Durham University

Durham E-Theses

Characterisation of silicon mis negative resistance devices

Clifton, Paul Alan

How to cite:

Clifton, Paul Alan (1989) *Characterisation of silicon mis negative resistance devices*, Durham theses, Durham University. Available at Durham E-Theses Online: http://etheses.dur.ac.uk/6434/

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in Durham E-Theses
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the full Durham E-Theses policy for further details.

Academic Support Office, The Palatine Centre, Durham University, Stockton Road, Durham, DH1 3LE
e-mail: e-theses.admin@durham.ac.uk Tel: +44 0191 334 6107
http://etheses.dur.ac.uk

CHARACTERISATION OF SILICON MIS NEGATIVE RESISTANCE DEVICES

by
Paul Alan Clifton
BSc (L'pool)

A thesis submitted for the
Degree of Doctor of Philosophy
in the University of Durham
June 1989

The copyright of this thesis rests with the author.
No quotation from it should be published without
his prior written consent and information derived
from it should be acknowledged.

3 1 OCT 1990

Declaration of Originality

I hereby declare that the work presented in this thesis has not previously been nor is currently submitted for any other degree.

The work reported in this thesis was carried out by the candidate, except where indicated by reference to other authors.

DEDICATION

To my parents.

ABSTRACT

Metal-insulator-semiconductor switches (MISS), in which the 'I' denotes some form of thin semi-insulating layer and the semiconductor part consists of a pn junction, are part of the general class of regenerative switching devices which includes the thyristor. The switching behaviour of the MISS derives from the ability of the MIS junction to exhibit current gain and to exist in two modes, deep depletion and inversion. In this thesis, a general model for the regenerative switching is proposed after investigating the properties of the MIS junction both theoretically and experimentally. Results from MIS diodes with tunnelling-thickness oxide I-layers indicate that interface states play a dominant role in their electrical behaviour and that the uniformity of the oxide is poor, giving rise to a large spread in the current-voltage characteristics. Subsequently, the epitaxial form of the MISS device is investigated and in particular the importance of isolation of the pn junction. It is concluded that spreading effects set a practical lower limit to the device dimensions, making the epitaxial form unsuitable for microelectronic applications. An alternative semi-insulator, 'silicon-rich oxide' (SRO) is introduced as an optional I-layer with possibly greater integrity than tunnel oxide. MIS diodes formed with SRO are shown to have very similar properties to tunnelling diodes. Large area devices fabricated using this material are surprisingly discovered to exhibit stable negative differential resistance (NDR). Although this discovery at first appears to be contrary to normal circuit stability criteria and to the regenerative feedback model itself, both of these points are resolved. It is shown that the frequency of oscillation of an unstable device is controlled by the external circuit. Then it is proposed that if this frequency is greater than the maximum frequency of operation of the regenerative mechanism, stable NDR is observed. In the final chapter, alternative lateral MISS structures which should overcome the geometrical limitations of epitaxial devices are discussed.

ACKNOWLEDGEMENTS

Thanks are due to my supervisor Dr M.J.Morant for obtaining the SERC quota award which allowed me to undertake these studies. Special thanks go to Steve Lavelle with whom I enjoyed many fruitful theoretical discussions and whose rigorous approach to device modelling has raised the level of our understanding of the MISS device. The clean room technician, Mr.J.Gibson is also acknowledged for his very capable management of the fabrication facilities and for assistance in device processing. Closer to home, I thank my partner Sue for her great patience during the period of writing this thesis and for giving me the personal stability to see it through. Finally I thank all my family and friends who gave me much needed encouragement.

CONTENTS

1. Introduction

Page

1.1 **Regenerative Switching Devices** 1

1.2 **The MISS** 2

1.3 **Thesis Outline** 4

2. Theory of the Non-Equilibrium MIS Diode

2.1 **Introduction** 6

2.2 **Equilibrium MIS Diodes** 6

2.2.1 **The MIS Diode Under Zero Bias.** 7

2.2.2 **MIS Diodes Under Reverse Bias.** 7

2.2.3 **Inversion in MIS Diodes.** 9

2.3 **The Non-Equilibrium MIS Diode** 10

2.3.1 **Disequilibrium In Reverse Bias.** 10

2.4 **Generation in the Surface Depletion Layer** 12

2.5 **Tunnelling Currents Through Thin Oxides.** 13

2.5.1 **Introduction.** 13

2.5.2. **Electron Tunnel Current.** 14

2.5.3.**Hole Tunnel Current.** 15

2.5.4. **Tunnelling Via Surface States.** 16

2.6 **Conclusions** 18

3. Fabrication and Characterisation of Tunnel Oxide MIS Diodes

3.1 Introduction 21

3.2 MIS Diodes with Tunnelling Thickness SiO_2 22

3.2.1. Aluminium Top Contact 22

3.2.2. The Aluminium-Oxide Interface 22

3.2.3. Stoichiometric SiO_2 23

3.2.4. The Oxide-Silicon Interface 23

3.2.5. The Silicon Surface 24

3.3 Theory of Thin Oxide Growth 25

3.4 Fabrication of Tunnel Oxide MIS Diodes. 27

3.4.1. Cleaning of Substrates 27

3.4.2 Growth of Field Oxide 28

3.4.3. Patterning of the MIS Windows 28

3.4.4. Growth of Tunnel Oxides 29

3.4.5 Evaporation and Patterning of the Aluminium Contacts 30

3.4.6 Back Contact Metallisation 31

3.5 Determination of Tunnel Oxide Thickness 31

3.5.1 Measurement of Ultra-Thin Oxides by Ellipsometry 31

3.5.2 Measurement of Accumulated Capacitance 32

3.6 Electrical Characterisation of MIS Diodes 33

3.7 Current-Voltage Characteristics 33

3.7.1 Forward Bias 34

3.7.2 Reverse Bias 38

3.7.3 The Reverse Saturation Current 38

3.8 Capacitance-Voltage Characterisation of MIS Diodes 41

4. A Theoretical Basis for the Study of MISS Devices

4.1 Introduction 47

4.2 A Steady State Model of the MISS 48

4.3 The pn Junction 49

4.3.1 Recombination in the pn Junction 50

4.3.2 Hole Diffusion Current. 50

4.3.3 Electron Diffusion Current. 52

4.4 The MIS Junction 53

4.4.1 Direct Tunnel Currents. 53

4.4.2 Generation in the Surface Depletion Region. 54

4.5 Formulation of the Steady State Model 54

4.6 A General outline of the Switching Mechanism. 55

4.6.1 The Off State. 56

4.6.2 The Switching Point. 56

4.6.3 The NDR Region. 57

4.6.4 The Holding Point. 58

4.6.5 The On State. 58

4.7 The Modes of Operation of the MISS Device 58

4.7.1 The 'Avalanche' Mode. 58

4.7.2 The 'Punch-Through' Mode. 59

4.7.3 The Generation-Controlled Mode. 59

4.7.4 The 'High Current' Mode. 60

4.7.5 A Synopsis of the pn-MIS Junction Interaction. 60

4.8 **Regenerative Switching in the MISS.** 61

4.8.1 **Introduction.** 61

4.8.1 **The Regenerative Feedback Model of the MISS.** 61

4.9 **A Proposed Model of the MISS Consistent with Thyristor Theory.** 63

4.9.1 **The MIS as an 'Inversion Base Bipolar Transistor'.** 64

4.9.2 **The Small-Signal Gain of a Tunnel Oxide IBT** 66

4.9.3 **The Regenerative Feedback Loop Gain.** 67

4.9.4 **Limitations of the Regenerative Feedback Model.** 68

4.10 **Conclusions** 69

5. Epitaxial Tunnel Oxide MISS Devices

5.1 **Introduction** 71

5.2 **Etching of V-grooves in {100} Silicon** 72

5.2.1 **Characterisation of the V-groove Etching Process** 74

5.2.2 **Measurement of Groove Depth** 75

5.2.3 **Photolithography on a Grooved Topology.** 75

5.2.4 **Fabrication of Isolated Epitaxial MISS Devices** 76

5.3 **Switching Characteristics of Isolated Epitaxial MISS Devices.** 77

5.3.1 **Measurement of MISS Switching Characteristics** 77

5.3.2 **The Myth of the Third State** 78

5.3.3 **Automated Measurement of MISS Switching Characteristics** 79

5.4 Initial Forming of MISS Switching Characteristics. 80

5.4.1 Possible Causes of Forming 81

5.4.2 Generation of Interface States in MIS Diodes 82

5.5 Area Effects in the Isolated MISS. 83

5.5.1 Measured Effects on Device Characteristics 83

5.5.2 Effects of Area on the Feedback Gain 85

5.5.3 Current Spreading in the Stripe-Contact Laser 87

5.5.4 Current Spreading in Epitaxial MISS Devices 88

5.5.5 Evaluation of the Current Spreading in MISS Devices. 89

5.5.6 The Spreading Length in 'Punch-Through' Mode Devices 91

5.5.7 The MIS Aspect Ratio and Depletion Layer Spreading 93

5.6 Conclusions 94

6. Deposition and Characterisation of Silicon Rich Oxide

6.1 Introduction 97

6.2 Application of SRO in MISS Devices 97

6.3 The Structure of SiO_x 98

6.4 Deposition of SRO 99

6.5 Determination of SRO Thickness and Composition 100

6.6 Annealing of SRO films 102

6.7 Electrical Characterisation of SRO MIS Diodes 102

6.7.1 Current-Voltage Characteristics 102

6.7.2 Capacitance-Voltage Characteristics 103

6.7.3 SRO MISS Devices 103

7. Stable Negative Differential Resistance

7.1 **Introduction; Large Area MISS Devices** 105

7.2 **Measurement of MISS Characteristics** 106

7.2.1 **The First Observation of Stable Negative Differential Resistance** 107

7.2.2 **Stability and Device Dimensions** 108

7.2.3 **Detection of Oscillations using an Oscilloscope.** 109

7.2.4 **Use of a Constant Current Source** 109

7.2.5 **The Effect of Capacitive Loading on NDR Stability** 110

7.3 **Circuit Stability and the Concept of Negative Differential Capacitance** 111

7.3.1 **Measurement of Negative Differential Capacitance** 114

7.4 **The Unstable State** 115

7.4.1 **Frequency of Oscillation** 115

7.4.2 **Oscillations in Small Area Devices** 116

7.4.3 **Very Low Frequency Oscillations** 117

7.4.4 **The Amplitude of Oscillation** 120

7.5 **Stable NDR and the Regenerative Feedback Mechanism** 123

7.6 **The Condition for Stable NDR** 124

7.7 **Conclusions** 125

8. Conclusions

8.1 **Negative Differential Capacitance** 128

8.2 **Applications of the MISS in Microelectronics** 129

8.3 **Limitations of the Epitaxial MISS** 129

8.4 **Lateral Forms of the MISS Device** 130

8.4.1 **MISIM** 130

8.4.2 **The Three-Terminal Lateral M-I-n-p$^+$ Switch** 131

8.5 **Future Trends in MISS Switches** 132

Appendix A 134

Appendix B 139

Appendix C 142

Appendix D 144

CHAPTER ONE

Introduction

1.1 Regenerative Switching Devices

The appearance of the first alloyed junction transistor in 1950 marked the beginning of the modern era of solid state electronics. For the first time, the function of amplification, which is fundamental to most electronic systems, could be achieved using a completely solid device. Transistors consequently superseded the bulky, unreliable and power-hungry vacuum tubes which had served for the previous 40 years. The new devices relied for their operation on the simultaneous transport of particles of both polarities (electrons and holes) through semiconducting material and were consequently termed 'bipolar'.

The arrival of the bipolar junction transistor (BJT) was followed very soon after, in 1952, by the Shockley diode (silicon controlled rectifier or thyristor). This four layer p-n-p-n structure brought the possibility of regenerative switching in the solid state and thus replaced the vacuum thyratron. Regenerative switching devices (RSDs) exhibit two stable states in their current-voltage characteristics due to the interaction of two amplifying mechanisms in the form of a feedback loop. A high impedance 'off'-state is separated from a low impedance 'on'-state by a zone of negative differential resistance (NDR) as shown in figure 1.1. The thyristor is therefore a solid state switch with memory; once switched on, the device remains in that state until its current supply is reduced below the holding current (I_H) indicated in the figure. Furthermore, the voltage at which the device switches (V_S) may be controlled by the application of a small 'gating' current through one of the intermediate layers in its p-n-p-n structure. The obvious application of such an effect is the control of the transmission of electrical power to a load such as a motor or light source and the principle market for thyristors is indeed power electronics. Although control of power has become very much the domain of the thyristor, its possible use as a digital memory device has not gone unnoticed. Any device with two stable conductive states is inherently capable of Boolean data retention. Moreover, the memory effect in an RSD is of a static nature. Once the

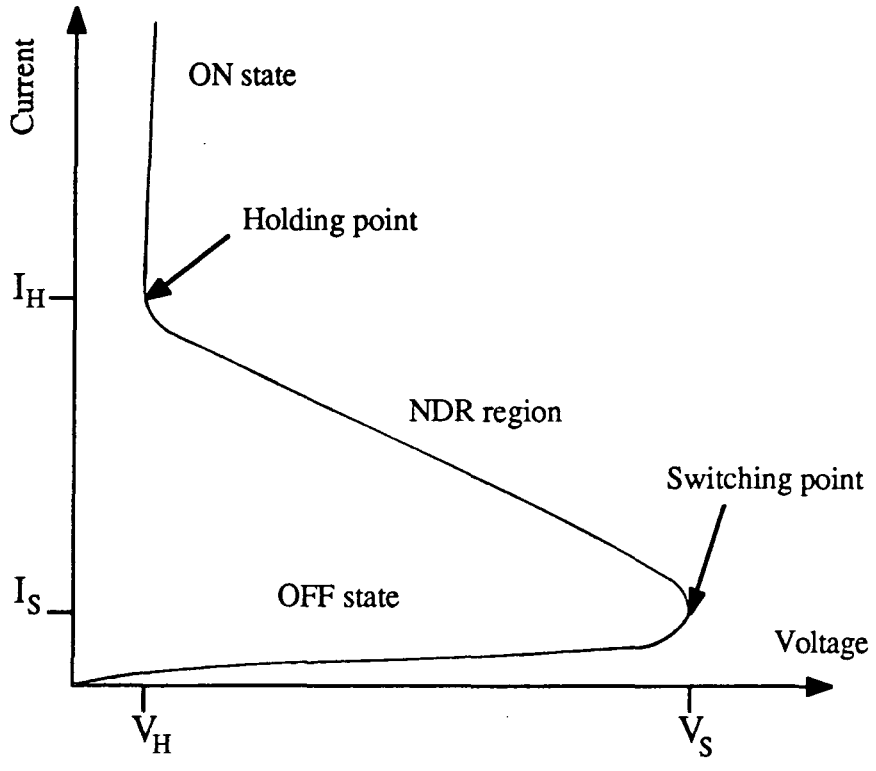


Figure 1.1; The 'S-type' or 'current-controlled' switching characteristic of a Regenerative Switching Device.

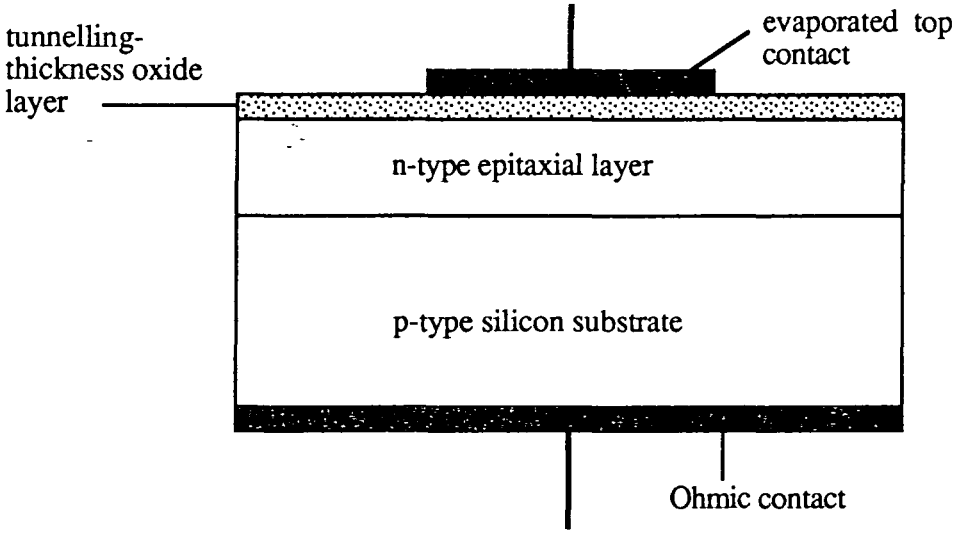


Figure 1.2; The basic structure of a conventional Metal-Insulator-Semiconductor Switch (MISS)

device is established in one or other of its conductive states, it remains in that state (assuming its electrical supply is maintained) without the need to be 'refreshed'. The lack of success of the thyristor as a memory device is in part due to its slow switching speed but perhaps of greater significance is the incompatibility of its four-layer structure with established integrated circuit fabrication techniques.

1.2 The MISS

Twenty years after the development of the thyristor, a new semiconductor device, of much simpler construction, was found to exhibit two states of conduction in the same manner as RSDs. Yamamoto and Morimoto[1] , found, quite by accident, that an epitaxial p-n junction on which they were working displayed a switching characteristic when under forward bias. They attributed this effect to the presence of a thin insulating layer between the top metal contact and the n-type layer. Subsequently the device became known as a 'Metal-Insulator-Semiconductor Switch' or MISS. The structure of a conventional MISS of this type is shown schematically in figure 1.2. At about that time, the discovery that a MIS junction might be capable of current amplification was also creating some interest.[2] However, it was not until four years later that it was suggested that the switching mechanism in the MISS was regenerative.[3]

The MISS was experimentally found to be much faster than the thyristor with turn-on times of a few nS reported[3] [4] . In addition, the simplicity of its construction made it ideally suited to microelectronic applications. For instance, in semiconductor memory, a single MISS (with two MOS transistors for loading and addressing) could replace the six MOS transistors required to form a static random access memory (SRAM) cell[5] Operated in its 'unstable' NDR region by a simple biasing circuit, the MISS may behave as a relaxation oscillator[5] with potentially large dynamic range and very small geometry.[6] Also, being highly sensitive to illumination, it can be turned on by an optical pulse and thus shows promise as an optoelectronic switching device.[7]

Why then, it must be asked, does the MISS remain little more than the subject of scientific curiosity 17 years after its discovery?

In part it might be said that the reason is historical. The BJT and subsequently the Schockley diode were **developed** at Bell Labs between 1945-50 on the basis of sound semiconductor theory by a team who had as a specific goal the realisation of a solid-state amplifier. In contrast, the MISS was **discovered**, with the result that experiment

preceded theory and has continued to do so ever since. Looking back to the late 70's it may be seen that theoretical progress was thwarted by the assertion of El-Badry and Simmons[8] that the interaction between the MIS and pn junctions was based on either punch through of their depletion regions or avalanching in the reverse biased MIS junction, both of which were well known first breakdown phenomena in BJTs. This tended to divert attention from the regenerative switching theories advanced the previous year.

Meanwhile, experimental development continued apace but without the guidance of a sound theoretical understanding. Kroger and Wegener[9] . reported successful use of a variety of semi-insulating layers; SiO_xN_y (15-60Å), amorphous Si (200-2000Å) polycrystalline Si (10^3-10^4Å) and amorphous Ge (40-400Å) as well as tunnelling thickness oxide, SiO_2 (20-45Å) and of several MIS contact metals including chromium, aluminium and molybdenum. In addition, they found no basic assymetry in the I-V characteristics of the complimentary M-I-n-p and M-I-p-n structures nor did it appear to matter whether the pn junction was epitaxial or diffused. Thus, by 1978 the **generality** of the switching phenomenon had been established.

Since then, a large amount of experimental data, often conflicting, has accumulated but without a sound theory to explain it. More recently, concepts of gain and regenerative feedback have enjoyed something of a revival in a series of review papers[10–12] . Meanwhile, however, silicon MOS technology has continued to advance at a remarkable pace. The single transistor RAM cell was developed in 1976[13] and 1 Megabit memory chips are now feasible, obviating any benefits the MISS might have brought in 1972. Most other proposed functions of the MISS, such as in oscillators and charge-transfer type devices, may also now be performed with speed and efficiency by MOS transistor circuitry.

It is against this historical background that this work was embarked upon in 1983. Although by no means all the questions have been answered, it is hoped that this thesis, by means of informed theoretical and experimental enquiry, goes some way towards resolving the remaining confusion which surrounds the MISS and points the way to future developments.

