J+1.0 !CENTER POINT = 2RD ONE ALONG IN GROUP OF 3
2460    END IF
```

```
2470  NEXT J
2480  !USE Y=MX+C TO CALCULATE Vt, EXTRAPOLATE TANGENT TO I=0
2490  Vd=.05*Pt !STEP HEIGHT*NUMBER OF STEPS
2500  Vt=(((Amax*Vd)-Idvte(Pt))/Amax)-.05 !Vt=Vgs-Vds/2
2510  PRINT "Vt=";Vt," Idvte=";Idvte(Pt);" Slope=";Amax;" Pt=";Pt
2520  RETURN
2530  !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
2540 Lists: !
2550  PRINTER IS 701
2560  Time=(Tim*.01)-.01
2570  PRINT "Vt=";Vt,"  Voltage=";Voltage,"Id=";Idx,"Time=";Time
2580  PRINTER IS CRT
2590  RETURN
2600  !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
2610 Listsv: !
2620  PRINTER IS 701
2630  Vtp=Vtp/(Ii-1)
2640  Vte=(Vte/(Ii-1))/2
2650  PRINT "Vtp average=";Vtp,"  Vte average=";Vte
2660  PRINT "-----------------------------------------"
2670  PRINTER IS CRT
2680  RETURN
2690  !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
2700 Endd: !
2710  PRINT "END"
2720  END
```

Appendix D

Process Simulation

Two process simulation programs follow. The first is for TMA SUPREM-3, and the second for TSUPREM-4.

```
TITLE          LL4B.DAT

INITIALIZE     <100> SILICON BORON=8E14
+              THICKNESS=2.0  DX=.001  XDX=0.02  SPACES=240

COMMENT        2. pad oxide aim for 350A
DIFFUSION      TIME=5 TEMPERATURE=950 DRYO2 HCL%=5
DIFFUSION      TIME=5 TEMPERATURE=950 STEAM HCL%=5
DIFFUSION      TIME=5 TEMPERATURE=950 DRYO2 HCL%=0

COMMENT        3. nitride deposition
DEPOSITION     NITRIDE THICKNESS=0.1 DX=0.005 SPACES=20
DIFFUSION      TIME=100 TEMPERATURE=800 INERT

COMMENT        8. field oxidation
DIFFUSION      TIME=5 TEMPERATURE=950 DRYO2 HCL%=5
DIFFUSION      TIME=30 TEMPERATURE=950 STEAM HCL%=5
DIFFUSION      TIME=930 TEMPERATURE=950 STEAM HCL%=0
DIFFUSION      TIME=5 TEMPERATURE=950 DRYO2 HCL%=0

COMMENT        12. and 13. 4:1 etch and nitride wet etch
ETCH           OXIDE
ETCH           NITRIDE

COMMENT        14. boron implant
IMPLANT        BORON DOSE=2E12 ENERGY=50

COMMENT        15. remove pad oxide
ETCH           OXIDE

COMMENT        17. sacrificial oxide
DIFFUSION      TIME=5 TEMPERATURE=950 DRYO2 HCL%=5
DIFFUSION      TIME=5 TEMPERATURE=950 STEAM HCL%=5
DIFFUSION      TIME=5 TEMPERATURE=950 DRYO2 HCL%=0

COMMENT        19. remove sacrificial oxide
ETCH           OXIDE
```

```
COMMENT         21. gate oxide (tonox)
DIFFUSION       TIME=20 TEMPERATURE=800 T.FINAL=900 INERT
DIFFUSION       TIME=5 TEMPERATURE=900 DRYO2 HCL%=0
LOOP            OPTIMIZE
ASSIGN          NAME=TM N.VALUE=0 LOWER=1 UPPER=100 OPTIMIZE
DIFFUSION       TIME=0+@TM TEMPERATURE=900 DRYO2 HCL%=5
EXTRACT         NAME=TOX THICKNESS LAYER=2 TARGET=0.011
L.END
DIFFUSION       TIME=10 TEMPERATURE=900 INERT
DIFFUSION       TIME=20 TEMPERATURE=900 T.FINAL=800 INERT

COMMENT         22. poly deposition
DEPOSIT         POLYSILICON THICKNESS=0.039 TEMPERATURE=600

COMMENT         aresnic implant
IMPLANT         ARSENIC DOSE=2E15 ENERGY=160

COMMENT         22. poly deposition
DEPOSIT         POLYSILICON THICKNESS=0.6 TEMPERATURE=600

COMMENT         23. implant to dope gate
IMPLANT         ARSENIC  DOSE=2E16 ENERGY=50

COMMENT         25. high temp
LOOP            OPTIMIZE
ASSIGN          NAME=TMP N.VALUE=1000 LOWER=1000 UPPER=1100 OPTIMIZE
DIFFUSION       TIME=30 TEMP=@TMP-150 T.FINAL=@TMP INERT
DIFFUSION       TIME=60 TEMP=@TMP INERT
DIFFUSION       TIME=30 TEMP=@TMP T.FINAL=@TMP-150 INERT
DIFFUSION       TIME=5 TEMPERATURE=950 INERT
DIFFUSION       TIME=5 TEMPERATURE=950 INERT
DIFFUSION       TIME=5 TEMPERATURE=950 INERT
EXTRACT         NAME=AJ NET ACTIVE X.EXTRACT LAYER=1
+               Y=0.0 TARGET=0.675
L.END

COMMENT         calculate poly sheet res, use zero bias analysis
ELECTRICAL
END

EXTRACT         NAME=SR LAYER=3 E.RESIST
EXTRACT         NAME=AR MIN.REGI=2 LAYER=1 E.RESIST
EXTRACT         NAME=AJ NET ACTIVE LAYER=1 X.EXTRACT Y=0
EXTRACT         NAME=TX LAYER=2 THICKNESS
EXTRACT         NAME=TP LAYER=3 THICKNESS
PLOT            ACTIVE BORON LINE=2 DEVICE=1/postscript
+               TITLE="Doping Profiles: Arsenic Device, Drain"
PLOT            ACTIVE ARSENIC ADD LINE=3
LABEL           LABEL="Arsenic"  START.RI LX.F=2.0 LINE=3 X=1.2 Y=3E20
LABEL           LABEL="Boron"  START.RI LX.F=2.0 LINE=2
LABEL           LABEL="Arsenic junction depth: "@AJ" um"
LABEL           LABEL="Drain sheet res.: "@AR" ohms/square"
LABEL           LABEL="Polysilicon thickness: "@TP" um"
LABEL           LABEL="Polysilicon sheet res.: "@SR" ohms/square"
LABEL           LABEL="Gate Oxide thickness: "@TX" um"
LABEL           LABEL="Anneal temperature : "@TMP
```

```
DEFINE        GDENS 12
LINE X        LOCATION=0.0 SPACING=(0.2/ GDENS )
LINE X        LOCATION=1.2 SPACING=(0.2/ GDENS )
LINE Y        LOCATION=0.0 SPACING=(0.2/ GDENS )
LINE Y        LOCATION=1.0 SPACING=(0.2/ GDENS )
LINE Y        LOCATION=2.0 SPACING=(0.4/ GDENS )

INITIALIZE    <100> BORON=8E14

METHOD        VERTICAL

DIFFUSION     TIME=5 TEMP=950 DRY HCL=5
DIFFUSION     TIME=5 TEMP=950 STEAM HCL=5
DIFFUSION     TIME=5 TEMP=950 DRY HCL=0

DEPOSIT       NITRIDE THICKNESS=.1
DIFFUSION     TIME=100 TEMP=800 INERT

DIFFUSION     TIME=5 TEMP=950 INERT
DIFFUSION     TIME=30 TEMP=950 INERT
DIFFUSION     TIME=930 TEMP=950 INERT
DIFFUSION     TIME=5 TEMP=950 INERT

ETCH          NITRIDE  ALL

IMPLANT       BORON DOSE=2E12 ENERGY=50

ETCH          OXIDE ALL

DIFFUSION     TIME=5 TEMP=950 DRY HCL=5
DIFFUSION     TIME=5 TEMP=950 STEAM HCL=5
DIFFUSION     TIME=5 TEMP=950 DRY HCL=0