1.3 Thesis Outline

In view of the stated generality of the switching phenomenon in MISS devices, it is intended to keep the discussion as general as possible. However, the use of tunnel oxide I-layers has been predominant in this, as in past studies. Tunnelling is perhaps the best understood of the possible mechanisms of conduction through insulating films and hence much of the theoretical discussion will concentrate on this MIS system. Despite this, it should hopefully remain apparent that the switching processes described do not depend on any particular conduction regime.

A first-order theoretical treatment of the MIS diode is given in chapter 2 followed by an experimental study of the tunnel oxide MIS. Chapter 4 then deals with the theoretical models of the MISS. Starting from a well established steady state description of the device, a more complete regenerative feedback model is proposed in which it is suggested that the MIS diode itself behaves as a bipolar transistor when switching occurs. This new approach conceptually brings the MISS into the fold of RSDs, alongside the thyristor and hopefully removes any remaining doubts about the nature of the switching mechanism.

In chapter 5, an experimental study of the epitaxial MISS is reported with a view to assessing the potential for scaling down the device for microelectronic applications. It is concluded that the epitaxial MISS is fundamentally unsuited to implementation in very large-scale integrated circuits and that tunnel oxide is unlikely to offer either reproducibility or stability in manufactured devices. An alternative semi-insulating material, silicon-rich oxide (SRO) is introduced in chapter 6. This material has already found widespread use in integrated circuits both as a passivation layer for high voltage bipolar I.C.s[14] and as a charge injection layer for electrically-alterable memories[15] but at the time of these studies, its application to MISS devices was novel.

Some remarkable experimental observations made in the course of assessing SRO layers for MISS devices are presented in chapter 7. Although these discoveries seem at first to conflict with the regenerative switching model previously espoused, the introduction of a new concept, negative differential capacitance, together with a consideration of frequency limitations of MISS devices lead to a successful resolution.

Finally, by way of a conclusion in chapter 8, new lateral forms of the device which are more efficient and more amenable to integration are proposed with the support of preliminary experimental results.

REFERENCES

1. T.Yamamoto and M.Morimoto, Appl. Phys. Letters **20** 2269 (1972)
2. R.A.Clarke and J.Shewchun, Solid State Electronics **14** 957 (1971)
3. T.Yamamoto, K.Kawamura and H.A.Shimizu, Solid State Electronics **19** 701 (1976)
4. H.Kroger and H.A.R.Wegener, Solid State Electronics **21** 655 (1978)
5. J.G.Simmons and A.El-Badry, *The Radio and Electronic Engineer* **48** 215 (1978)
6. A.G.Nassibian, R.B.Calligaro and J.G.Simmons, Solid State Electronics **22** 149 (1978)
7. A.El-Badry and J.G.Simmons, Solid State Electronics **20** 963 (1977)
8. H.Kroger and H.A.R.Wegener, Solid State Electronics **21** 643 (1978)
9. I.Zólomy, Solid State Electronics **28** 537 (1985)
10. K.Board, Phys. Rev. D: Appl. Phys **12** 1595 (1985)
11. J.G.Simmons and G.W.Taylor, Solid State Electronics **29** 287 (1986)
12. C.N.Ahlquist, J.R.Breivogel, J.L.McCollum, W.G.Oldham and A.L.Renninger, IEEE J. Solid-State Ccts **SC-11** 570 (1976)
13. T.Matsushita, N.Oh-uchi, H.Hayashi and H.Yamoto, Appl. Phys. Letters **35** 549 (1979)
14. D.J.DiMaria and D.W.Dong, J. Appl. Phys. **51** 2722 (1980)

CHAPTER TWO

Theory of the Non-Equilibrium MIS Diode

2.1 Introduction

Several variants of the MIS switching device have been reported, as discussed in the introduction to this thesis. A feature common to all these structures is the reverse biased MIS diode where the 'I' represents some form of thin semi-insulating layer. When the I-layer is sufficiently conductive, communication becomes possible between the sea of electrons up to the Fermi level in the metal and the conduction and valence bands in the semiconductor. Net fluxes of electrons and holes therefore become possible when the device is under bias and the concentration of minority carriers in the MIS depletion layer may depart from its thermal equilibrium value. It will become clear that it is this ability of the MIS diode to depart from thermodynamic equilibrium which forms the basis of the switching phenomenon in MISS devices. For tunnelling-thickness oxide layers the transition from equilibrium to 'disequilibrium' behaviour occurs for thicknesses below about 60Å, at which quantum mechanical tunnelling becomes significant. Polysilicon, SIPOS, SRO and SnO_x are examples of materials that exhibit sizeable conduction and thus provide for non-equilibrium MIS diodes with I-layer thicknesses in the range 200 to 2000Å. In order to appreciate the importance of deviations from thermodynamic equilibrium, an equilibrium MIS diode with a non-transparent insulating layer (a MIS capacitor) will first be considered.

2.2 Equilibrium MIS Diodes

Silicon dioxide layers thicker than a few hundred Ångstroms present an almost infinite impedance to current flow and are fundamental to gate control in MOS transistors (MOST). Due to the immense importance of the MOST to the world microelectronics industry, the metal-oxide-semiconductor structure has become perhaps the most widely researched and technologically developed of all solid state devices.

2.2.1 The MIS Diode Under Zero Bias.

The energy band diagram of a MIS diode with a non-conductive I-layer (not necessarily oxide) and n-type semiconductor is shown in figure 2.1. In general, at zero bias the bands in the semiconductor will not be flat, figure 2.1(a), due to the combination of metal-semiconductor work function difference and the electrostatic effect of charges existing both at the insulator-semiconductor (I-S) interface and within the insulator itself.

The situation can be described by;

$$V_{FB} = \phi_m - \phi_S - \frac{Q_{ss}}{C_{ins}} - \frac{1}{C_{ins}} \cdot \int_0^{d_{ins}} \rho . dx \tag{2.1}$$

where V_{FB} is the voltage required to establish the flat band condition shown in figure 2.1(b). ϕ_m and ϕ_s are the metal and semiconductor work functions, Q_{ss} the density of interface trapped charge and ρ the bulk density of trapped charge in the oxide.

2.2.2 MIS Diodes Under Reverse Bias.

A MIS diode may be considered to b under reverse bias when a potential difference is applied across its terminals in the sense which causes repulsion of majority carriers from the semiconductor surface. As such, a negative bias would have to be applied to the metal of a M-I-Si(n) structure to achieve this condition.

The effect of a reverse bias is then to deplete the semiconductor surface of majority carriers, leaving a space charge layer of uncompensated donor ions as shown in figure 2.1(c). As a consequence, the semiconductor bands are bent in this region by an amount $\psi(x)$ which is given by Poissons equation;

$$\frac{\delta^2 \psi(x)}{\delta x^2} = -\frac{\rho(x)}{\epsilon_s \epsilon_o} = -\frac{q}{\epsilon_s \epsilon_o}\left(N_D{}^+ + p(x) - n(x)\right) \tag{2.2}$$

Here, $N_D{}^+$ is the concentration of fixed, positively charged ionised donors in the depleted region which, at room temperature, may be considered as simply equal to the doping concentration, N_D. $p(x)$ and $n(x)$ represent the volume densities of free holes and electrons as functions of distance in the depletion layer. As the I-layer prevents any

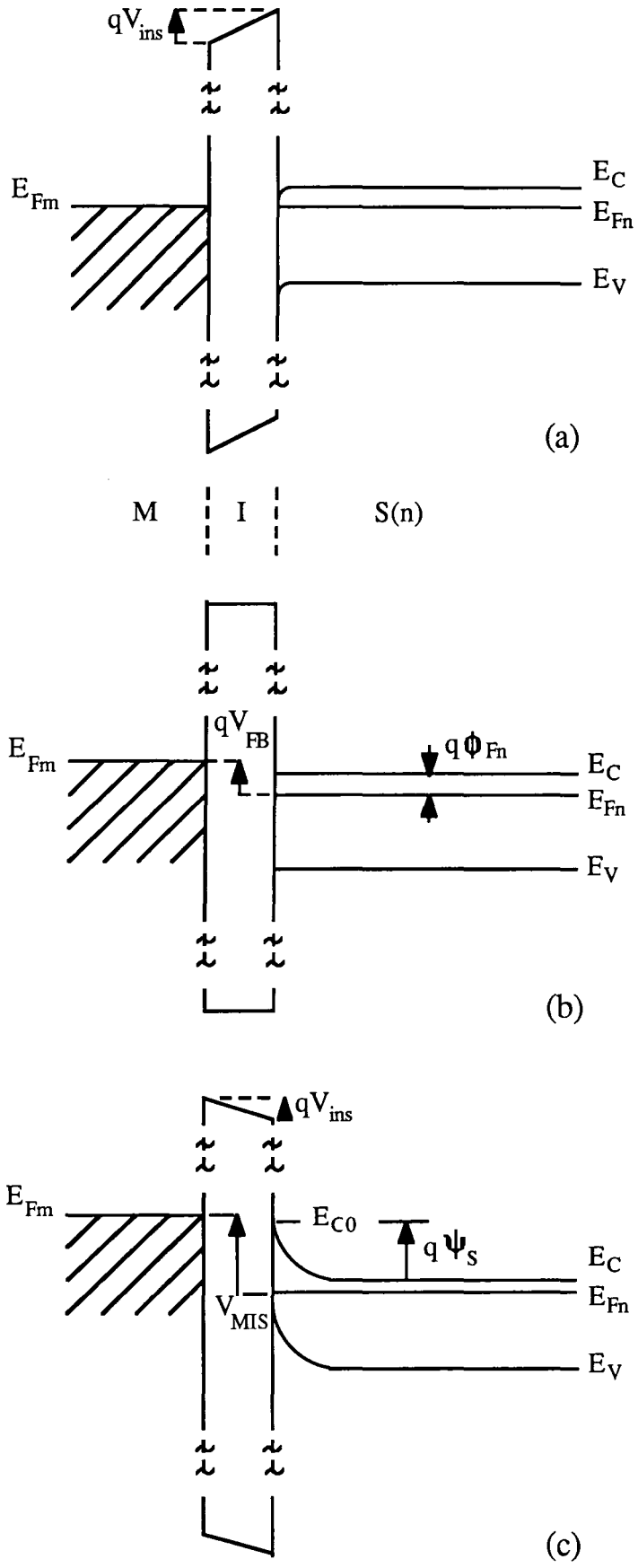


Figure 2.1; Band diagram of a MIS(n) diode under conditions of (a) Zero bias, (b) Flat bands and (c) Reverse bias.

steady state current flow, the carriers remain in thermal equilibrium with the lattice. Their quasi Fermi levels thus remain equal and identical to the Fermi level of the neutral n-type bulk, E_{Fn}, which lies flat throughout the surface depletion region. In such a case, the electron and hole densities in the depletion layer are given respectively by;

$$n(x) = n_i.\exp\left(\frac{E_{Fn} - E_i(x)}{kT}\right) = N_D.\exp\left(\frac{\psi(x)}{V_T}\right) \tag{2.3}$$

$$p(x) = n_i.\exp\left(\frac{E_i(x) - E_{Fn}}{kT}\right) = \frac{{n_i}^2}{N_D}.\exp\left(\frac{-\psi(x)}{V_T}\right) \tag{2.4}$$

where $\psi(x)$ is negative and $V_T \equiv kT/q \approx 25\text{mV}$ at room temperature.

In n-type material at room temperature, ${n_i}^2/N_D \ll N_D$ such that, prior to the onset of surface inversion, $p(x) \ll N_D$. Also, for values of $\psi(x) > 3.V_T$, $n(x) \rightarrow 0$ which allows expression (2.2) to be simplified to;

$$\frac{\partial^2\psi(x)}{\partial x^2} = -\frac{q}{\epsilon_s.\epsilon_o}.N_D \tag{2.5}$$

This corresponds to the so-called 'depletion approximation' in which the distribution of ionised donors is assumed to be a step-function, their density falling abruptly to zero at the depletion edge defined as $x = W_S$.

The total potential across the MIS structure, V_{MIS}, is the sum of the drops across the insulator and surface depletion region plus the correction for the flat band voltage;

$$V_{MIS} = \psi_S + V_{ins} + V_{FB} \tag{2.6}$$

ψ_S is defined as the surface potential and represents the total band bending at the semiconductor surface. It can be obtained by integration of (2.5), taking as boundary conditions the limits $\psi(0) = \psi_S$ and $\psi(W_S) = 0$;

$$\psi_S = \frac{q}{2.\epsilon_s.\epsilon_o}.N_D.{W_S}^2 \tag{2.7}$$

V_{ins}, the voltage dropped across the insulator, is related to the field at the semiconductor surface, $\mathcal{E}_s$, by the so-called 'charge neutrality equation' which is an expression of

continuity of the displacement vector at the I-S interface;

$$\epsilon_{ins}.\epsilon_o.\frac{V_{ins}}{d_{ins}} = \epsilon_s.\epsilon_o.\mathcal{E}_s - Q_{ss} \tag{2.8}$$

An approximate expression for the surface field, $\mathcal{E}_s$, is derived in Appendix B. In the absence of any inversion charge, this becomes simply;

$$\mathcal{E}_s = -\left(\frac{2kT.N_D}{\epsilon_s\epsilon_o}\right)^{1/2}.\left(\frac{\psi_S}{V_T}\right)^{1/2} \tag{2.9}$$

for a MIS diode in depletion.

2.2.3 Inversion in MIS Diodes.

As the applied potential across a reverse biased MIS diode is raised, ψ_S will increase and the depletion region will continue to extend until the onset of strong inversion. This is considered to occur when the concentration of minority carriers at the semiconductor surface becomes equal to the majority carrier concentration in the bulk. Thus, substituting $p(0) = n_{n0} = N_D$ into (2.4);

$$\psi_S(\text{strong inversion}) = V_T.\ln\left(\frac{{N_D}^2}{{n_i}^2}\right) = \frac{2}{q}.(E_{Fn} - E_i) = 2.\phi_{Fn} \tag{2.10}$$

The presence of an inversion layer at the surface prevents any further growth of the depletion layer thickness with increasing reverse potential. In inversion, any increase in oxide field is satisfied by an increase in the inversion layer charge, $p(0)$ rather than in the number of exposed donor ions. Thus the surface potential reaches a maximum;

$$\psi_{max} = \psi_S(\text{strong inversion}) = 2.\phi_{Fn} \tag{2.11}$$

and the second term becomes important in the expression for the surface field from Appendix B;

$$\mathcal{E}_s(\text{inversion}) = -\left(\frac{2kT.N_D}{\epsilon_s\epsilon_o}\right)^{1/2}.\left(\frac{\psi_{max}}{V_T} + \frac{p(0)}{N_V}\right)^{1/2} \tag{2.12}$$

In conclusion then, for an equilibrium MIS diode;

(i) No net currents flow and the carriers remain in thermal equilibrium with the semiconductor.

(ii)In the steady state, the semiconductor surface becomes inverted when the applied bias is sufficient to give rise to a surface potential equal to $2\phi_{Fn}$.

(ii) V_{ins}, ψ_S and $p(0)$ are related through the charge neutrality condition.

2.3 The Non-Equilibrium MIS Diode

When the insulating layer in a MIS structure is sufficiently conductive, passage of electrons between the semiconductor and the metal becomes possible and net currents may flow. In the case of a tunnelling thickness oxide on silicon, figure 2.2, electrons making a tunnel transition between the metal and the semiconductor conduction band constitute an electron tunnel current, J_{NT} while those tunnelling into the valence band are equivalent to a hole current, J_{PT} in the reverse direction. It is also possible for electrons to tunnel in either direction between the metal and bound states at the I-S interface. This flux will constitute a net hole or electron current, J_{ST}, depending on the type and energy of the interface states and on the position of the metal Fermi level. Equivalent currents to J_{PT}, J_{NT} and J_{ST} also pass through silicon-rich oxide, the other semi-insulating material of interest in this thesis. Before describing the relevant conduction theories, the electrostatic consequences of leakage through the insulator will first be considered.

2.3.1 Disequilibrium In Reverse Bias.

In a MIS(n) diode, leakage of holes through the insulating layer will cause the hole density in the depletion layer, as given by (2.4), to fall below its thermal equilibrium value[1] . As such, the quasi Fermi level, E_{Fp}, which describes the hole density will become separated from the bulk Fermi level E_{Fn} in the depletion layer. The amount of splitting of the Fermi levels, defined as $\xi = (E_{Fp} - E_{Fn})$ (figure 2.2) is thus a measure of the degree of thermodynamic disequilibrium in this region and is sometimes referred to as the 'disequilibrium potential'[2] .

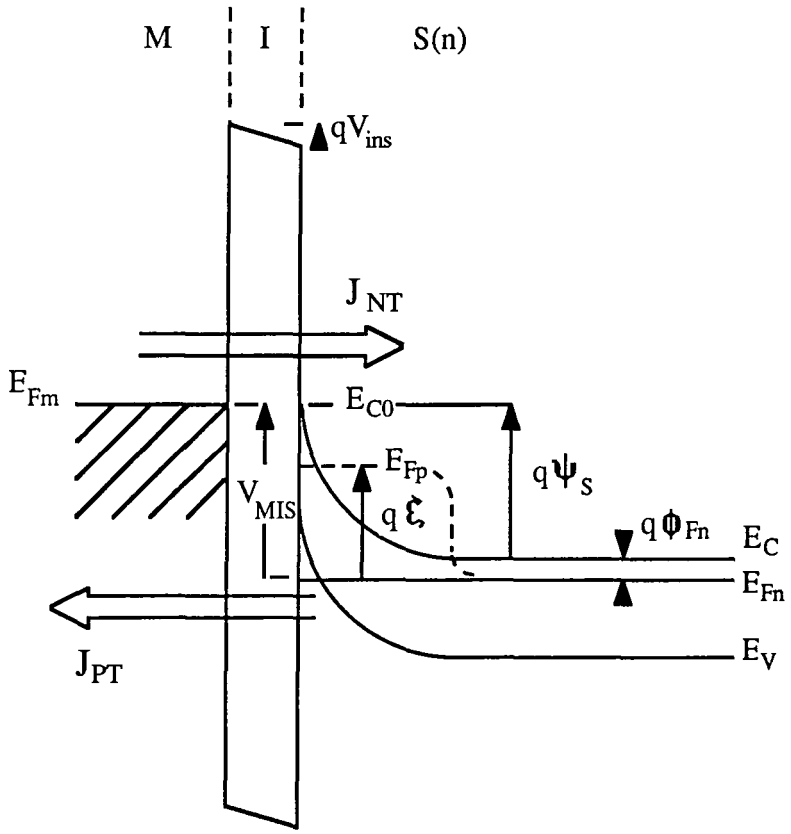


Figure 2.2; Band diagram of a deeply-depleted non-equilibrium MIS(n) diode, showing the band-to-band electron and hole tunnel currents.

The carrier concentrations in the depletion layer become;

$$n(x) = n_i.\exp\left(\frac{E_{Fn} - E_i(x)}{kT}\right) = N_D.\exp\left(\frac{-\psi(x)}{V_T}\right) \tag{2.13}$$

$$p(x) = n_i.\exp\left(\frac{E_i(x) - E_{Fp}}{kT}\right) = \frac{{n_i}^2}{N_D}.\exp\left(\frac{\psi(x) - \xi}{V_T}\right) \tag{2.14}$$

The magnitude of ξ will be a function of the detailed balance between the rates of supply of holes to and from the depletion region. For a MIS diode in isolation, the main sources of holes are thermal and optical generation in the depletion layer. As such the hole current may be termed 'semiconductor limited'. The contribution from the neutral semiconductor bulk will consist of only those holes generated within a diffusion length of the depletion edge ($x = W_S$) which are collected by the surface field before they recombine.

However, any source of extra hole current will alter the detailed balance existing at the MIS junction and, if sufficiently large, may force the condition at the semiconductor surface closer to thermal equilibrium. For example, a MIS(n) diode biased into deep depletion in the dark may be strongly influenced by illumination. This can increase the supply of holes to the surface by generation of carriers in the depletion region and potentially give rise to a degree of surface inversion with a consequent contraction of the depletion region.

Departure from thermal equilibrium allows the reverse biassed MIS to enter a deeply depleted state where ψ_s exceeds the limit normally set by strong inversion, (2.10). Assuming the supply of minority carriers remains limited, the depletion region is able to extend further into the neutral region as the reverse bias is increased. In this case, the growth is eventually limited by either;

(i) breakdown of the insulating layer at high field as, from (2.8), $\mathcal{E}_s$ and hence V_{ins} will also be increasing;

(ii) where N_D exceeds about 10^{16}cm^{-3}, avalanching in the high depletion field near the surface or;

(iii) where N_D exceeds about 10^{18}cm^{-3}, band-to-band tunneling (Zener breakdown) due to the high surface field.

Condition (ii) in fact corresponds to the switching condition in bistable MIS diodes which have been reported by Hayashi[3] and Lai et. al.[4] .

2.4 Generation in the Surface Depletion Layer

In the absence of injection of minority carriers by any other means, the minority current flow in a MIS diode will be principally controlled by generation of holes in the surface depletion region. Generation and recombination occur via traps in semiconductors and both mechanisms are favoured for traps lying at mid-gap, such that their energy $E_t = E_i$. In neutral material in equilibrium, the rate of generation is identical to the rate of recombination. However, where a field exists such as in a depletion region, generated carriers are swept away before they can recombine and a net generation current arises.

From classical Schockley-Read-Hall (SRH) statistics the rate of recombination/ generation is given by;

$$U_{GR} = \frac{1}{\tau_{GR}} . \frac{p.n - {n_i}^2}{p + n + 2n_i} \tag{2.15}$$

Here τ_{GR} is the recombination lifetime defined as $\tau_{GR} = (\sigma.v_{th}.N_t)^{-1}$ where σ is the capture cross section of the traps, v_{th} the thermal velocity and N_t the trap concentration. The generation current I_G, is in general obtained by the integration;

$$I_G = q.\int_0^{\infty} U.dx \tag{2.16}$$

Assuming the generation process occurs uniformly throughout the depletion region, a simple solution for I_G is then;

$$I_G = \frac{q.n_i.W_S}{2.\tau_{GR}} \tag{2.17}$$

However, it has been pointed out[2] that since the recombination process is greatest for traps located at mid-gap, it is probably more realistic to consider only those traps with energy $E_t = E_i$. Thus, in a deeply depleted non-equilibrium MIS diode, nearly all the generation will occur in the zone where E_i lies between E_{Fn} and E_{Fp}, as shown in figure 2.3. As a consequence, I_G should only be evaluated between these limits;

$$I_G = q.\int_{\chi_n}^{\chi_n+\xi} U.\frac{dx}{d\psi}.d\psi \tag{2.18}$$

where $\chi_n = E_g/2q - \phi_{Fn}$. This integral has been solved[2] to obtain an approximate

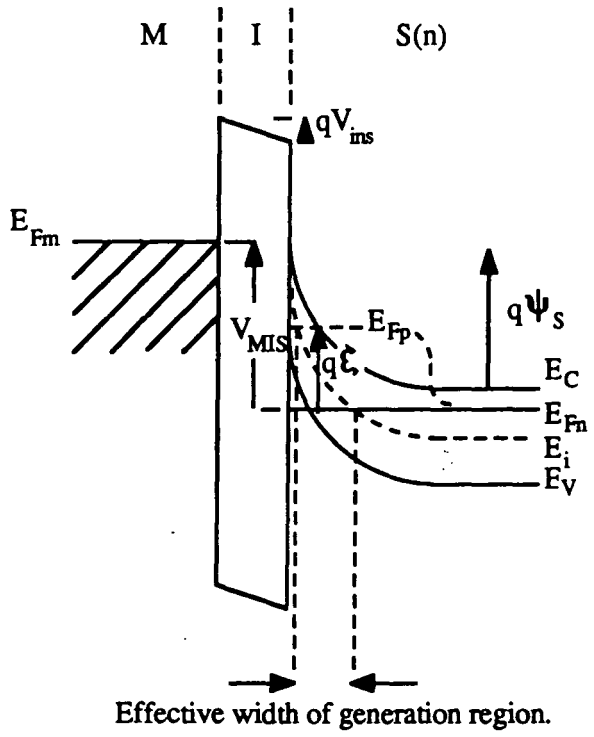


Figure 2.3 ; The effective width of the generation region in the surface depletion layer which is a function of the degree of disequilibrium, as represented by ξ.

expression;

$$I_G = q.\frac{n_i.L_D}{\tau_{GR}}.\frac{1-\exp(-\xi/V_T)}{1+\exp(-\xi/2V_T)}.\left[min\left(\frac{\psi_S}{V_T}-1,\frac{\chi_n}{V_T}+\frac{\xi}{V_T}-1\right)^{1/2}-\left(\frac{\chi_n}{V_T}-1-\ln 2\right)^{1/2}\right] \tag{2.19}$$

where *min* implies the minimum of the two arguments applies.

As a consequence, the generation current will in general be less than would arise if the generation process was occurring uniformly throughout the depletion region, as given by expression (2.17).

2.5 Tunnelling Currents Through Thin Oxides.

2.5.1 Introduction.

Initial studies of the nature of tunnelling in metal-tunnel oxide-semiconductor diodes were undertaken in the early seventies[5-8] when there was some interest in the current multiplication effect such junctions may provide. It was appreciated early in the development of the theory that tunnelling between the metal and interface states could not be neglected in relation to direct tunnelling to the semiconductor bands[5] . The first order theory to be presented here is essentially that of Green and Shewchun[5] and is subject to a number of approximations;

(i) The energy barrier presented by the thin oxide is considered to be trapezoidal, as represented in figure 2.2. Image force effects[9] and non-abrupt interfaces, which tend to reduce the barrier height at the edges by rounding the sharp changes in potential, are neglected.

(ii) The tunnelling probability is assumed to be independent of the field in the insulator and thus of the shape of the barrier. In most theoretical works, it is ascribed a constant value corresponding to that of an ideal rectangular barrier with zero field.

(iii) Electrons tunnel independently so there is no interaction between their wavefunctions.

(iv) The WKB approximation for the tunnelling transmission coefficient is invoked.

(v) Reflections of electrons at the insulator-semiconductor and insulator-metal interfaces are neglected.

2.5.2. Electron Tunnel Current.

The net flux of electrons from the metal to the semiconductor conduction band, as derived by Green and Shewchun[5] , consists of the forward flux, $J_{m\rightarrow c}$, less that in the reverse sense, $J_{c\rightarrow m}$ giving;

$$I_{NT} = A_e.T^2.\exp(-\chi_e{}^{1/2}.d_{ins}).\left[F_1\left(\frac{E_{Fn}-E_{C0}}{kT}\right) - F_1\left(\frac{E_{Fm}-E_{C0}}{kT}\right)\right] \tag{2.20}$$

where

$$A_e = q.\frac{m_{te}.k^2}{2.\pi^2.\hbar^3} \tag{2.21}$$

is the Richardson constant for electrons with m_{te}, the effective electron mass transverse to the insulator barrier. F_1 are Fermi-Dirac integrals of order one and may be approximated by exponentials[5] if the argument is negative

$i.e. F_1(x) \rightarrow \exp(x) \quad \forall\, a < 0.$

The term $\exp(-\chi_e{}^{1/2}.d_{ins})$ represents the tunnel transition probability where $(\chi_e^{1/2}.d_{ins})$ is the tunnel attenuation factor in which χ_e represents the nominal barrier height faced by the electrons.

From the band diagram, figure 2.2, it can be seen that $(E_{Fn}-E_{C0}) = -q.(\psi_S+\phi_{Fn})$ and $(E_{Fm}-E_{C0}) = q(V_{MIS}-V_{FB}-\psi_S-\phi_{Fn})$. Consequently, the expression for the electron tunnel current becomes;

$$I_{NT} = -I_{NT_0}\left[\exp-\left(\frac{\psi_S+\phi_{Fn}}{V_T}\right) - \exp-\left(\frac{\psi_S+\phi_{Fn}-V_{MIS}+V_{FB}}{V_T}\right)\right] \tag{2.22}$$

where $I_{NT_0} = A_e.T^2.\exp(-\chi_e{}^{1/2}.d_{ins})$.

In the case of reasonably large reverse bias, such that $\psi_s \gg V_T$, the first exponential term may be neglected. Then, substituting for V_{MIS} from (2.6), I_{NT} may be approximated by;

$$I_{NT}(\text{rev}) = -I_{NT_0}.\exp\left(\frac{V_{ins}-\phi_{Fn}}{V_T}\right) \tag{2.23}$$

The majority carrier current in reverse bias is thus principally a function of the insulator potential drop V_{ins} which is given by equation (2.8).

Under a forward bias V_{fwd} however, the second term in (2.22) may be neglected for $V_{fwd} > 3.V_T$ and the electron current becomes, in terms of the various device potentials;[6]

$$I_{NT}(\text{fwd}) = -I_{NT_O}.\exp-\left(\frac{\psi_s + \phi_{Fn}}{V_T}\right) \tag{2.24}$$

For the purpose of converting this expression into an empirical diode equation, Card and Rhoderick[6] introduce the thermionic, or Schottky barrier, potential $\phi_{Bn} = \psi_s^0 + \phi_{Fn}$, where ψ_s^0 is the surface potential under zero bias, such that;

$$\psi_s + \phi_{Fn} = \psi_s^0 + \Delta\psi_s + \phi_{Fn} = \phi_{Bn} + \Delta\psi_s \tag{2.25}$$

They also define an ideality factor n which allows the change in surface potential, $\Delta\psi_s$, to be expressed in terms of the total applied potential;

$$n = \left[\frac{\Delta\psi_s}{V_{fwd}}\right]^{-1} = \left[1 + \frac{C_{dep} + C_{ss}^S}{C_{ox} + C_{ss}^M}\right] \tag{2.26}$$

Here, C_{dep} and C_{ox} are respectively the depletion layer and oxide capacitances and $C_{ss}^M = q.N_{ss}^M$ and $C_{ss}^S = q.N_{ss}^S$ are the interface state capacitances corresponding to states in equilibrium with the metal (M) and semiconductor conduction band (S). This relation for n may be determined from the equivalent circuit for the MIS diode shown in figure 2.4. Equation (2.24) may then be reduced to;[6]

$$I_{NT}(\text{fwd}) = -I_{NT_0}.\exp\left(\frac{-\phi_{Bn} + V_{fwd}/n}{V_T}\right) \tag{2.27}$$

2.5.3. Hole Tunnel Current.

The expression for the hole tunnel current is similarly composed of two terms;

$$I_{PT} = A_h.T^2.\exp-\chi_h{}^{1/2}.d_{ins}.\left[F_1\left(\frac{E_{V0} - E_{Fp}}{kT}\right) - F_1\left(\frac{E_{V0} - E_{Fm}}{kT}\right)\right] \tag{2.28}$$

where A_h is the Richardson constant for holes, given by an equivalent expression to (2.21).

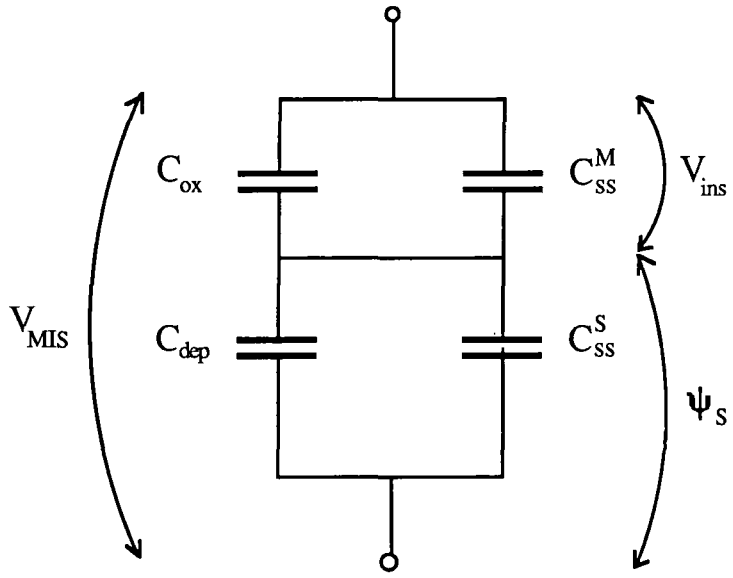


Figure 2.4; Small-signal equivalent circuit of the tunnelling MIS diode showing the partition of the potential across the device.

Noting that the hole concentration at the silicon surface is given by the expression $p(0) = N_V \exp[(E_{V0} - E_{Fp})/kT]$, where N_V is the density of states in the valence band, (2.28) may be stated alternatively as;

$$I_{PT} = I_{PT_0} . \frac{p(0)}{N_V} . \left[1 - \left(\frac{E_{Fp} - E_{Fm}}{kT} \right) \right] \tag{2.29}$$

where $I_{PT_0} = A_h . T^2 . \exp -\chi_h{}^{1/2} . d_{ins}$ As such, the minority carrier current under reverse bias is dependent on the equilibrium concentration of minority carriers at the surface which is in turn a function of the disequilibrium at the surface represented by ξ. The latter is determined by the detailed balance between the rates of their supply to and removal from the surface as discussed in section 2.3.1.

2.5.4. **Tunnelling Via Surface States.**

The electrostatic effect of surface state charge on the band bending in the MIS system has already been discussed in section 2.2.1. Surface states also play a part in charge transfer within the device through two mechanisms which are represented diagramatically in figure 2.5;

(i) They provide a spatially concentrated region of recombination-generation centres which contribute to the total current without involving a tunnel transition.

(ii) They can allow tunnelling transitions between the metal and the semiconductor which constitute a current in parallel with the band-to-band tunnel currents I_{NT} and I_{PT}.

Surface states are categorised as donor or acceptor type depending on whether they are associated with electron or hole trapping. In general, they are characterised by thermal capture cross sections σ_e and σ_h for electrons and holes respectively and by the distribution of their density $N_{ss}(E)$ within the semiconductor energy gap. An electron in a donor trap may tunnel to an empty state in the metal above the Fermi level or an electron may tunnel out of the Fermi sea in the metal and recombine with a hole in an acceptor state. These tunnelling transitions will have associated probabilities, P_T^{ss}.

σ_e and σ_h may take a range of values depending on the oxide thickness and technological parameters such as crystal orientation, surface preparation prior to oxidation

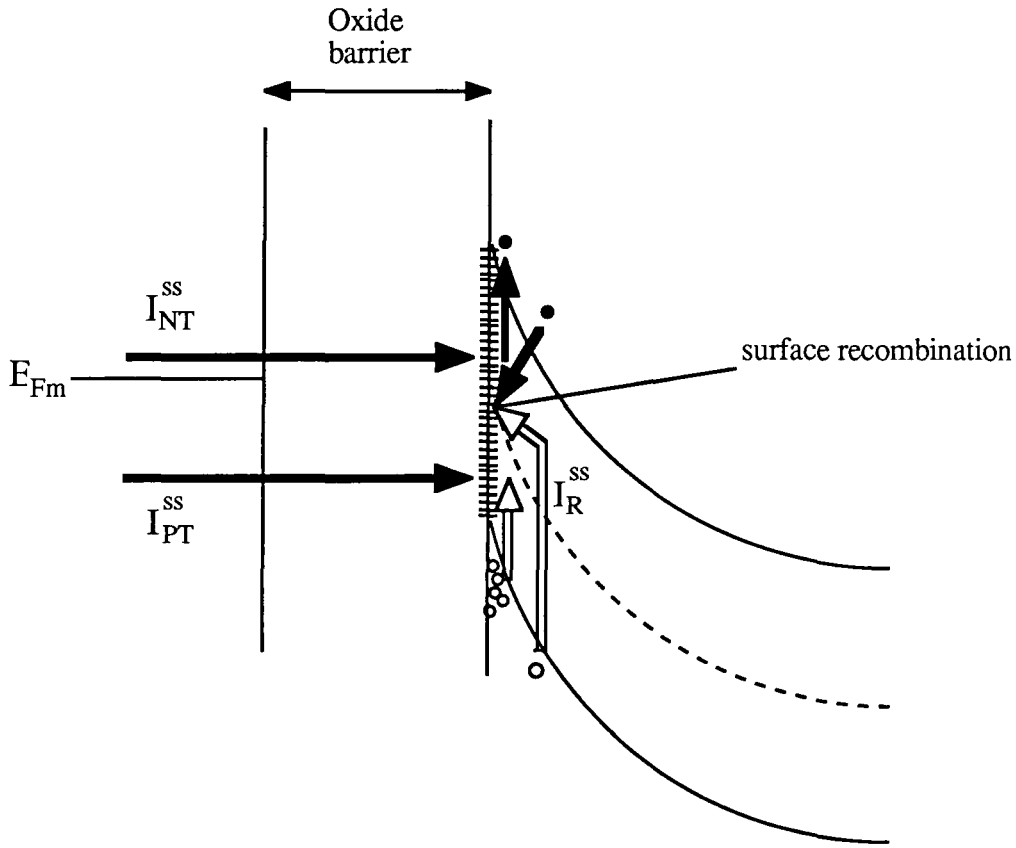


Figure 2.5; Charge transfer mechanisms via the interface involving surface states.

and the type of metal used. They do appear however, to be independent of the surface state energy[10] .

Although the distribution of N_{ss} is continuous, characteristic peaks which appear to be related to the metal used for the top contact have been reported by some workers[11] The magnitude of N_{ss} also depends largely on technological parameters and is found to decrease rapidly with increasing insulator thickness. The latter observation may be interpreted as further evidence that metal diffusion from the top contact does indeed make a contribution.

For the purpose of modelling the process of tunnelling via traps, certain generalisations may be made. For instance, practical MIS diodes are found experimentally to exhibit strong peaks in N_{ss} near the conduction and valence band edges. As such, it is reasonable to treat the distribution as two discrete peaks close to E_C and E_V which correspond to donor and acceptor states respectively.

The tunnelling probability for the traps, P_T^{ss}, may be expressed empirically as;

$$P_T^{ss} \approx P_0 . \exp(-\chi^{1/2} . d_{ins}) \tag{2.30}$$

in which Lundström and Svensson[12] have calculated a value of $10^{13} sec^{-1}$ for the pre-exponential factor P_0.

If f_t and f_m are respectively the occupation functions of a surface state at an energy E_t and in the metal at the same energy, then the tunnel current may be given[13] as;

$$I_{ss} = qN_{ss}P_T^{ss} \int_{E_1}^{E_2} (f_t - f_m) dE_t \tag{2.31}$$

f_t takes the following forms[13] , depending on whether the traps are donors or acceptors;

$$f_t(\text{donor}) = \frac{\sigma_e . v_{th} . n(0) + P_T^{ss} . f_m}{\sigma_e . v_{th} . (n(0) + n_1) + P_T^{ss}} \tag{2.32}$$

$$f_t(\text{acceptor}) = \frac{\sigma_h . v_{th} . p(0) + P_T^{ss} . f_m}{\sigma_h . v_{th} . (p(0) + p_1) + P_T^{ss}} \tag{2.33}$$

The surface carrier concentrations, $n(0)$ and $p(0)$ are given by (2.14) and (2.13) for $x = 0$ and n_1 and p_1 are the concentrations when the corresponding quasi Fermi levels are at the trap energies E_t.

Sarrabayrouse et. al.[14] have solved the set of equations (2.30) to (2.33) for both donor and acceptor traps. Figure 2.6 after reference [14] shows $I_{NT}^{ss} = I_{ss}$(electrons) normalised to $N_{ss}^{D} = N_{ss}$(donors) for three values of oxide thickness. The saturated regions of the curves correspond to a tunnel-limited regime where the donor traps are in equilibrium with the semiconductor. In this regime, the current is only slightly affected by the applied bias but it is exponentially dependent on the insulator thickness. Before the tunnel-limited regime is reached, the current flow via donor states is limited by thermal excitation to the semiconductor conduction band and takes the form[14] ;

$$I_{ss}(\text{electrons}) = \mathrm{I_{NT}^{ss}} = \mathrm{I_{NT_0}^{ss}} \cdot \exp\left(\frac{\mathrm{E_{Fn} - E_{C0}}}{\mathrm{kT}}\right) \cdot \left[\exp\left(\frac{\mathrm{V_{MIS}}}{\mathrm{kT}}\right) - 1\right] \tag{2.34}$$

As such, the surface state-assisted current can not be distinguished from the band-to-band current given by equation (2.22). The same authors also show that the magnitude of the surface state current will in general exceed the band to band current for realistic surface state densities. Figure 2.7 shows the required value of N_{ss}(donor) for $I_{NT}^{ss} > I_{NT}$.

A similar result is obtained for the hole current via acceptor traps where again it is shown that $I_{PT}^{ss} > I_{PT}$ for reasonable values of N_{ss}(acceptor).