ETCH          OXIDE ALL

DIFFUSION     TIME=20 TEMP=800 T.FINAL=900 INERT
DIFFUSION     TIME=10 TEMP=900 DRY HCL=0
DIFFUSION     TIME=12 TEMP=900 DRY HCL=5
DIFFUSION     TIME=60 TEMP=900 INERT
DIFFUSION     TIME=20 TEMP=900 T.FINAL=800 INERT
ETCH          OXIDE ALL
DEPOSIT       OXIDE THICKNESS=0.011 SPACES=1

DEPOSIT       POLYSILICON THICKNES=0.039 SPACES=4

DEPOSIT       PHOTORES THICKNES=1.0 SPACES=5
ETCH          PHOTORES RIGHT P1.X=0.8 P2.X=0.8

IMPLANT       ARSENIC DOSE=2E15 ENERGY=160

ETCH          PHOTORES ALL
DEPOSIT       POLYSILI THICKNESS=0.6 SPACES=6

IMPLANT       ARSENIC DOSE=2.0E16 ENERGY=50

DIFFUSION     TIME=30 TEMP=922 T.FINAL=1072 INERT
DIFFUSION     TIME=60 TEMP=1072 INERT
DIFFUSION     TIME=30 TEMP=1072 T.FINAL=922 INERT

ETCH          OXIDE RIGHT P1.X=0.8 P2.X=0.8
ETCH          POLYSILI ALL
ETCH          OXIDE RIGHT P1.X=0.8 P2.X=0.8

DEPOSIT       PHOTORES THICKNESS=0.6 SPACES=6
ETCH          PHOTORES RIGHT P1.X=0.8 P2.X=0.8
STRUCTURE     OUTFILE=KEEPASR
```

Appendix E

Program to Analyse POT Devices

The following gives the key sections of a program in HP BASIC, written to control an HP 4062 Semiconductor Parameter Test System, and KLA Automatic Wafer Prober. This conducts subthreshold and time dependent dielectric tests on POT devices.

```
830   !-------------------------------------------------------------
840
850 Auto:! first set up chip positions to be tested
851   Open=0
860   DIM Cdatax(50),Cdatay(50)
870   DATA 4,5,6,8,8,8,9,10,10,10,10,10       !CX
880   DATA 10,10,10,1,9,10,2,3,4,5,6,7      !CY
910   !
911   FOR I=1 TO 12
920   READ Cdatax(I)
930   NEXT I
940   FOR I=1 TO 12
950   READ Cdatay(I)
960   NEXT I
1000  !
1001  FOR Chip=1 TO 12
1010  !1st CONDUCT SYMMETRY ANALYSIS
1023  FOR Tyyy=1 TO 20
1024  Prog$(1)="C"
1026  Prog$(2)=VAL$(Cdatax(Chip))
1027  Prog$(3)=VAL$(Cdatay(Chip))
1028  Prog$(4)="T"
1029  Prog$(5)="2"
1030  Prog$(6)=VAL$(Tyyy)
1031  Prog$(7)="7"
1032  Prog$(8)="N1"
1033  Prog$(9)=" "
1034  Prog$(10)=" "
1035  Prog$(11)=" "
1038 GOTO Menu_execute !must leave this for next loop to do test
```

```
1042 Next_test1: !return to for next loop at this label
1045  NEXT Tyyy
1050  Prog$(1)="7S"
1051  Prog$(2)="N4"
1052 GOTO Menu_execute !must leave this for next loop to do test
1053 Next_test4: !return to for next loop at this label
1059  !
1062  !TDDB ON GATE/DRAIN OVERLAP
1063  Drainrupt=1
1064  Sourcerupt=0
1065  Gdover=0
1067  FOR Tyyy=Tysymm-4 TO Tysymm+5
1068  Gdover=Gdover+.05
1069  Prog$(1)="C"
1070  Prog$(2)=VAL$(Cdatax(Chip))
1071  Prog$(3)=VAL$(Cdatay(Chip))
1072  Prog$(4)="T"
1073  Prog$(5)="2"
1074  Prog$(6)=VAL$(Tyyy)
1075  Prog$(7)="3"
1076  Prog$(8)="3S"
1077  Prog$(9)="N2"
1078  Prog$(10)=" "
1080 GOTO Menu_execute !must leave this for next loop to do test
1081 Next_test2: !return to for next loop at this label
1084  NEXT Tyyy
1284  !
1285  !TDDB SOURCE/GATE OVERLAP
1286  Drainrupt=0
1287  Sourcerupt=1
1288  Gdover=0
1290  FOR Tyyy=Tysymm+4 TO Tysymm-5 STEP -1
1291  Gdover=Gdover+.05
1292  Prog$(1)="C"
1294  Prog$(2)=VAL$(Cdatax(Chip))
1295  Prog$(3)=VAL$(Cdatay(Chip))
1296  Prog$(4)="T"
1297  Prog$(5)="2"
1298  Prog$(6)=VAL$(Tyyy)
1299  Prog$(7)="3"
1300  Prog$(8)="3S"
1301  Prog$(9)="N3"
1302  Prog$(10)=" "
1303  Prog$(11)=" "
1304 GOTO Menu_execute !must leave this for next loop to do test
1305 Next_test3: !return to for next loop at this label
1309  NEXT Tyyy
1312  !
1313  NEXT Chip
1314  ASSIGN @Path TO *
1315  RETURN
1316 Skip_auto: !
1317  !-------------------------------------------------------------
```

```
2086  !-----------------------------------------------------------------
2088 Start_test7:! SUBTHRESHOLD
2089             ! TWO Vt SWEEPS, ONE FROM DRAIN, THE OTHER SOURCE
2090  Point=12
2091  MAT Rdng= (0)
2092  MAT Inp= (0)
2093  Test_code$="G"
2094  Forrevdiff(Ty)=0
2095  Devbust(Ty)=0
2096  Num=2!Number of sweeps for characteritics
2097  !NB...Num is used in curve_plot, line 5230, and must be given a value.
2098  Voidx=Tx
2099  IF Voidx MOD 2.0=0 THEN
2100 Txeven7:                 !1st determin whether the ty is odd or even
2101  Void=Ty                 !MOD only returns real result for real arguments.
2102  IF Void MOD 2.0=0. THEN !MOD returns the remainder of a division.
2103  Drain=10
2104  Gate=11                 !Even
2105  Source=14
2106  Ssub=44
2107  ELSE                    !Odd
2108  Drain=10
2109  Gate=11
2110  Source=14
2111  Ssub=44
2112  END IF
2113  ELSE
2114 Txodd7:                  !1st determin whether the ty is odd or even
2115  Void=Ty                 !MOD only returns real result for real arguments.
2116  IF Void MOD 2.0=0. THEN !MOD returns the remainder of a division.
2117  Drain=11
2118  Gate=14                 !Even
2119  Source=16
2120  Ssub=44
2121  ELSE                    !Odd
2122  Drain=11
2123  Gate=14
2124  Source=16
2125  Ssub=44
2126  END IF
2127  END IF
2128  !
2129  Vd=.5
2130  Vb=0.
2131  Vstart=.1
2132  Vstop=.4
2133  Vstep=(Vstop-Vstart)/(Point-1)
2134  Compliance=1.E-3
2135  Integ_time=2
2136  !
2137  Init_system
2138  Set_smu(Integ_time)
2139  Connect(FNSmu(1),Gate)
2140  Connect(FNSmu(2),Drain)
```