2.6 Conclusions

It has been shown that where the insulating layer in a MIS structure is slightly conductive, minority carriers are able to escape from the semiconductor to the metal when the diode is reverse biased. Thus, whereas the surface of an insulating MIS diode becomes inverted when the surface potential reaches a value of $2\phi_{Fn}$, the surface of a leaky diode need not invert and its depletion region may continue to extend as the reverse bias is increased. This condition of deep depletion corresponds to a departure from thermodynamic equilibrium at the surface which is represented by a splitting of the majority and minority carrier quasi-Fermi levels. The density of minority carriers in the depletion region is then determined by their rates of supply from the semiconductor and loss to the metal. A return towards equilibrium at the surface, corresponding to the growth of an inversion layer, may be brought about if the flux of minority carriers to the surface is enhanced by some means. It has been noted in this chapter that such an increase in hole flux may result from avalanche multiplication when the field in the surface depletion region is high. It will be shown in chapter 4 that a forward biased pn

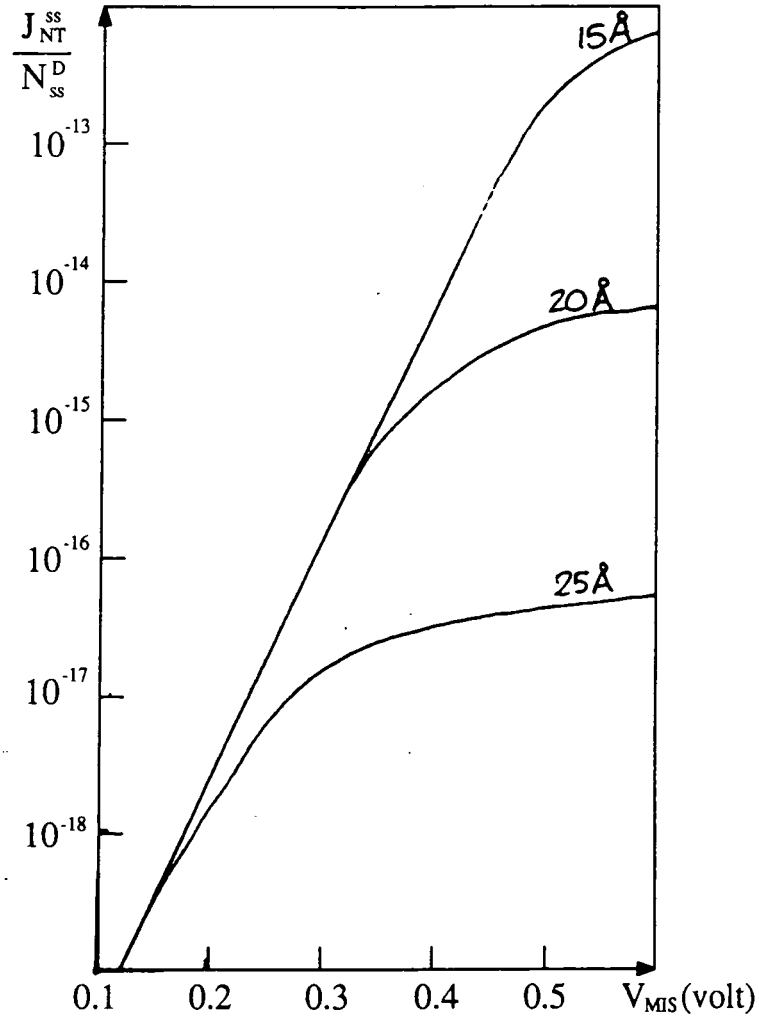


Figure 2.6; The ratio of trap-assisted tunnel current to interface state density calculated as a function of the applied forward bias across a MIS diode for several values of oxide thickness (after reference 14).

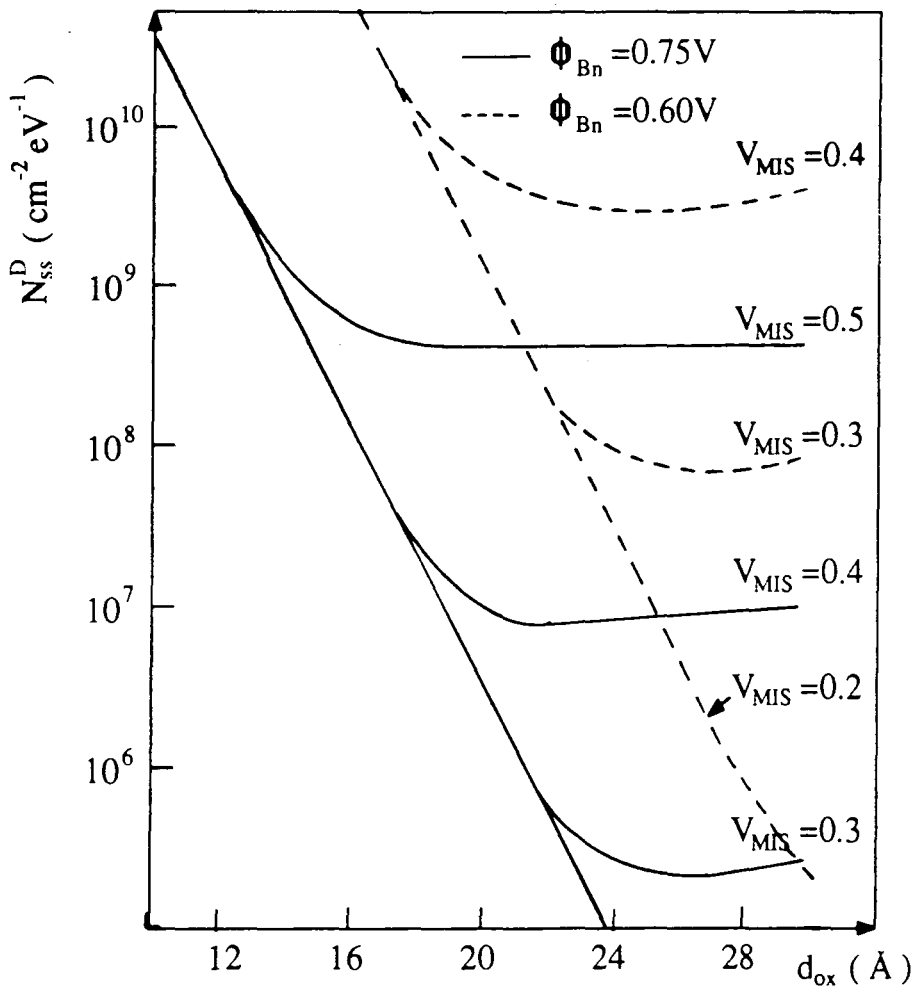


Figure 2.7; Calculated values of interface state density required for trap-assisted tunnelling to exceed band-to-band tunnelling assuming a capture cross-section, $\sigma_n = 10^{-16} cm^2$ (after reference 14).

junction in close proximity to the MIS may be a highly effective source of hole flux and that this combination forms the basis of the MISS device.

Although a first order theory of tunnelling through very thin silica layers has been outlined in this chapter, the precise nature of the conduction mechanism through the insulating layer is not critical to the non-equilibrium behaviour of the MIS diode. In chapter 6 of this thesis, an alternative semi-insulator, 'silicon rich oxide' will be introduced and shown to exhibit essentially the same characteristics. What the particular choice of insulator will affect, however, is the exact relationship between the majority and minority carrier currents and the field in the insulating layer. As a final point, it has been noted that the influence of interface states on the passage of carriers between the semiconductor and the metal may not be neglected.

REFERENCES

1. R.A.Clarke and J.Shewchun, Solid State Electronics **14** 957 (1971)

2. A.C.F.Fiore de Mattos, *These d'Etat No.1158*, Universite Paul Sabatier, Toulouse, France (1984)

3. Y.Hayashi, Appl. Phys. Letters **37** 407 (1980)

4. S.K.Lai, P.V.Dressendorfer, T.P.Ma and R.C.Barker, Appl. Phys. Letters **38** 41 (1981)

5. M.A.Green and J.Shewchun, Solid State Electronics **17** 349 (1974)

6. H.C.Card and F.M.Rhoderick, J.Phys.D; Apll. Phys. **4** 1589 (1971)

7. S.Kar and W.E.Dahlke, Solid State Electronics **15** 221 (1972)

8. M.A.Green, F.D.King and J.Shewchun, Solid State Electronics **17** (1974)

9. A.Tugulea and D.Dascalu, J. Appl. Phys. **56** 2823 (1984)

10. S.Kar and W.E.Dahlke, Solid State Electronics **15** 221 (1972)

11. S.Kar and W.E.Dahlke, Appl. Phys. Letters **18** 401 (1971)

12. I.Lundström and C.Svensson, J. Appl. Phys. **43** (1972)

13. L.B.Freeman and W.E.Dahlke, Solid State Electronics **13** 483 (1970)

14. G.Sarrabayrouse, J.Buxo and D.Esteve, phys. stat. sol. (a) **46** 185 (1978)

CHAPTER THREE

Fabrication and Characterisation of Tunnel Oxide MIS Diodes

3.1 Introduction

In the introductory chapter, it was noted that switching has been observed in MISS devices incorporating a great variety of semi-insulating films. By far the most widely investigated of these has been tunnelling-thickness oxide grown by thermal oxidation. In fact, the MISS represents only a minor example of the technological importance of tunnel oxides in a range of applications. For example, metal tunnel-oxide silicon solar cells provide very high energy conversion efficiency, exceeding 24%[1] and bipolar transistors with MIS emitters exhibit the highest current gain yet achieved, of the order of 25,000[2] . EEPROM static memory transistors which incorporate tunnelling oxides as a means of injecting and removing charge from a floating gate, are of great commercial significance.

The importance of tunnel oxides can be seen to have arisen from a combination of historical and technological factors. Early research in MIS diodes grew out of an interest in the effects of the interfacial oxide which normally exists in Schottky barriers formed on silicon and can provide current gain. In parallel with a growing interest in tunnelling films, the technology and scaling of MOS transistors to sub-micron dimensions is requiring that gate oxides be used with thicknesses of the order of 100Å.

The majority of studies of tunnel oxide MIS diodes have been concerned with aluminium - thermally grown SiO_2 - silicon structures, perhaps because the technologies are readily available. The choice of Al as the top contact metal, although by no means scientifically justified, is also made in the present work. All the tunnel oxide MIS diodes and most of the MISS switching devices reported in this thesis are of quite small dimensions ($10 \times 10\mu m$ to $30 \times 180\mu m$) and have been prepared using a mask set designed and fabricated for this purpose by the author. Device 'active areas' are defined in a layer of thick (0.5 to 1.0 μm) oxide by photolithography and wet etching. Subsequent to tunnel oxide growth, aluminium is evaporated and top contacts and associated tracks and bond pads are defined.

Before describing in some detail the process used for the fabrication of such devices, some consideration is given to the nature of thin oxides and their growth mechanism.

3.2 MIS diodes with tunneling thickness SiO_2

Despite their apparent simplicity, metal- tunnel oxide - silicon structures may be considered to comprise at least six different regions, only three of which may be considered as 'bulk' metal, stoichiometric SiO_2 and silicon. These six regions are represented in figure 3.1 for an Al-SiO_2 diode and may be briefly summarised as follows;

3.2.1. Aluminium Top Contact

An important characteristic of the contact metal is its work function which is a factor in determining the potential barrier faced by tunnelling electrons.

3.2.2. The Aluminium-Oxide Interface

Good adhesion between a metal and SiO_2 normally implies that some form of chemical reaction[3] between the two has occurred. For this reason, gold, which is not reactive with SiO_2 at normal deposition temperatures, is a notably poor adherent to SiO_2 . In contrast, at an aluminium-oxide interface, the following solid state reaction takes place[3] [4] ;

$$4.Al + 3.SiO_2 \Longrightarrow 2.Al_2O_3 + 3.Si \tag{3.1}$$

providing a chemical bond which makes aluminium a good material to use for interconnect metallisation on integrated circuits. Auger electron analysis (AES)[5] and X-ray photoelectron spectroscopy (XPS)[4] studies of this interface have confirmed the presence of Al_2O_3 and free Si precipitates for room temperature deposited films. In the case of tunnel oxides however there is cause for concern because by means of this reaction, Al is able to penetrate a substantial distance into the SiO_2 [3] . Furthermore the penetration of Al_2O_3 has been observed to increase with time, even at room temperature and to extend to a depth of 100Åor more into the SiO_2 layer when annealed at between 300-400°C [6] One consequence of the formation of an Al_2O_3 layer is a reduction in the work function difference, ϕ_{ms} of about 0.15eV[3] but obviously in the case of tunnel thickness oxides, a more important result could be the complete conversion of the SiO_2 into a mixed layer of Al_2O_3 and Si.

3.2.3. Stoichiometric SiO_2

Assuming that the aluminium does not penetrate completely through the oxide layer, a layer of stoichiometric SiO_2 will remain. This layer is considered theoretically to present a potential barrier through which electrons must tunnel and the height of this barrier is assumed to be equivalent to that of bulk SiO_2 . However, oxide layers grown at low temperatures have been shown by EerNisse[7] to be under considerable biaxial compressive stress, of the order of 7.10^8 Pascal. Irene et. al.[8] have accounted for the build up of strain on the basis of the low viscosity of the SiO_2 during growth. At temperatures below about 900°C the viscosity of the growing oxide layer is too great for the 120% increase in free volume required by the transition from crystalline Si to amorphous SiO_2 to be accomodated by flow. As a consequence, oxide layers grown in the temperature range 600-800°C are found to be substantially denser than fully relaxed SiO_2 as well as containing a high level of residual strain[8] . The increased density results in a higher measured value of refractive index (1.48 at 600°C and 1.475 at 700°C) than is expected at more normal growth temperatures (1.462 at 1000°C).

Although it has been shown that the strain may be relaxed by subsequent annealing, this process is very slow at temperatures below 900°C unless H_2O is present.[8] Thus the tunnel oxides in the present work, which were grown in a dry ambient at around 750°C , must be expected to remain in a strained condition. Under these circumstances, it would seem likely that other physical properties of the oxide layer, such as its effective bandgap, will also be modified.

In conclusion it is unlikely that even the stoichiometric component of the MIS 'sub-structure' is treated realistically in the tunnelling model of Chapter Two.

3.2.4. The Oxide-Silicon Interface

In general, this interface is characterised by a transition region between the Si and the stoichiometric SiO_2 and by the presence of traps at or near the interface.

There is some disagreement as to the extent of the transition layer which has an intermediate composition, SiO_x where $0 < x < 2$. Although it is theoretically possible for amorphous SiO_2 to terminate on Si within one bond length[9] , experimental investigations have revealed a non-stoichiometric layer of between 5 and 25Ådepending on the technique used. Thus in the case of tunnel oxides, there is some concern that the transition layer may constitute the greater part of the whole oxide layer[10]

Evidence for a relatively abrupt interface has been obtained by Feldman[11] using He backscattering and Krivanek et. al.[12] using high resolution transmission electron microscopy (HRTEM) both of whom report a transition within about 5Å of the interface. Also, diStefano[13] using field-dependent internal photoemission spectroscopy found insignificant (0.15%) distortion of the oxide conduction band beyond 4Å of the interface, implying that non-stoichiometry is negligible at this distance. Importantly, Krivanek[12] observed with HRTEM both short and long range modulation of the silicon surface. Monatomic steps (3.2Å high on {111} and 1.4Å on {100} surfaces) were seen every 20-40Åas well as more pronounced undulations 3-10Åin height with a period of 200-600Å.

It is significant that most of the evidence for a broad transition region has been gained using ESCA (20Å)[14] and Auger surface analysis with depth profiling (28Å)[15] Both of these techniques average the composition both in-plane and normal to the interface due to the wide beam and the sputtering action used and may thus misinterpret an undulating interface as being non-stoichiometric[15] .

Roughening of the SiO_2 -Si interface has since been correlated with an increase in the density of dangling bonds represented by Si^{+3} or $Si_3 \equiv Si\cdot$ which are thought to be the main cause of interface states in the SiO_2 - Si system[16] . Annealing in N_2 at 1050°C has been shown to reduce surface roughness very considerably[17] and to improve the electrical breakdown properties of the oxide. Dressendorfer et. al. also demonstrated that an anneal in N_2 at 825°C subsequent to a 10 min oxidation at the same temperature reduced the interface state density N_{ss} by a factor of 2 to 3. They also showed that a post-metallization anneal in 10:90 H_2:N_2 forming gas further reduced N_{ss} by a similar factor.

Diffusion of metal from the top contact has also been found to give rise to interface states in MOS systems where the oxide is tunnelling thickness. Kar and Dahlke[18] detected peaks in the interface state distribution characteristic of the particular metals used. This link has not always been corroborated by other workers however.

3.2.5. **The Silicon Surface**

In addition to complexities of strain and non-stoichiometry in the oxide transition layer, disorder has been detected in the surface of the silicon itself. Using He backscattering, Sigmon et. al.[19] found evidence that Si atoms are displaced from their surface

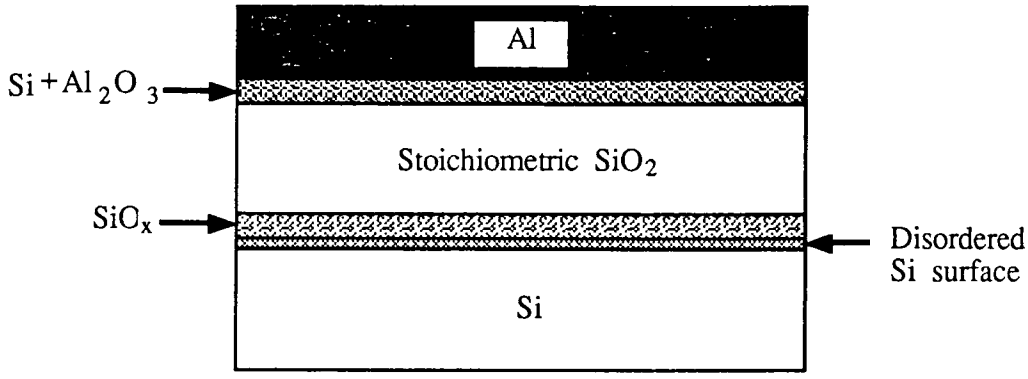


Figure 3.1; Section through an Al-SiO_2-Si MIS diode indicating the various regions discussed in the text.

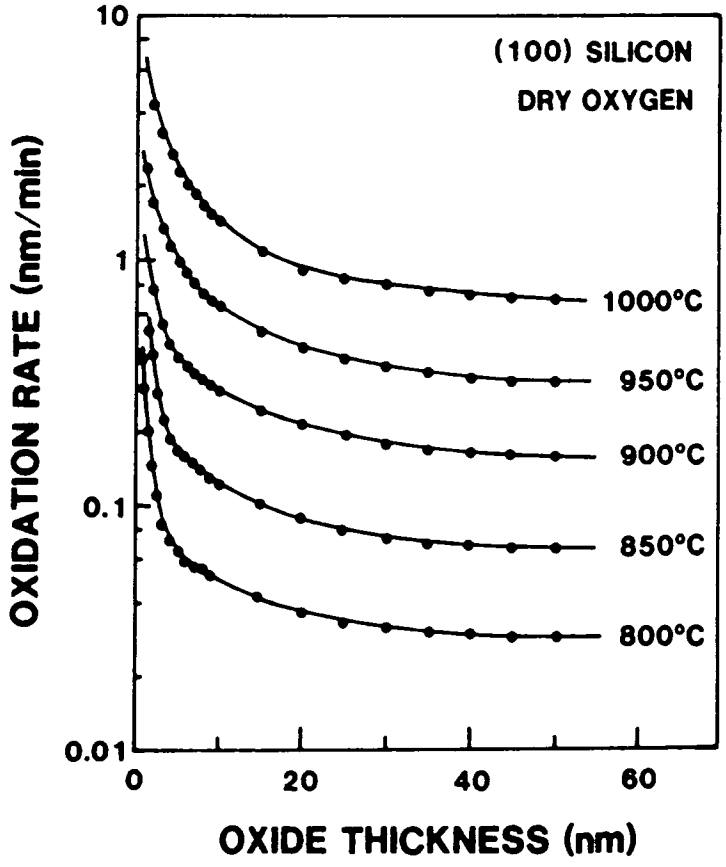


Figure 3.2; Oxidation rate of {100} silicon in dry oxygen after reference [24].

sites to accomodate the structural transition from crystalline to low density amorphous. Disorder, as well as contributing to the density of interface states, may be expected to cause broadening of the trap distribution in energy. Interestingly, the degree of surface disorder has been found to be twice as great for {111} as for {100} oriented silicon[20] in line with the measured values of N_{ss} for these two surfaces.

In conclusion then, it has to be recognized that a tunnel oxide MIS is in reality a far more complex structure than is treated in the theoretical model of chapter 2. It is clearly naïve to treat the potential barrier presented by the oxide as trapezoidal or to neglect the presence of interface states.

3.3 Theory of Thin Oxide Growth

A well established model for the growth of oxides on silicon is that due to Deal and Grove[21] which accounts quite well for the growth of layers thicker than several hundred Ångstroms. In this model, the growth rate is described by a 'linear-parabolic' relationship;

$$d_{ox}^2 + A.d_{ox} = B.(t + \tau) \tag{3.2}$$

in which τ is included to account for the presence of a nascent oxide (of thickness x_i) prior to oxidation and is defined as;

$$\tau = \frac{x_i^2 + A.x_i}{B} \tag{3.3}$$

For thicknesses less than around 200Å, growth is considered to be limited by the rate of surface reaction and thus to be linear with time, independent of the oxide depth. Expression (3.2) then reduces to a linear relationship;

$$d_{ox} = \frac{B}{A}(t + \tau) \tag{3.4}$$

At atmospheric pressure, the linear rate constant is given by;

$$\frac{B}{A} = \frac{k_s.C^*}{N_1} \tag{3.5}$$

where k_s is the temperature-dependent interfacial reaction constant, C^* is the equilibrium concentration of the oxidising species in the oxide and N_1 is the number of oxidant

molecules required per unit volume of oxide formed. For greater thicknesses, diffusion of the oxidising species becomes the rate limiting mechanism and the growth rate becomes parabolic.

Clearly, in the growth of tunnelling thickness oxides the parabolic regime is never reached and it might be expected that growth would proceed linearly at a rate determined only by the temperature, orientation of substrate and oxidising species present. Although C^* is to some extent a function of temperature, the activation energy for the oxidation process, E_A, predominantly reflects the temperature dependence of k_s and relates to the energy required to break a Si-Si bond. As such, the activation energy is also influenced by the crystal orientation and for the principle silicon surfaces used follows a progression $E_A\{100\} > E_A\{111\} > E_A\{110\}$. The factor k_s is also dependent on the crystal orientation of the silicon on the basis of the availabilty of silicon-silicon bonds at the surface, as accounted for by Ligenza[22] . It has since been confirmed experimentally that the growth rate, γ, does indeed follow the sequence $\gamma_{\{110\}} > \gamma_{\{111\}} > \gamma_{\{100\}}$.

However, in practice it is found that the initial growth rate is considerably higher than predicted by the simple linear rate expression. For example, recently Massoud et. al. have measured a substantial excess growth rate in dry O_2 over more than 100Å[23] Figure 3.2 after reference [23] shows measured growth rates on {100} silicon in dry oxygen for a range of temperatures. The excess rate observed, in this case up to about 3 to 4nm, has been found to vary exponentially with thickness. Irene first proposed a model[24] suggesting that channels form in the relatively open structure of the amorphous SiO_2 network which may extend right through to the interface. It is possible then for the oxidising species to 'stream' through such 'micropores' and produce the anomalously high growth rate.

It is becoming increasingly clear though that the initial growth rate is highly sensitive to the preparation of the silicon surface prior to oxidation. Early work by Raider and Flitsch[25] showed that a freshly cleaved silicon surface oxidised more rapidly than a sample etched in HF. Such a result is perhaps not surprising since cleaving reveals a highly reactive free surface to the ambient. More recently, Schwettman et. al.[26] found that pre-oxidation treatment of silicon with NH_4OH-H_2O_2 caused a reduction in growth rate but treatment with H_2SO_4-H_2O_2 caused an enhancement. Grunthaner and Maserjian[27] first showed that a 'real' silicon surface (i.e. one that has not been prepared *in vacuo*) oxidises more quickly after an etch in HF. The acid is presumed to

have the effect of removing the existing layer of nascent (atmospheric) oxide which would normally inhibit growth. Interestingly, in a more detailed study of surface preparation, Gould and Irene[28] have found the growth rate to be affected by the condition of the surface to thicknesses of at least 275nm, which is beyond the limits of the linear rate regime.

In conclusion then, it is clear that the growth of ultra-thin oxides on silicon is critically dependent on the nature of the silicon surface and of the oxide layer existent prior to controlled thermal oxidation.

3.4 Fabrication of Tunnel Oxide MIS Diodes.

The process adopted at Durham for the growth of ultra thin oxides consists of thermal oxidation in pure dry O_2 at a nominal growth temperature of 750°C . The single walled oxidation tube used is of stabilized fused quartz, 6 cm in diameter and 1.5 metre long and is shrouded in a 'Purox' alumina tube, figure 3.3. This furnace is dedicated to the growth of dry oxides (both tunnel thickness and for MOS gates) and is constantly purged with nitrogen. Both the nitrogen and oxygen supplies are obtained from pressurised cylinders and dried by passage through columns containing calcium alumino-silicate molecular sieve. Unfortunately, the three-zone furnace had to be used with the upstream zone not functioning and thus provided a flat zone (±1°C) of only about 10 cm as indicated in figure 3.3.

In view of the great importance of surface preparation to the subsequent growth process, as discussed in the previous section, a meticulously clean and controlled procedure has been established. All equipments used for processing are dedicated to a particular stage in the process and only MOS grade chemicals are used.

3.4.1. Cleaning of Substrates

As-received wafers would be expected to have contaminated surfaces due to adsorption of impurities from the air, particularly organics, and due to handling. The first cleaning step is therefore de-greasing in boiling 1,1,1-trichloroethane for five minutes which is repeated for a further five minutes in fresh solvent. Subsequently, substrates are etched for 20 minutes in a boiling 1:1 mixture of H_2SO_4:H_2O_2 which is freshly prepared on each occasion and then they are washed in re-circulating de-ionised water (d.i. H_2O) for at least an hour.

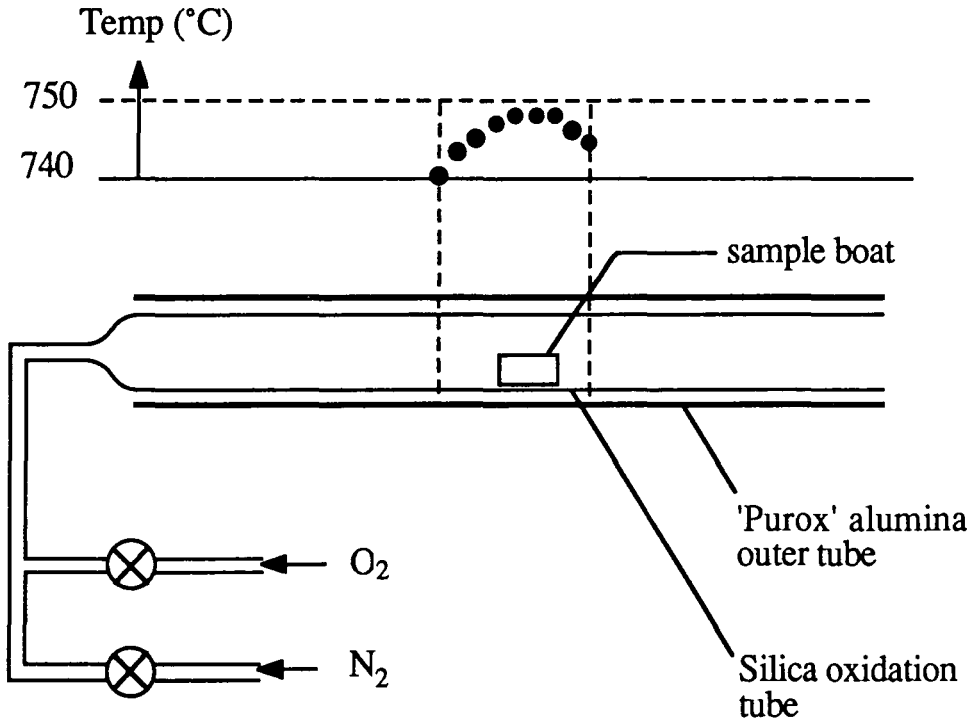


Figure 3.3; Schematic diagram of the dry oxidation furnace used for the growth of tunnelling-thickness oxides.

Sample position	Time (mins)	Gas admitted	Flow (l/min)
Mouth	10	dry N_2	1.0
Flat zone	10	dry N_2	1.0
Flat zone	10	dry O_2	1.0
Flat zone	Growth time	dry N_2 + wet O_2	0.5 + 0.7
Flat zone	10	dry O_2	1.0
Flat zone	10	dry N_2	1.0
Mouth	10	dry N_2	1.0

Table 3.1

Sequence for growth of field oxides by wet oxidation.

3.4.2 Growth of Field Oxide

Prior to growth of field oxide, substrates are taken from the wash bath and etched in buffered HF (a 4:1 mixture of NH_4OH: HF) until they become hydrophobic which indicates that any surface oxide has been removed. They are then rinsed and left in recirculating d.i.H_2O for 20 minutes. On removal from the bath, the substrate surfaces are no longer hydrophobic, which implies that a nascent oxide layer is present. Nascent oxides are reported to be of the order of 10-20Å in thickness[15] and to be of mixed composition, with a considerable carbon content. However, the presence of a thin oxide is tolerated prior to growth of the less critical thick field oxide. Growth of thick oxides is achieved using a 'wet' oxidation process at around 980°C. In this process, the oxidation tube is supplied with oxygen wetted by bubbling through a temperature-controlled bath of d.i.H_2O . The procedure is outlined in Table 3.1. A 240 minute oxidation provides an oxide approximately 10,000Å thick, as determined by its colour.

3.4.3. Patterning of the MIS Windows

The windows in which the tunnel oxide is to be grown are defined photolithographically using mask 4 of the set described in Appendix C. After hard baking the resist, the pattern is etched using buffered HF and the resist removed with a propietary stripping agent. The sample is then washed for an hour in recirculating d.i.H_2O .

Sample position	Time (mins)	Gas admitted	Flow (l/min)
Mouth	10	N_2	1.0
Flat zone	10	N_2	1.0
Flat zone	Growth time	O_2	1.0
Flat zone	10	N_2	1.0
Mouth	10	N_2	1.0

Table 3.2

Sequence for growth of tunnel oxides
by dry oxidation

3.4.4. **Growth of Tunnel Oxides**

Just prior to loading a sample into the mouth of the oxidation tube, it is dipped briefly in a 10% solution of HF to remove any oxide grown during the washing procedure. After this it is rinsed several times in free flowing d.i. H_2O , dipped in an ultrasonic bath of high purity grade IPA and blown dry with filtered dry nitrogen. It is then laid face up on a silica boat which is loaded into the mouth of the furnace tube. Although delay between removal from the final HF etch and loading into the dry ambient of the furnace mouth is minimised, it must still be accepted that growth of a nascent oxide, which is spontaneous in air, is inevitable.

The procedure after loading is listed in table 3.2. Samples are retained in the mouth for a period of time to ensure complete desorption of surface moisture. Then they are pushed to the centre of the flat zone at a rate of 0.3 cm/second where they are left to thermalize for 10 minutes.

The oxidation phase is initiated by simultaneously decreasing the N_2 flow rate to zero and raising the O_2 flow to 1.0 litre/minute and it is terminated by reversing the procedure. As such the oxidation period is not determined precisely. Due to the large 'dead space' delay, oxidation must commence some time, t_1 after admitting oxygen and continue for a similar period, t_2 after restoring the nitrogen flow. If a 'gas packet' model may be assumed in which the volume of O_2 admitted during the growth period constitutes a moving column of gas with minimal interdiffusion with the surrounding N_2, then the two delays t_1 and t_2 may be assumed to be equal. Knowing the flow rate (35 cm/min) and the volume of the dead space, t_1 and t_2 may be estimated to be about

two minutes.

Intermixing of the gases is not negligible however[29] , especially when the flow velocity is low as in the present case. Thus at the beginning and end of the growth cycle, oxidation must occur in an ambient varying in composition between 100%O_2 and 100%N_2. The sample remains in the hot zone for 10 minutes after stopping the O_2 flow to allow a 100% N_2 ambient to be restored and is then pulled slowly to the furnace mouth.

On removal from the furnace, samples are quickly loaded into the aluminium evaporator which is then evacuated. On some occasions, a bare silicon control wafer is included in the growth run to provide a sample for ellipsometric evaluation.

3.4.5 Evaporation and Patterning of the Aluminium Contacts

In all the devices reported in this thesis, the top contact metal is pure aluminium deposited by electron beam evaporation of a 99.999% purity source. Samples are mechanically clamped by one edge to suitable jigs and then loaded into the evaporator face down so as to be exposed to the source. The evaporator is pumped by a completely oil-free vacuum system in the following sequence;

(i) diaphragm pump to approximately 10 Torr.

(ii) first sorption pump to below 1 Torr.

(iii) second sorption pump to 10 millTorr and below.

(iv) ion pumped to 10^{-6} Torr.

The base pressure is further improved to 2.10^{-7} Torr by degassing the aluminium source with several short bursts of e-beam current with the shutter closed.

Aluminium is deposited onto thin oxides in four or five intermittent bursts of about 7 seconds each with at least 15 minutes delay for cooling between them. The intention is to avoid excessive heating of the Al-SiO_2 interface and thus minimise the surface reaction, as was discussed in section 3.2.2. Although the sample is not deliberately heated (each phase of the evaporation commences at room temperature), neither is it cooled. As such, there is some concern that the latent heat of evaporation released by the deposited aluminium will significantly raise the temperature of the sample surface. However, no means were available to assess whether penetration of the surface oxide by aluminium was occurring.

3.4.6 Back Contact Metallisation

All the MIS diodes fabricated have been on n-type silicon with low doping of the order of 2.10^{15} cm^{-3}. Ohmic contacts to such substrates are not so easily achieved as they are to p-type or heavily doped n-type material. However, contacts of reasonably low resistance are achieved at Durham using evaporated gold doped with antimony. The antimony assists in the formation of an electrical contact by diffusing slightly into the silicon surface, making it strongly n-type. Although the Au-Si(n^+) contact so formed does have a Schottky barrier, the surface depletion layer is so thin that it breaks down due to tunnelling and near ohmic conduction results.

3.5 Determination of Tunnel Oxide Thickness

3.5.1 Measurement of Ultra-Thin Oxides by Ellipsometry

Despite the claims of several authors that the thickness of tunnel oxides may be determined to a fraction of an Ångstrom by ellipsometry[30,31] , it is doubtful whether such a degree of accuracy can really be expected using this technique. The principles of ellipsometry are outlined in Appendix D where a plot of the (refractive index, thickness), (n_r, d_{ins}) locii in the plane of the ellipsometric angles (ψ,Δ) is shown. By inspection of these locii it can be seen that the best thickness resolution is achieved for SiO_2 (assuming n = 1.46) for values close to 140Å and 420Å. For tunnelling-thickness oxide, both n_r and d_{ins} are very strong functions of ψ and Δ and even small errors in determining these angles will lead to large error in the value of d_{ins} obtained. Certainly, using the home built ellipsometer at Durham, a resolution of better than about ±10Å is not expected. Even though measurement error may be as low as ±0.02° if several measurements are taken in two zones, other experimental errors will be significant, such as in the initial calibration of the instrument. Despite these poor expectations, however, ellipsometric measurements of tunnel oxide thickness have occasionally been made during the course of this study and are included for completeness.

Other points of importance in the use of ellipsometry include;

(i) A value for the refractive index is sometimes assumed *a priori* in order to determine d_{ins}. However, as was pointed out in section 3.2.3, the refractive index of very thin SiO_2 layers grown on silicon is generally different to the value of 1.462 which applies to bulk stoichiometric oxide.

(ii) In most ellipsometric investigations, a simple two-layer model is assumed in which a homogeneous layer of SiO_2 overlies a flat reflective substrate. No account is taken of the presence of a transition layer.

(iii) The measurement is made over an area of oxide corresponding to the size of the light beam employed which is about 4mm diameter in the present work. Thus the value for d_{ins} obtained represents only a mean over that area. The only indication of uniformity is the degree to which the beam is extinguished at the points of cross-polarisation (see Appendix D) and the sharpness of the nulls achieved. In general, the nulls are quite broad and the beam is not well extinguished.

3.5.2 Measurement of Accumulated Capacitance

A MIS(n) diode may be biased into accumulation by applying a positive voltage to the metal, as indicated in figure 3.4. In this condition, any increase of the applied bias is dropped across the insulating layer such that the field lines are terminated on electrons in the potential well of the accumulation layer at the semiconductor surface. The diode thus behaves like a parallel plate capacitor with two sheets of charge separated by the oxide layer and the capacitance per unit area may be given essentially by;

$$C'_{accum} = C'_{ins} = \frac{\epsilon_{ins} \cdot \epsilon_0}{d_{ins}} \tag{3.6}$$

However, several limitations of this technique as applied to tunnel oxides must be noted;

(i) As in the case of ellipsometry, a homogeneous layer model is assumed and the measurement obtained is a mean value for the area of oxide determined by the size of the metal contact. There is no indication of the degree of inhomogeneity or non-uniformity of the oxide layer.

(ii) An error is introduced by the fact the centroid of the accumulation charge layer is not localised precisely at the i-s interface[32] due to the quantum mechanical effect of confining the electron wave functions in a potential well, as indicated in figure 3.4.

(iii) Because accumulation corresponds to forward bias, the direct current that flows in tunnelling MIS diodes may be too great for the instrument used.

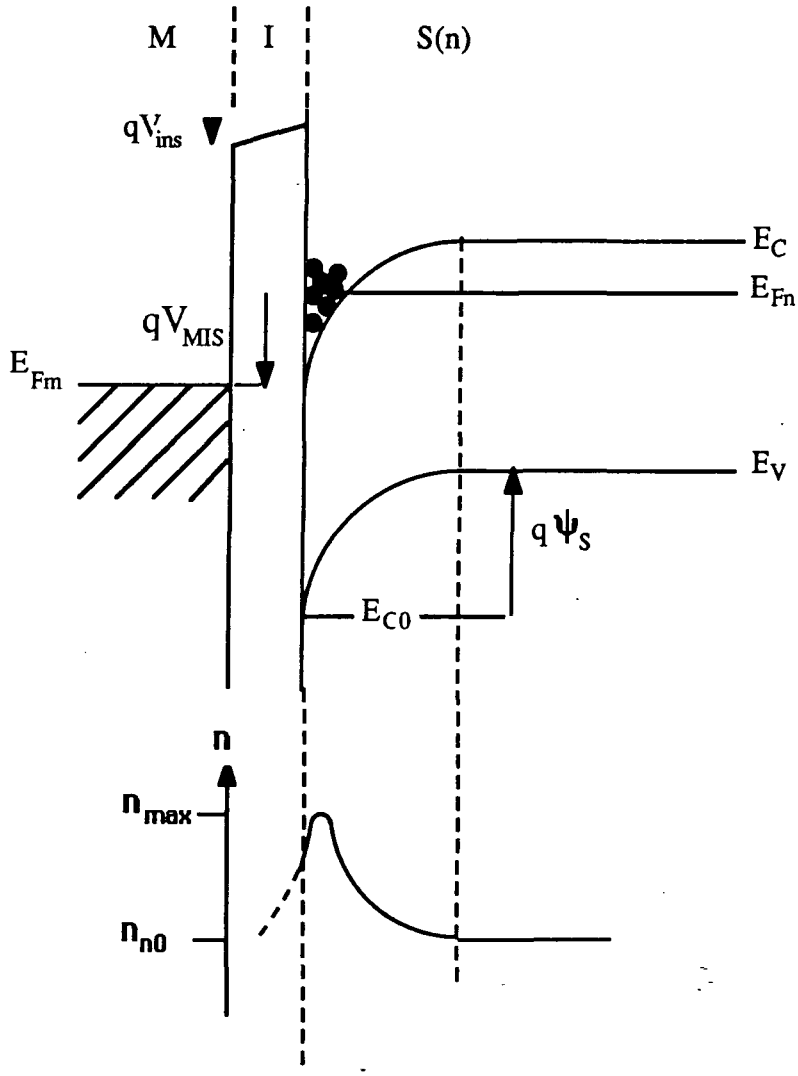


Figure 3.4; Band diagram of a MIS(n) diode in accumulation showing the quantum confinement of the accumulation layer some distance from the I-S interface.

Sample	Si crystal orientation	Oxidation time (min)
A	{111}	8
B	{100}	8
C	{111}	6

Table 3.3

Oxidation time and substrate orientation of the three MIS samples studied.

3.6 Electrical Characterisation of MIS Diodes

After fabrication, MIS diodes have been measured on-wafer and no attempt has been made to separate individual chips and mount and bond them to packages. Whole wafers were held on a vacuum chuck which provided contact to the substrate and individual devices were contacted using a mechanical probing station. Because each of the MIS diodes under study was provided with its own bond pad, it was possible to use tungsten needle probes which ensured penetration of the aluminium metallisation (which forms an insulating surface oxide) and thus good electrical contact.