```
2141  Connect(FNGnd,Source)
2142  Connect(FNSmu(3),Ssub)
2143  Force_v(Drain,Vd,0,9.0E-2)
2144  Force_v(Ssub,Vb,0,9.0E-2)
2145  Set_iv(Gate,1,0,Vstart,Vstop,Point,0,0,Compliance)
2146  Sweep_iv(Drain,2,0,Rid7(*),Rvg7(*))
2147  Disable_port
2148  Connect
2149  !
2150  Init_system
2151  Set_smu(Integ_time)
2152  Connect(FNSmu(1),Gate)
2153  Connect(FNGnd,Drain)
2154  Connect(FNSmu(2),Source)
2155  Connect(FNSmu(3),Ssub)
2156  Force_v(Source,Vd,0,9.0E-2)
2157  Force_v(Ssub,Vb,0,9.0E-2)
2158  Set_iv(Gate,1,0,Vstart,Vstop,Point,0,0,Compliance)
2159  Sweep_iv(Source,2,0,Ris7(*),Rvg7(*))
2160  Disable_port
2161  Connect
2162  !
2166  Forrevdiff(Ty)=(LGT(ABS(Rid7(Point))))-(LGT(ABS(Ris7(Point))))
2167  Forrevdiff(Ty)=ABS(Forrevdiff(Ty))
2169  IF (Rid7(1)>2.E-9) AND (Ris7(1)>2.E-9) THEN Devbust(Ty)=1 !SC
2170  IF (Rid7(Point)<2.E-9) AND (Ris7(Point)<2.E-9) THEN Devbust(Ty)=1 !OC
2171  !
2173!!PRINT Rid7(Point),Ris7(Point),Devbust(Ty),Forrevdiff(Ty)
2175  RETURN
2176  !-----------------------------------------------------------------------
2177 Start_test7s:! SORT RESULTS OF SUB-THRESHOLD
2189  FOR Tysort=10 TO 20
2190  IF (Devbust(Tysort)=1) THEN GOTO Next_device1
2191  IF (1.7<Forrevdiff(Tysort)) THEN
2192  IF (2.8<Forrevdiff(Tysort)) THEN
2193  Tysymm=Tysort-10
2194  GOTO End_of_sorting
2195  END IF
2196  Tysymm=Tysort-9
2197  GOTO End_of_sorting
2198  END IF
2199 Next_device1: NEXT Tysort
2200  !
2201  !
2202  FOR Tysort=10 TO 1 STEP -1
2203  IF (Devbust(Tysort)=1) THEN GOTO Next_device2
2204  IF (1.7<Forrevdiff(Tysort)) THEN
2205  IF (2.8<Forrevdiff(Tysort)) THEN
2206  Tysymm=Tysort+10
2207  GOTO End_of_sorting
2208  END IF
2209  Tysymm=Tysort+9
2210  GOTO End_of_sorting
2211  END IF
2212 Next_device2: NEXT Tysort
2213 End_of_sorting:  !
2214 RETURN
2215 !-------------------------------------------------------------
```

```
1745  !---------------------------------------------------------------
1746 Start_test3:! STRESS TEST     TDDB
1747             ! vg=0 vsub=0 vs=Float  vd=CONSTANT
1748  Point=300  !10 MINUTES MAX
1749  Trupt=0
1750  Charge3=0
1751  MAT Inp= (0)
1752  MAT Rdng= (0)
1753  Num=1!Number of sweeps for characteritics
1754  !NB...Num is used in curve_plot
1755  Voidx=Tx
1756  IF Voidx MOD 2.0=0 THEN
1757 Txeven3:                 !1st determin whether the ty is odd or even
1758  Void=Ty                 !MOD only returns real result for real arguments.
1759  IF Void MOD 2.0=0. THEN !MOD returns the remainder of a division.
1760  Drain=10
1761  Gate=11                 !Even
1762  Source=14
1763  Ssub=44
1764  ELSE                    !Odd
1765  Drain=10
1766  Gate=11
1767  Source=14
1768  Ssub=44
1769  END IF
1770  ELSE
1771 Txodd3:                  !1st determin whether the ty is odd or even
1772  Void=Ty                 !MOD only returns real result for real arguments.
1773  IF Void MOD 2.0=0. THEN !MOD returns the remainder of a division.
1774  Drain=11
1775  Gate=14                 !Even
1776  Source=16
1777  Ssub=44
1778  ELSE                    !Odd
1779  Drain=11
1780  Gate=14
1781  Source=16
1782  Ssub=44
1783  END IF
1784  END IF
1785  !
1786  IF Drainrupt=1 THEN
1787  PRINT "DRAIN STRESS"
1788  !STRESS DRAIN/GATE OVERLAP
1789  Vd=10.0
1790  Vg=0.
1791  Vb=0.
1792  !
1793  Compliance=1.E-4
1794  Integ_time=2 !        INTEGRATION
1795  Init_system
1796  Set_smu(Integ_time)  !INTEGRATION
1797  Connect(FNSmu(1),Gate)
1798  Connect(FNSmu(2),Drain)
1799  Connect(FNSmu(3),Ssub)
```