Current-voltage and capacitance-voltage characteristics have been obtained from MIS diodes on each of three sample wafers. These three samples represent two different growth runs and two substrate orientations and are summarised in Table 3.3.

3.7 Current-Voltage Characteristics

Current-voltage (I-V) characteristics of MIS diodes have been measured using both a simple X-Y pen plotter and an automated measurement system. The latter incorporates a Keithley PIC485 picoammeter and DMM175 digital voltmeter with IEEE488 bus in a computer controlled system which is shown schematically in figure 3.5. The picoammeter is sensitive down to 1pA and offers 5 digit resolution. Applied bias was provided by a CIL120 D/A converter and was generally incremented in steps of 0.1V. Each data point (I,V) obtained was taken as the mean of fifty consecutive measurements at a constant bias. Figure 3.6 shows the forward and reverse characteristics of random sets of MIS diodes corresponding to each of the three types. These were obtained in the dark and at room temperature.

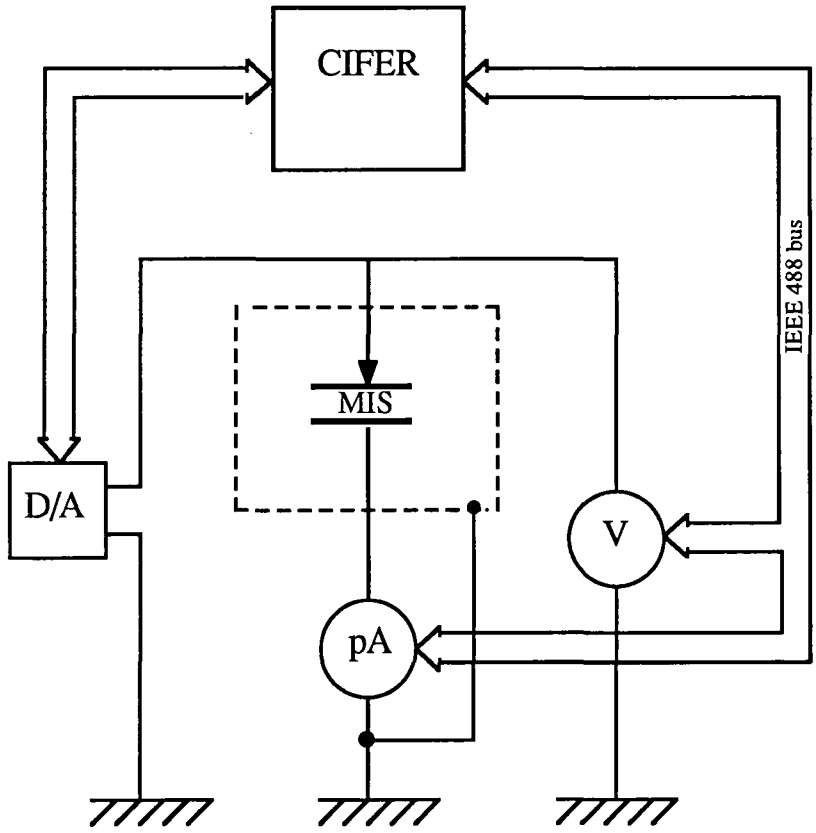


Figure 3.5; Computer-controlled measurement circuit for I-V characterisation of MIS tunnel diodes.

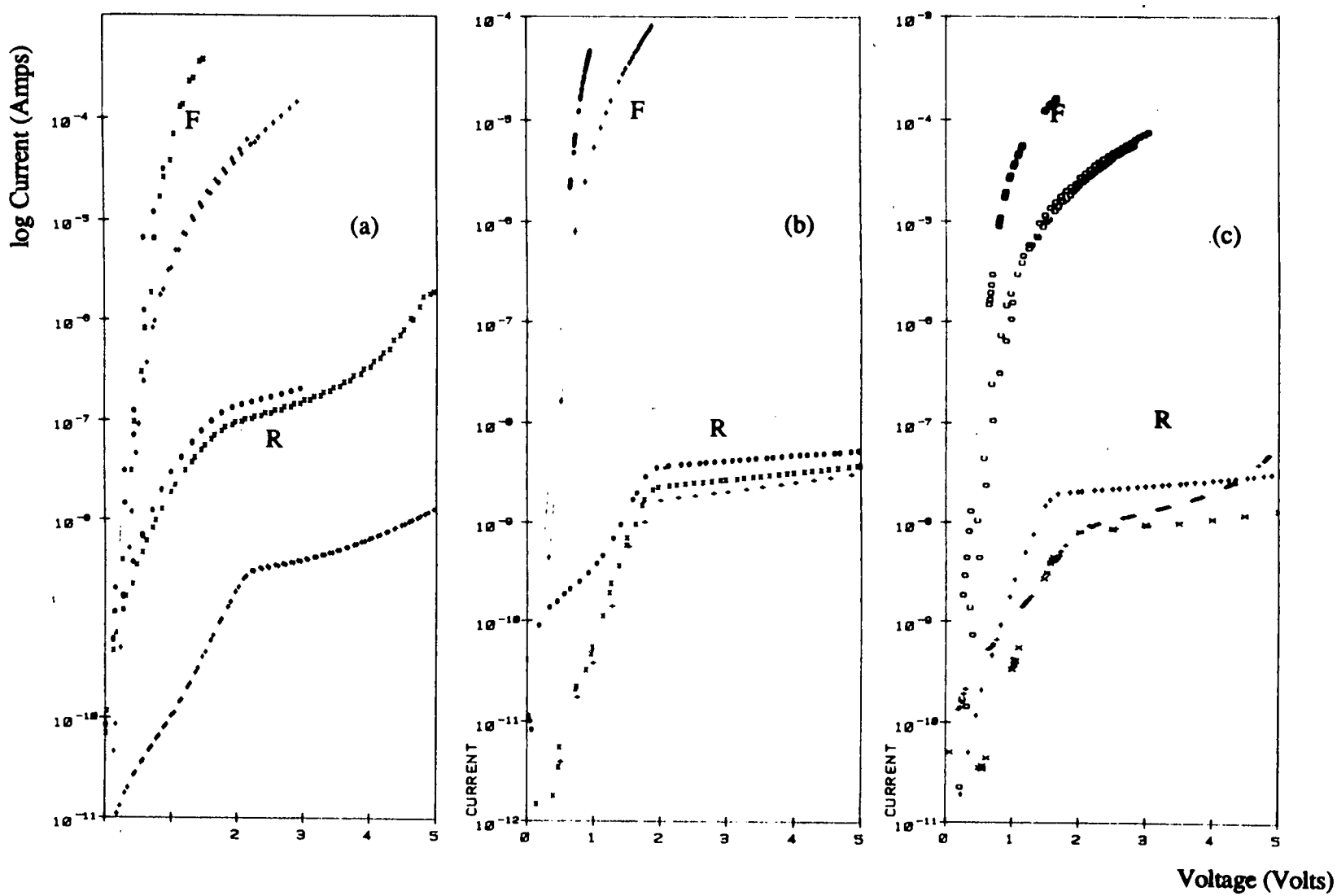


Figure 3.6; Forward and reverse bias I-V characteristics of tunnel oxide MIS diodes (a) type A, (b) type B and (c) type C, obtained in the dark at room temperature.

3.7.1 Forward Bias

It can be seen that in all three diode types, the forward current rises exponentially with voltage over five decades up to about 0.7V. The reduction in the rate of rise of current with voltage for larger forward bias may be explained by the onset of ohmic effects. Linear I-V pen plots obtained from several diodes of type C are shown in figure 3.7. Here, the current is seen to become a linear function of voltage when it exceeds about 1 mA. Such behaviour is typical of ohmic current limitation due to the combined resistances of the semiconductor bulk and the back contacts. The effective total resistance, as obtained from the slopes of these plots, is in the range 200Ω to 400Ω.

The exponential rise before the onset of ohmic effects has been observed by other authors[33-35] and was first explained by Card and Rhoderick[33] on the basis of standard tunnelling theory. It is first important to note that Al-SiO_2-Si(n) diodes are majority carrier devices[36] . Due to the small work function difference between semiconductor and metal, the thermionic barrier faced by electrons is small and electron tunnelling to the conduction band is strongly favoured under forward bias. The total forward current may therefore be expected to be due to electron tunnelling;

$$I_{fwd} = I_{NT} \tag{3.7}$$

Now, the electron tunnel current I_{NT} is given for forward biases greater than $3V_T$ by equation (2.27);

$$I_{NT} = -I_0 . \exp\left(\frac{V_{fwd}}{n.V_T}\right) \tag{3.8}$$

where, in the absence of tunnelling via interface states;

$$I_0 = I_{NT_0} . \exp\left(-\frac{\phi_{Bn}}{V_T}\right) = A_e.T^2 . \exp(-\chi_e^{1/2}.d_{ins}) . \exp\left(\frac{-\phi_{Bn}}{V_T}\right) \tag{3.9}$$

and the ideality factor n is given by equation (2.26);

$$n = -\left[\frac{\Delta\psi_s}{\Delta V_{fwd}}\right]^{-1} = \left[1 + \frac{C_{dep} + C_{ss}^S}{C_{ox} + C_{ss}^M}\right] \tag{3.10}$$

It may then be expected from these relations that the slope of the logI-linV straight line should correspond to 1/n and the intercept of the straight line extrapolated to V=0 should yield a value for I_0.

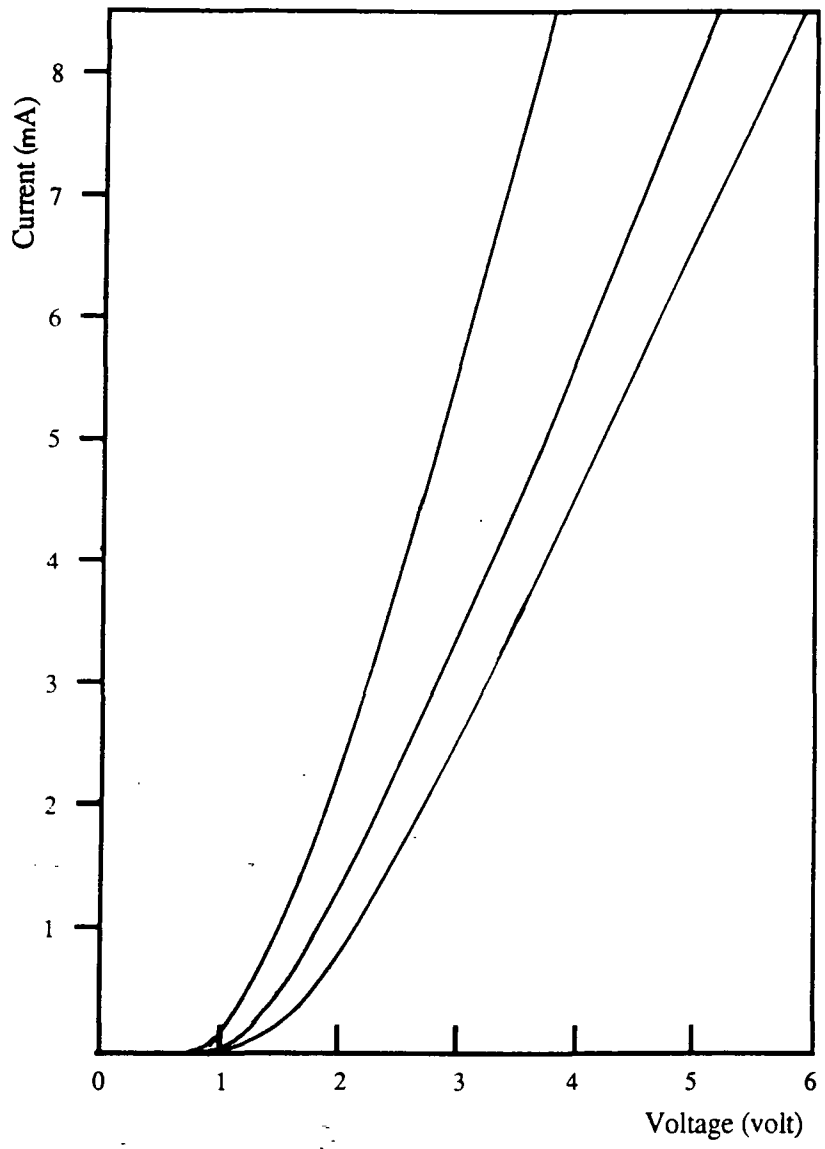


Figure 3.7; Linear I-V plots obtained from a variety of MIS diodes of type 'C' under forward bias.

Sample	Description	J_0(A/cm^2)	n	J_{sat} (A/cm^2)
A	{111}, 8 min	$1.3\,10^{-7}$	2.45	$5\,10^{-4}$
B	{100}, 8 min	$6.0\,10^{-8}$	2.11	$2.5\,10^{-4}$
C	{111}, 6 min	$4.0\,10^{-6}$	2.10	$8\,10^{-3}$

Table 3.4

Mean pre-exponential current density, J_0 and ideality factor n for the three diode types under forward bias with the corresponding reverse saturation current density, J_{sat}.

Mean values of I_0 obtained from the intercepts of several plots are given in Table 3.4 in terms of current density (A/cm^2) for the three diode types. They are comparable to values obtained by other authors[35] [37] and would appear to depend on oxidation time and substrate orientation.

Because I_0 is a function of both the tunnelling attenuation factor ($-\chi_e^{1/2}.d_{ox}$) and the barrier height for thermionic emission, ϕ_{Bn}, it is rather difficult to interpret these results. However, it may be considered that for very thin films ($d_{ox} \leq 15$Å) the tunnelling attenuation is negligible and equation (3.8) reduces to the thermionic current equation for a Schottky barrier. In the other limit, where the film is relatively thick ($d_{ox} \geq 25$Å) the tunnelling term may be expected to dominate.

A {111}-oriented control sample, oxidised alongside samples A and B, was measured by ellipsometry. However, the readings obtained; $\psi = 12.65$ and $\Delta = 165.47$, do not fall on any of the locii on the ellipsometry chart (Appendix D). Despite this, a value of $d_{ox} = 45$Å was extracted by assuming a refractive index of 1.46 and finding a best fit using the fitting program (Appendix D). In the light of the poor data and the discussion of section 3.5.1, this result must obviously be regarded as very approximate. It does serve as an indication however, that the oxide thickness is perhaps large enough for tunnelling attenuation to dominate the conduction.

Taking from Bagnoli and Nannini[35] a value for A_e of 110 A.cm^{-2}.K^{-2}, the total exponent ($\chi_e^{1/2}.d_{ox} + \phi_{Bn}/V_T$) may be calculated to be 0.766, 0.784 and 0.683 for MIS types A, B and C respectively. It remains then to determine whether this variation arises from differences in (a) the oxide thickness or (b) the barrier potential which is a function of the interface charge density. If it is assumed that only the former is important in

these cases and a value of 0.922/Å[36] is taken for $\chi_e^{1/2}$, oxide thicknesses of 34.6Å, 35.5Åand 30.9Åare obtained.

Comparing types A and C, the difference in J_0 may be expected to result from the different oxidation times due to a change in thickness, d_{ox}. In general, the difference between the oxide thicknesses may be obtained, assuming no change in the other parameters in (3.9), from;

$$\chi_e^{1/2}.(d_{ox}(\mathrm{C}) - d_{ox}(\mathrm{A})) = \ln \frac{J_{NTO}(\mathrm{A})}{J_{NTO}(\mathrm{C})} \tag{3.11}$$

Thus an increase in d_{ox} of only 2Å will result in more than a six-fold reduction in I_{NT}! Given that the forward current is so strongly affected by changes in oxide thickness, the spread in the measured curves under forward bias is not surprising.

Comparing diode types A and B, however, a reason for the difference in J_0 is not so obvious since the two samples were oxidised together and in close proximity to each other. It would appear that orientation is also important in determining J_0. Three effects of orientation on the properties of MIS diodes are well documented;

(i) The rate of oxidation for thin oxides is found to be slightly greater on the {111} as compared to the {100} oriented surface of silicon[25] [38] , as discussed in section 3.3.

(ii) The tunnelling probability between metal and semiconductor states is greater for the {100} oriented surface[36] for which the constant energy surface of the semiconductor states is centered on k=0 and the area of the shadow overlap between metal and semiconductor states is considerably greater than for the {111} orientation.

(iii) The density of interface states N_{ss}, attributed mainly to dangling bonds, is known to be about a factor of 3 lower on the {100} oriented surface[39] mainly because of the reduced density of surface atoms.

Effects (i) and (ii) would be expected to reduce J_0 for the {111} sample due to the respective consequences of higher d_{ox} and higher χ_e in equation (3.9). Only the third effect would tend to enhance J_0 through the addition of an interface trap-assisted tunnelling component $J_{NT_O}^{ss}$, as discussed in chapter 2. It might be concluded that the increase in J_{NT_O} due to N_{ss} for the {111}-oriented surface is more important than the reduction due to the other two effects.

If it is the case that trap-assisted tunnelling is significant for type A diodes, then it might be assumed that it may not be neglected for type C diodes either. As a consequence, the term I_{NT_O} in (3.9) should be replaced by the sum $I_{NT_O} + I^{ss}_{NT_O}$ and the values of d_{ox} determined earlier need to be re-assessed. Although it is difficult to estimate the contribution of $I^{ss}_{NT_O}$, an indication of this factor may be deduced from early work on MIS tunnelling by Clarke and Shewchun[34] . They found, quite reproducibly, that the 'effective tunnelling thickness', d_T, of an oxide on {100}silicon was related to the thickness determined from the accumulated capacitance, d_{ox} by;

$$d_T = 0.6\, d_{ox} \tag{3.12}$$

In their discussion, they assumed that this reduction arose from the existence of thickness non-uniformities in the oxide layer. This conclusion is questionable however, since, although significant non-uniformities are likely to arise (as discussed in section 3.2.2), they would be expected to affect the measured value of d_{ox} equally. It is therefore suggested instead that the factor of 0.6 is more likely due to the additional contribution to I_{NT_O} from trap-assisted tunnelling which Clarke and Shewchun chose to neglect.

The non-ideality factor, n, of each MIS diode type has been obtained from the measured slopes in figure 3.6 using the formula;

$$n = 17.376.\frac{\Delta V}{\Delta(\log_{10} I)} \tag{3.13}$$

The values calculated, in the range 2.1 to 2.5 (Table 3.4), are rather larger than reported by Card[33] . However, a large value does confirm that true tunnelling behaviour is manifest since conduction through pinholes would lead to a value more typical of a Schottky barrier, about 1.05.[40]

From equation (3.10) it can be seen that a value for n of about 2 would imply;

(i) The majority of the interface states are in communication with the the semiconductor such that N^S_{ss} is large and N^M_{ss} is small.

(ii) The oxide capacitance is comparable with the sum of the depletion layer and interface state capacitances.

Both these conditions are satisfied if the oxide thickness is approaching the tunnelling limit for MIS diodes of about 45Å and the interface state density is large.

3.7.2 Reverse Bias

Of more concern in a study of MIS switching devices is the reverse behaviour of the MIS diodes. The reverse bias (metal negative) plots in figure 3.6 are characterised in general by a large spread in the measured current values and by a distinct saturation effect, as expected. The initial steep exponential rise reaches a limit for all three diodes at around 1.8V after which the current continues to increase only very gradually in saturation. The onset of saturation is particularly sharp for diode types A and B and is evidence that the oxide thickness lies in the range 20Åto 45Å[41] Clarke and Shewchun[34] suggested that a MIS diode will remain in equilibrium (and thus not show any saturation) for oxide thicknesses greater than 33Å but here again the specification of a 'critical' value of the oxide thickness by different authors is subject to the large experimental errors inherent in its determination. Given the poor definition of the oxide interfaces, as discussed in sections 3.2.2 and 3.2.4, consideration of oxide thickness in absolute terms is not considered fruitful.

3.7.3 The Reverse Saturation Current

It is worthwhile now to consider in greater detail the nature of the reverse saturation current of the MIS diodes since this has a direct bearing on the 'off' state behaviour of the MIS switching devices. In reverse bias, most authors consider the majority carrier current, I_{NT} from equation (2.23) to be negligible since the potential dropped across the oxide is expected to be very small in the absence of a significant equilibrium hole density at the interface. Therefore, most of the total current is expected to be carried by holes.

According to the theoretical treatment of chapter 2, the exponential rise up to saturation corresponds to the situation where the holes are in equilibrium with the semiconductor. The quasi-Fermi level of the holes, E_{Fp}, is close to that of the majority carriers, E_{Fn} such that $\xi \sim 0$ and the current in this regime is tunnel-limited. Under the equilibrium condition $\xi = 0$, equation (2.28) reduces to;

$$I_{PT} = I_{PT_0} . \exp\left(\frac{q\psi_s - E_g + q\phi_{Fn}}{kT}\right) \tag{3.14}$$

which, again using the ideality factor n from (2.26) and the relation in (2.25), may be

further reduced to;

$$I_{PT} = I_{PT_0} . \exp\left(\frac{q.\phi_{Bn} + q.V/n - E_g}{kT}\right) \tag{3.15}$$

The saturation point corresponds to the transition from equilibrium to non- equilibrium conditions. As such, the reverse current in saturation becomes controlled by the semiconductor, as discussed in chapter 2. Under the assumption that the bulk of the total current is carried by holes, the saturated region will be controlled mainly by the rate of generation of holes in the depletion region which provides a current I_G;

$$I_{sat} \approx I_G \tag{3.16}$$

The importance of the generation current is demonstrated by the result shown in figure 3.8 where the effect of illumination is clearly to enhance the current in saturation by introducing an additional generation component. This effect has been confirmed in all types of MIS diode measured and imposes the restriction that, for consistency, all MIS and MISS characteristics must be measured in the dark.

However, despite the obvious influence of generation on the saturated current, the reverse bias I-V plots obtained from the three MIS diodes do not support a model which considers generated holes to be the dominant carrier. In fact, the measurements deviate from such a model in two important respects;

(i) If it was semiconductor limited, the reverse current ought to be a function only of the silicon properties and thus independent of the oxide layer thickness. Samples A and C were both fabricated on parts of the same {111} n-type wafer yet their reverse saturation curves are different. In addition, devices of equal size on any particular sample show a considerable spread in their measured reverse saturation currents.

(ii) It is noted that the rise in current in saturation is exponential with voltage which does not agree with the expression for I_G, equation (2.17), which predicts a $I \propto V^{1/2}$ relationship.

These anomalies may be resolved by assuming that the electron tunnel current (and consequently the oxide layer) does, after all, play a role in reverse saturation. The same conclusion was reached by Clarke and Shewchun[34] who similarly found that a semiconductor-limited hole current could not account for the measured value of saturation current. They also suggested that it is the electron tunnel current which

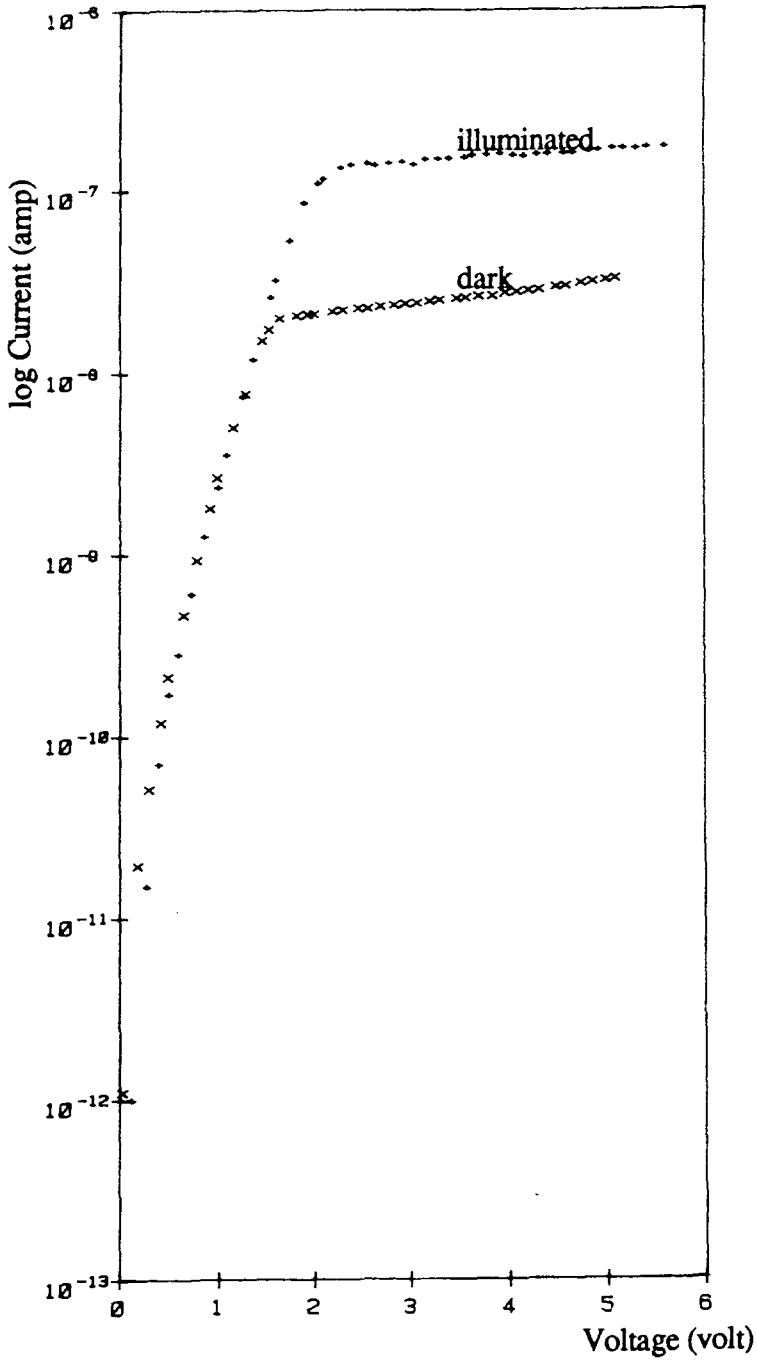


Figure 3.8; The effect of an arbitrary level of illumination on the reverse saturation current of a tunnel oxide MIS diode.

saturates, at a level determined by the oxide potential and thus by the minority carrier density $p(0)$ at the silicon surface. Since in the deep-depletion, non-equilibrium regime $p(0)$ is only a weak function of the reverse bias through $I_G = I_{PT} = I_{PT0}.p(0)$, the oxide potential is virtually constant and the conduction band at the interface, $E_C(0)$ is effectively pinned to the metal Fermi level, E_{Fm}.

Of course, at any quiescent point in the saturated region, an increase in $p(0)$ due to some increase in I_G (arising, for example, from heating or illumination of the sample) will shift the pinning position of $E_C(0)$ relative to E_{Fm} and thus lead to a larger oxide potential and hence to a larger quiescent electron tunnel current, I_{NT}. Or, as stated by Clarke and Shewchun[34] , 'reducing the limitations on the production rate of minority carriers forces the reverse current to saturate but at a slightly greater magnitude'- a new steady-state hole density is established, large enough to affect the majority carrier flow but still too small to invert the surface.

Considering equation (3.8);

$$I_{NT} = -I_{NTo}.\exp\left(\frac{V_{ox} - \phi_{Fn}}{V_T}\right) \tag{3.17}$$

for I_{NT} to be significant, V_{ox} must have a value of at least the order of V_T, that is $\sim$ 25mV at room temperature. The value of V_{ox} is given by the charge neutrality equation;

$$\epsilon_{ox}.\epsilon_0.\frac{V_{ox}}{d_{ox}} = \epsilon_s.\epsilon_0.\mathcal{E}_s - Q_{ss} \tag{3.18}$$

where the field in the silicon surface is given from Appendix B as approximately;

$$\mathcal{E}_s = \left(\frac{2.q}{\epsilon_s.\epsilon_0}\right)^{1/2}.\left[N_D.\psi_s + V_T.p(0)\right]^{1/2} \tag{3.19}$$

To briefly summarise the discussion in Appendix B, $\mathcal{E}_s$ has two components; the field due to the ionised donors in the depletion layer and the field due to the free holes in the inversion layer. Since it is expected that the minority carrier density will be negligible in non-equilibrium MIS diodes, it must be assumed that the field arises from the exposed donor ions. This may be confirmed to be the case by a simple estimation of V_{ox}. Putting the values $N_D = 2 \times 10^{15}\text{cm}^{-3}$, $\psi_S = -5V$ and $d_{ox} = 40$Å into equations (B.7) and

(3.18);

$$\mathcal{E}_S = \left(-\frac{2.q.N_D.\psi_S}{\epsilon_s.\epsilon_o}\right)^{1/2} \sim 5.5 \times 10^6 V/m \longrightarrow V_{ox} = \frac{\epsilon_s}{\epsilon_{ox}}.d_{ox}.\mathcal{E}_S \sim 66mV \qquad (3.20)$$

To summarise then, the slight rise in the saturation current with increasing reverse bias may be attributed to the growth of the depletion layer. Thus, this slope is independent of the level of illumination. The additional increase in saturation current due to illumination, manifested as a shift of the I-V curve along the I-axis, may by contrast be attributed to a supplemented minority carrier population at the interface, $p(0)$. As such, the MIS diode is behaving somewhat like a bipolar transistor in that it is exhibiting the potential for current gain.

3.7.4 Capacitance-Voltage Characterisation of MIS Diodes

Capacitance versus voltage (C-V) curves from MIS diodes have been measured by the author at LAAS, Toulouse where a HP4192A impedance analyser was available. This instrument was used to measure capacitance over the frequency range 1kHz to 2MHz. Connections to devices were made by mechanically probing to bond pads, in the same manner as for the I-V measurements. The large parasitic capacitance of the mechanical assembly and associated wiring was first eliminated by adjusting the zero offset on the capacitance meter with the probe held well away from the contact pad. However, the parasitic capacitances of the bond pad and interconnects could not be removed. This capacitive component has been measured independently by severing the connection to a MIS diode and found to be of the order of 2 pF. As such, the offset in the MIS capacitance readings is certainly not negligible. In principle, to extract useful information such as d_{ins} and the distribution of N_{ss} from capacitance measurements, the device area should be large and parasitic effects negligible.

Figure 3.9 shows two sets of C-V curves obtained from MIS diodes of type 'C'. The first set (fig.3.9(a)) represent the combined capacitance of two diodes (numbers 4 and 5) which were connected in parallel and thus had a total nominal area of 10,800 μm^2. The second set were obtained from diode 4 alone (measured area 5670μm^2) but on a different chip. In both cases there is clearly a very large frequency dispersion effect on the capacitance in the depleted and accumulated regions. The curves are shifted to

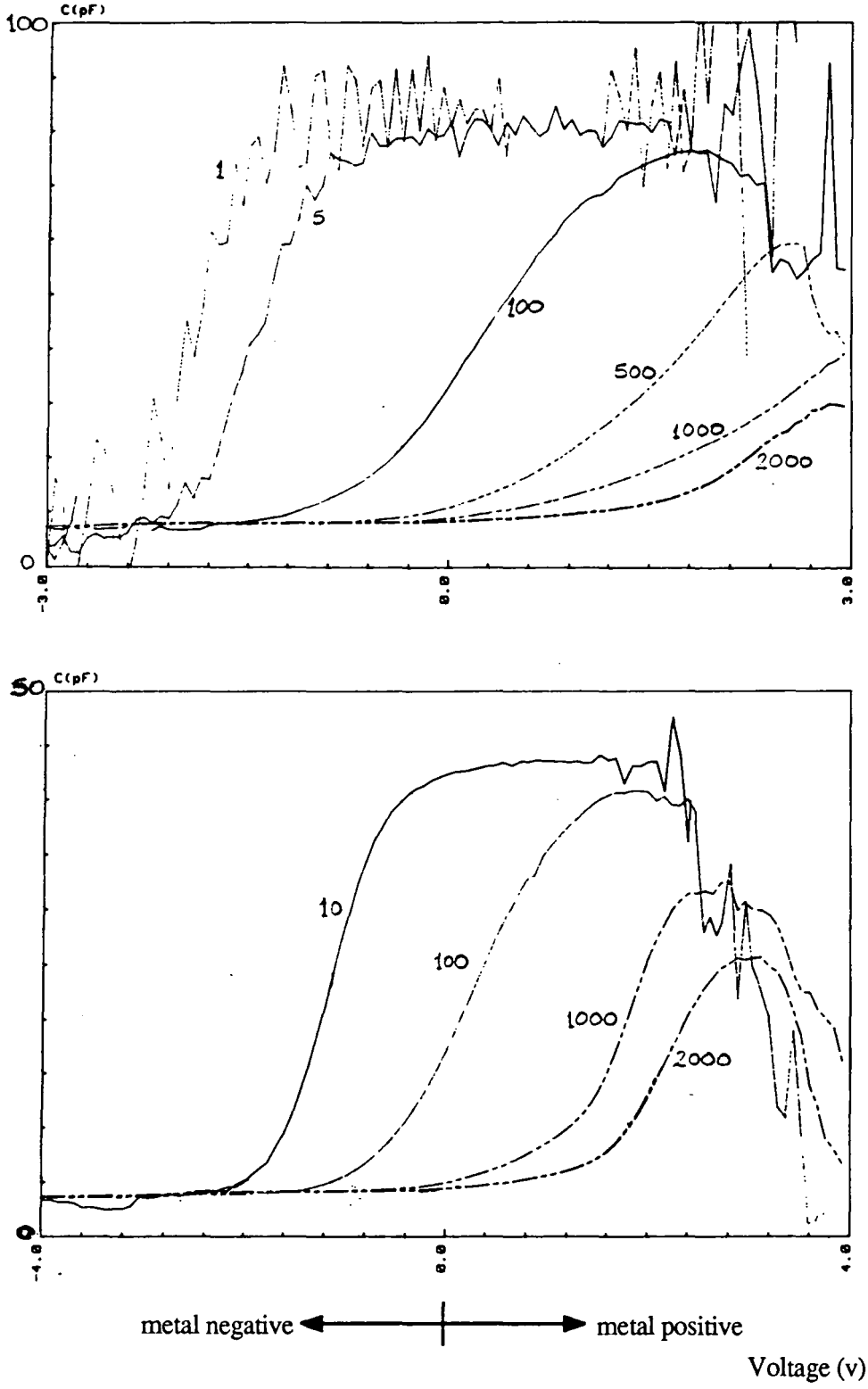


Figure 3.9 ; Capacitance-voltage characteristics of MIS diodes of type 'C';
(a) diodes 4 and 5 in parallel, chip B4,
(b) diode 5, chip A4
Frequency (kHz) is the parameter.

larger reverse bias (metal negative) as the frequency is reduced and there is a marked fall in the maximum capacitance measured in forward bias (metal positive) as the frequency is increased.

The phenomenon of 'anomalous dispersion' of the accumulation capacitance has been noted by other authors but usually in relation to insulating layers on compound semiconductors, GaAs[42] and InP[43] which have a notoriously poor interface. There are several possible explanations for the observed dispersion;

(i) Frequency dispersion of the dielectric constant of the insulating layer[43] .

However, this suggestion would imply that ϵ_{ox} actually decreases with increasing frequency and thus does not appear to be reasonable for an SiO_2 layer.

(ii) Series resistance.

Any series resistance, R_S, in the device under study (arising from bulk and contact resistances) will give rise to a reduction in the measured capacitance value at high frequencies, according to the relation;

$$C_{correct} = \frac{(G_{meas}^2 + \omega^2 C_{meas}^2).C_{meas}}{[G_{meas} - (G_{meas}^2 + \omega^2 C_{meas}^2).R_S]^2 + \omega^2.C_{meas}^2} \tag{3.21}$$

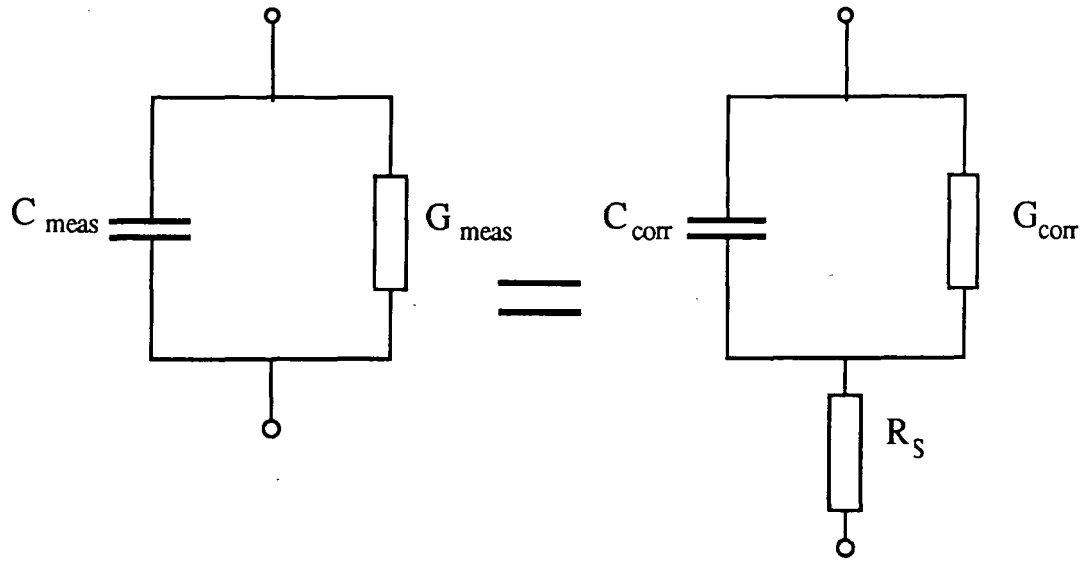
which is obtained from the equivalent circuit of figure 3.10(a). However, even after correcting for C_{meas} using $R_S = 200\Omega$ and conductance data which was obtained simultaneously with the capacitance readings, very little difference is made to the high frequency C-V curve, as shown in figure 3.10(b). The effect of R_S would appear to be insufficient to account for the observed dispersion.

(iii) The contribution of interface state capacitance in accumulation.

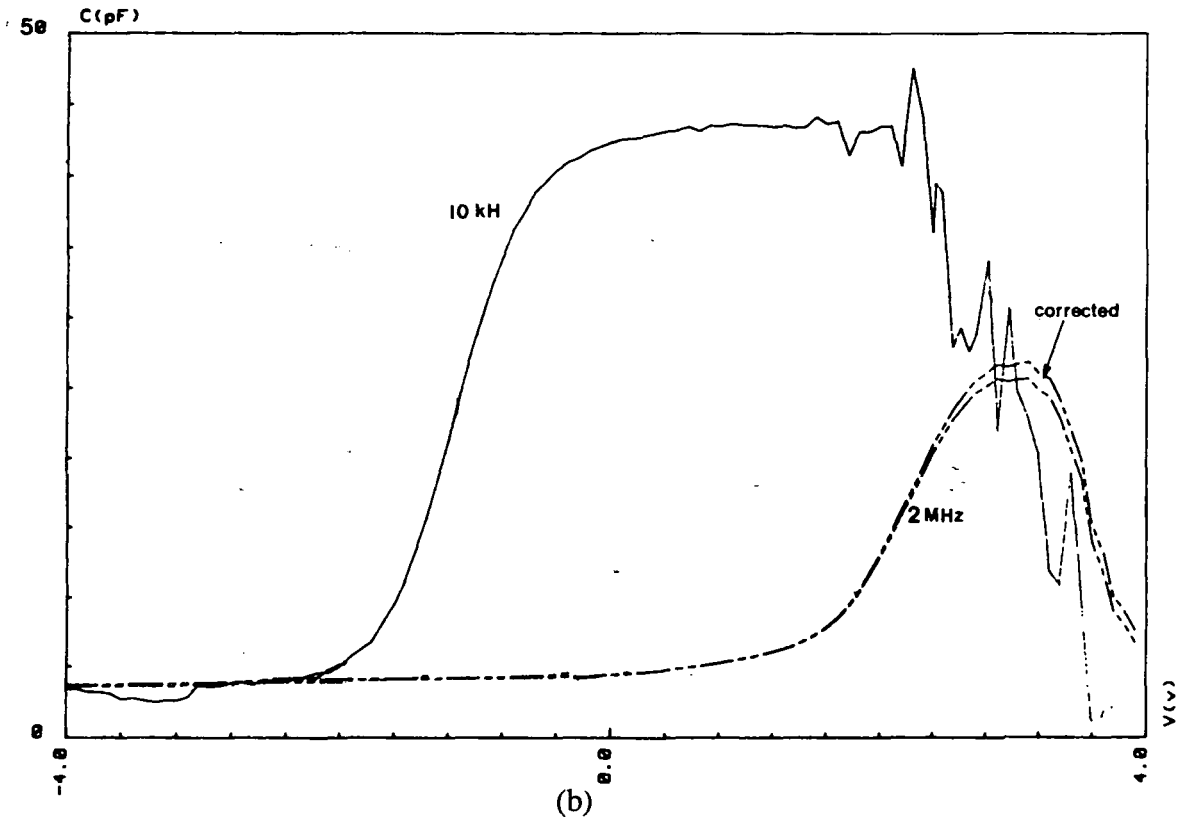
Sawada and Hasegawa[42] were able to account for the dispersion of the **accumulated** capacitance observed in metal-oxide-GaAs diodes on the basis of a high interface state density, of the order of $10^{13}cm^{-2}$. However, in the present case, the high 'n' factors obtained for the tunnel oxide MIS diodes (section 3.7.1) do not support such a conclusion. Recalling equation (2.26);

$$n = \frac{\Delta V}{\Delta \psi_S} = 1 + \frac{C_{dep} + C_{ss}^S}{C_{ox} + C_{ss}^M} \tag{3.22}$$

n values in the range 2.1 to 2.5 would suggest $C_{dep} + C_{ss}^S \geq C_{ox} + C_{ss}^M$ which only re-inforces the approximation $C_{ox} \sim C_{accum}$.