```
1800  Force_v(Ssub,Vb,0,Compliance)
1801  Force_v(Gate,Vg,0,Compliance)
1802  !
1803  Charge3=0.
1804  MAT Rig3= (0.)
1805  MAT Rtm3= (0.)
1806  Start_time=TIMEDATE
1807  Force_v(Drain,Vd,0,Compliance)
1808  FOR I=2 TO Point
1809  Measure_i(Gate,Rig3(I),0)
1810  Rtm3(I)=TIMEDATE-Start_time
1811  Pointcounter=I
1812  IF ABS(Rig3(I))>1.E-8 AND I>5 THEN
1813  Charge3=0
1814  GOTO Endtddbdrain3
1815  END IF
1816  Charge3=Charge3+(Rig3(I)*(Rtm3(I)-Rtm3(I-1)))
1817  WAIT 2.0
1818  NEXT I
1819 Endtddbdrain3: !
1820  Disable_port
1821  Connect
1822  END IF
1823  !
1824  IF Sourcerupt=1 THEN
1825  PRINT "SOURCE STRESS"
1829  !STRESS SOURCE/GATE OVERLAP
1835  Vs=10.0
1836  Vg=0.
1837  Vb=0.
1838  !
1839  Compliance=1.E-4
1840  Integ_time=2 !       INTEGRATION
1841  Init_system
1842  Set_smu(Integ_time)  !INTEGRATION
1843  Connect(FNSmu(1),Gate)
1844  Connect(FNSmu(2),Source)
1845  Connect(FNSmu(3),Ssub)
1846  Force_v(Ssub,Vb,0,Compliance)
1847  Force_v(Gate,Vg,0,Compliance)
1848  !
1849  Charge3=0.
1850  MAT Rig3= (0.)
1851  MAT Rtm3= (0.)
1852  Start_time=TIMEDATE
1853  Force_v(Source,Vd,0,Compliance)
1854  FOR I=2 TO Point
1855  Measure_i(Gate,Rig3(I),0)
1856  Rtm3(I)=TIMEDATE-Start_time
1857  Pointcounter=I
1858  IF ABS(Rig3(I))>1.E-8 AND I>5 THEN
1859  Charge3=0
1860  GOTO Endtddbsource3
1861  END IF
```

```
1863  Charge3=Charge3+(Rig3(I)*(Rtm3(I)-Rtm3(I-1)))
1864  WAIT 2.0
1865  NEXT I
1866 Endtddbsource3: !
1867  Disable_port
1868  Connect
1869  END IF
1870  !
1871  Point=Pointcounter
1876  Trupt=Rtm3(Pointcounter)
1886!!PRINT Pointcounter,Rig3(Pointcounter),Trupt,Charge3
1890  RETURN
1891  !-------------------------------------------------------------


2943 Store_res_qbd: !
2944  Filename$="ASW8C"
2945  IF Open=0 THEN
2946  CREATE ASCII Filename$,20
2947  ASSIGN @Path TO Filename$
2948  Open=1
2949  END IF
2950 Already_open: !
2955 OUTPUT K1$ USING "#,.DD";Gdover
2956 OUTPUT K2$ USING "#,D.4DE";Trupt
2957 OUTPUT K3$ USING "#,D.4DE";Qbd
2958 K2o$=K1$&","&K2$&","&K3$
2959 OUTPUT @Path;K2o$
2960  RETURN
2961  !-------------------------------------------------------------
```