(a)



(b)

Figure 3.10; (a) Equivalent circuits showing the measured and corrected C and G components of a tunnel MIS diode containing series resistance.
(b) Low and high frequency C-V curves from figure 3.9 with the high frequency curve corrected for R_S=200Ω.

(iv) Incomplete accumulation.

The remaining possibility is that the peaks of the measured capacitance plots do not after all correspond to true accumulation of the semiconductor surface. If the semiconductor is still depleted over the voltage range -2 to 4 V, then the contribution of C_{ss}^S may be important. Indeed, C-V characteristics of a very similar appearance to those in figure 3.9 have been reported by Kar and Dahlke[40] for tunnel oxide MIS diodes. A comparison is made between their results and those of the present study in figure 3.11 which is a plot of the peak capacitance against measurement frequency. In their discussion, Kar and Dahlke point out that the large dispersion indicates that the traps are tunnelling rather than recombination controlled[40] . As such, they are in equilibrium with the semiconductor and therefore effectively in parallel with C_{dep}.

The equivalent circuit of a tunnel MIS diode containing a frequency-dependent interface state component is shown in figure 3.12(a). In this circuit, the capacitance due to interface states which are in equilibrium with the semiconductor ($C_{ss}^S(\omega)$) is in parallel with the depletion layer capacitance (C_{dep}). ($G_{TS}+G_{TE}$) represents the D.C. (frequency independent) conductance due to tunnelling both via interface traps and directly from band-to-band. The circuit may be reduced to that of figure 3.12(b) in the high frequency limit where the trapped charge is unable to follow the time-varying potential and the associated capacitance component C_{ss}^S is removed.

Following this analysis, it is possible to extract an estimate of C_{ss}^S. If it is assumed that the interface states are unable to follow the measurement signal at high frequency, then the high frequency C-V curve represents only the depletion layer capacitance in series with the oxide capacitance.

At low frequency (LF), C_{ss}^S may be expressed as a function of ψ_S;

$$C_{ss}^S(\psi_S) = C_P(\psi_S, \mathrm{LF}) - C_{dep}(\psi_S) \tag{3.23}$$

At high frequency (HF), $C_{ss}^S \rightarrow 0$, leaving;

$$C_{dep}(\psi_S) \sim C_P(\psi_S, \mathrm{HF}) \tag{3.24}$$

Thus, an estimate of C_{ss}^S may be obtained from;

$$\begin{aligned} C_{ss}^S(\psi_S) &= C_P(\psi_S, \mathrm{LF}) - C_P(\psi_S, \mathrm{HF}) \\ &\sim C_{meas}(\psi_S, 10\mathrm{kHz}) - C_{meas}(\psi_S, 2\mathrm{MHz}) \end{aligned} \tag{3.25}$$

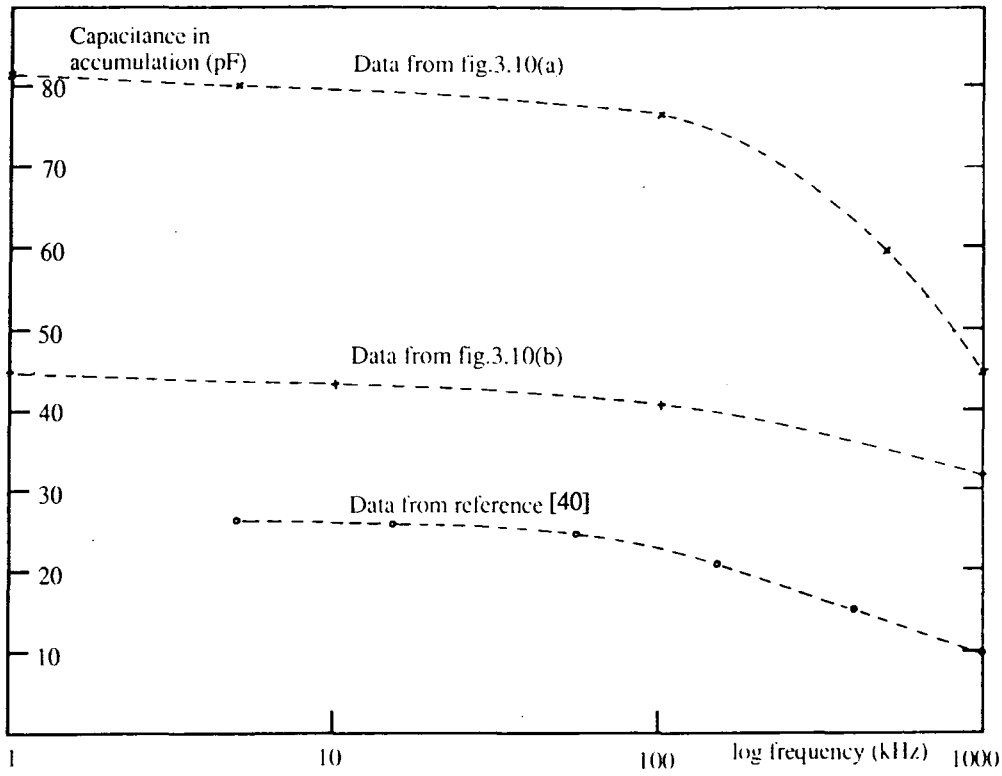
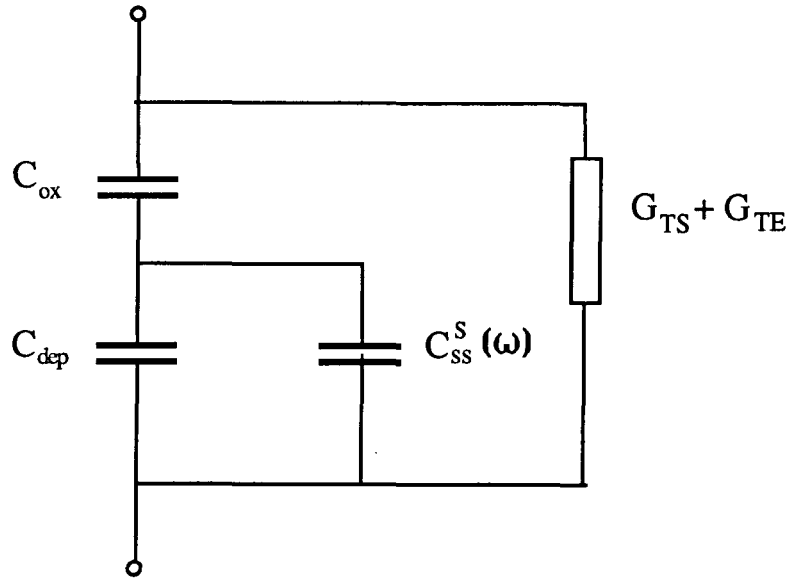
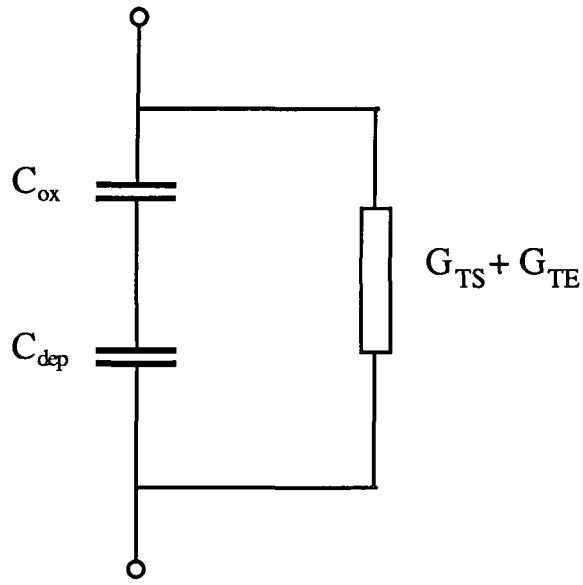


Figure 3.11; Frequency dispersion of the measured capacitance of tunnel oxide MIS diodes biased into accumulation.



(a)



(b)

Figure 3.12; Equivalent circuit of a tunnel MIS diode in which the interface traps are tunnelling-controlled (in equilibrium with the semiconductor);
(a) at intermediate frequency
(b) at high frequency.

Taking measured values of capacitance at 0V from figure 3.9(b),

$$C_{ss}^{S} \sim 40pF \longrightarrow N_{ss} = \frac{C_{ss}^{S}}{q.(\text{Area})} = 1.39 \times 10^{13}\text{cm}^{-2} \tag{3.26}$$

In addition, the shift of the C-V curves along the voltage axis, ΔV, allows an estimation of N_{ss} to be made using the relation $q.N_{ss} = \Delta V \times C_{ox}$. Again using the data from figure 3.9(b), a value of 1.75×10^{13} cm^{-3} is obtained. These are indeed very high densities for an SiO_2-Si interface and support the conclusion made in section 3.6.1 that interface states are making a strong contribution to the characteristics of these devices.

3.8 Conclusions

MIS diodes of the type aluminium–tunnel-oxide–n-type silicon have been fabricated and shown to exhibit saturation in their reverse current-voltage characteristics, indicative of a semiconductor-limited current regime. Under this condition, the diodes are considered to be in a non-equilibrium, deeply-depleted state characterised by fermi-level splitting in the depletion layer.

The form of both the forward and reverse bias I-V characteristics suggests that electrons are the principal current carriers in these devices. This is in agreement with the theoretical result[36] that MIS junctions with a low value of work-function difference, ϕ_{ms} should be majority carrier diodes. Because the electron tunnel current is exponentially dependent on the oxide potential, its magnitude is directly controlled by the field at the surface of the semiconductor. Under reverse bias, this field is in turn a function of the depth of the depletion layer (the amount of revealed ionised donor atoms) and of the equilibrium density of free holes at the I-S interface. The former gives rise to the gradual increase in the saturation current as the reverse bias is increased and the depletion layer is extended more deeply into the silicon. The result of an increase in the latter however, is an upward shift of the reverse characteristic, in which the magnitude of the reverse current increases uniformly throughout the saturated region. It is this effect which is fundamental to the behaviour of the MIS diode as a current amplifier. Any increase in the hole flux to the semiconductor surface (due for instance to illumination by photons with $h.\nu > E_g$) establishes a greater concentration of holes at the I-S interface and thus an enhancement of the electron tunnel current.

REFERENCES

1. R.B.Godfrey and M.A.Green, Appl. Phys. Letters **34** 790 (1979)
2. M.A.Green and R.B.Godfrey, IEEE Electron Device Lett. **EDL-4** 225 (1983)
3. T.W.Hickmott, J. Appl. Phys. **51** 4272 (1980)
4. S.Roberts and P.J.Dobson, J. Phys. D **14** L17 (1981)
5. Y.E.Strausser and K.S.Majumder, J. Vac. Sci. Technol. **15** 238 (1978)
6. J.Borel, Revue de Physique Appl. **13** 587 (1978)
7. E.P.EerNisse, Appl. Phys. Letters **35** 8 (1979)
8. E.A.Irene, E.Tierney and J.Angilello, J. Electrochem. Soc. **129** 2594 (1982)
9. S.T.Pantelides, J. Vac. Sci. Technol. **14** 965 (1977)
10. J.Shewchun, J.Dubow, A.Myszkowski and R.Singh, J. Appl. Phys. **49** 855 (1978)
11. L.C.Feldman, I.Stensgaard, P.J.Silverman and T.E.Jackman, in *The physics of SiO_2 and its interfaces*, ed. S.T.Pantelides (Pergamon, 1978)
12. O.L.Krivanek, D.C.Tsui, T.T.Sheng and A.Kamgar, *ibid.*
13. T.H.diStefano, J. Vac. Sci. Technol. **13** 856 (1976)
14. R.A.Clarke, R.L.Tapping, M.A.Hopper and L.Young, J. Electrochem. Soc. **122** 1347 (1975)
15. J.F.Wager and C.W.Wilmsen, J. Appl. Phys. **50** 874 (1978)
16. M.H.Hecht, P.J.Grunthaner and F.J.Grunthaner, in *17th Int. Conf. on the Physics of Semiconductors*, edited by M.Averous (1984)
17. A.H.Carim and A.Bhattacharyya, Appl. Phys. Letters **46** 872 (1985)
18. S.Kar and W.E.Dahlke, Solid State Electronics **15** 221 (1972)
19. T.W.Sigmon, W.K.Chu, E.Lugujjo and J.W.Mayer, Appl. Phys. Letters **24** 105 (1974)
20. P.Offerman, J. Appl. Phys. **48** 1890 (1977)
21. B.E.Deal and A.S.Grove, J. Appl. Phys. **36** 3770 (1965)
22. J.R.Ligenza, J. Phys. Chem. **65** 2011 (1961)

23. H.Z.Massoud, C.P.Ho and J.D.Plummer, Stanford University Tech. Rep. TR DXG501-82 (1982)

24. E.A.Irene, J. Electrochem. Soc. **125** 1708 (1978)

25. S.I.Raider and R.Flitsch, J. Electrochem. Soc. **122** 413 (1975)

26. F.N.Schwettman, K.L.Chiang and W.A.Brown in *Extended Abstracts of the Electrochemical Society Spring Meeting*, **78-1** 688 (1978)

27. F.J.Grunthaner and J.Maserjian, IEEE Trans. Nucl. Sci., **NS-24** 2108 (1977)

28. G.Gould and E.A.Irene, J. Electrochem. Soc. **134** 1031 (1987)

29. G.Sarrabayrouse, private communication.

30. F.LHsueh, L.Faraone and J.G.Simmons, Solid State Electronics **27** 499 (1984)

31. R.B.Calligaro, S.Moustakas, J.Dell and A.G.Nassibian, IEEE Proc I **128** 174 (1981)

32. D.Schmitt-Landsiedel, K.R.Hofmann, H.Oppolzer, G.Dorda and P.Pongrantz in 'Insulating Films on Semiconductors', editors J.F.Verweij and D.R.Wolters (North-Holland, 1983)

33. H.C.Card and E.H.Rhoderick, J.Phys. D **4** 1589 (1971)

34. R.A.Clarke and J.Shewchun, Solid State Electronics **14** 957 (1971)

35. P.E.Bagnoli and A.Nannini, Solid State Electronics **30** 1005 (1987)

36. M.A.Green and J.Shewchun, Solid State Electronics **17** 349 (1974)

37. K.K.Ng and H.C.Card, J. Appl. Phys. **51** 2153 (1980)

38. G.A.Haas and H.F.Grey, J. Appl. Phys. **46** 3885 (1978)

39. B.E.Deal, J. Electrochem. Soc. **121** 198C (1974)

40. S.Kar and W.E.Dahlke, Solid State Electronics **15** 221 (1972)

41. J.Ruzyllo, K.Kucharski and A.Jakubowski, Solid State Electronics **23** 1041 (1980)

42. T.Sawada and H.Hasegawa, Electronics Letters **12** 471 (1976)

43. A.J.Grant, D.C.Cameron, L.D.Irving, C.E.Greenhalgh and P.R.Norton, *'Insulating Films on Semiconductors 1979'* (ed. G.G.Roberts and M.J.Morant, Inst. of Physics 1979)

CHAPTER FOUR

A Theoretical Basis for the Study of MISS Devices

4.1 Introduction

The generality of the switching phenomenon in MISS devices, as discussed in the introductory chapter, suggests that a theoretical treatment of the device operation should also be kept as general as possible. As such, the basic S-shaped characteristic should ideally be accounted for, at least qualitatively, by a model which is not limited to any particular semi-insulator conduction mechanism or device structure. Although the effects of interface states in supplementing the tunnelling currents in the MIS structure are thought to be significant, these and high injection effects on the pn junction currents, will be left out of the analysis for simplicity. It is not the intention here to try and refine the model in order to obtain a fit to experimental results. The object of this discussion is to elucidate the mechanism behind the switching effect. It is quite clear from the literature that there are essentially two fundamentally different approaches to modelling the MIS switch. One is a steady-state treatment which assumes (a) there is continuity of all the components of the hole and electron fluxes and (b) that all distributions of charge and of potential within the device are invariant with time. The second model considers the device as a regenerative feedback system in which a current feedback loop is active. In this analysis, based on thyristor theory, the switching mechanism is associated with the condition of unity loop gain.

A 'first order' steady-state description of the two terminal MISS based mainly on the work of Habib and Simmons[1] and Fiore de Mattos and Sarrabayrouse[2] will first be outlined. The steady state model has recently been developed further and investigated extensively by Lavelle[3] in parallel with the author's experimental investigations and references will be made to his thesis.

The regenerative feedback mechanism as applied to the MISS will then be discussed with reference to the variety of published work on this matter.[4-7]

4.2 A Steady State Model of the MISS

In describing the steady state, direct current, mode of operation of the MISS, it is first important to set some pre-conditions.

(i) The steady state mode of operation is defined as that condition whereby the spatial distribution of holes and electrons within the device can be considered constant with time. Any change in bias is assumed to be imposed more slowly than the slowest physical process occurring in the device. It does not necessarily imply that the carriers are in thermal equilibrium.

(ii) As MISS devices exhibit 'S'-type NDR, their I-V characteristics are effectively current-controlled. This implies that for any given current through the device there must exist a unique solution for the potential drop across it. As a consequence for modelling the device, the fundamental independent parameter is the total current.

(iii) The stated equations will be 'first order' on the basis that the generality of the switching effect suggests that it should be accounted for, at least qualitatively, without any dependence on second order phenomena. It is acknowledged however that other physical effects such as the presence of interface states and recombination at surfaces will affect the results quantitatively.

(iv) All effects of device configuration such as pn and MIS areas, lateral or epitaxial structure, will be neglected initially and the problem will be reduced to one dimension, as indicated in figure 4.1. A unit cross-sectional area will be assumed which removes the distinction between current and current density and the depletion regions at both the MIS and pn junctions will be assumed to extend only perpendicular to the junction planes.

(v) Only the m-i-n-p^+ structure with a tunnel oxide i-layer will be treated here, although the model is adaptable to the m-i-p-n^+ structure and to other semi-insulating layers with appropriate modifications.

(vi) Any multiplication due to avalanching in the surface depletion region which might arise at high fields is neglected.

(vii) High injection effects in the surface depletion and 'neutral' base regions will not be considered.

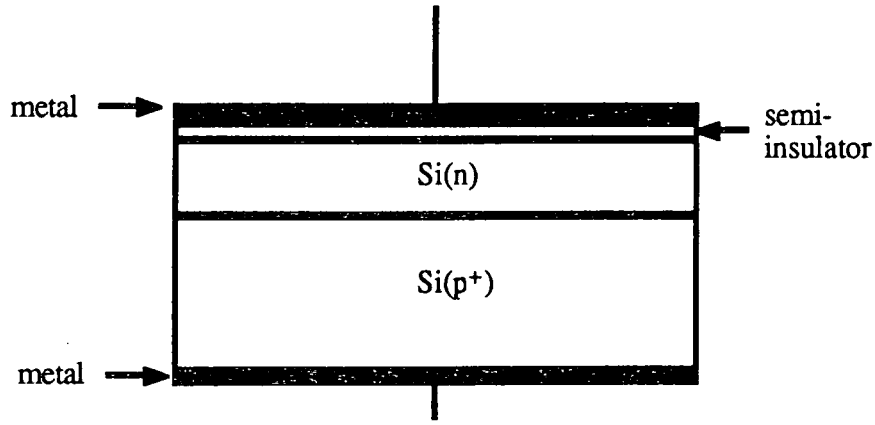


Figure 4.1; Schematic cross-section of a one-dimensional epitaxial MISS device.

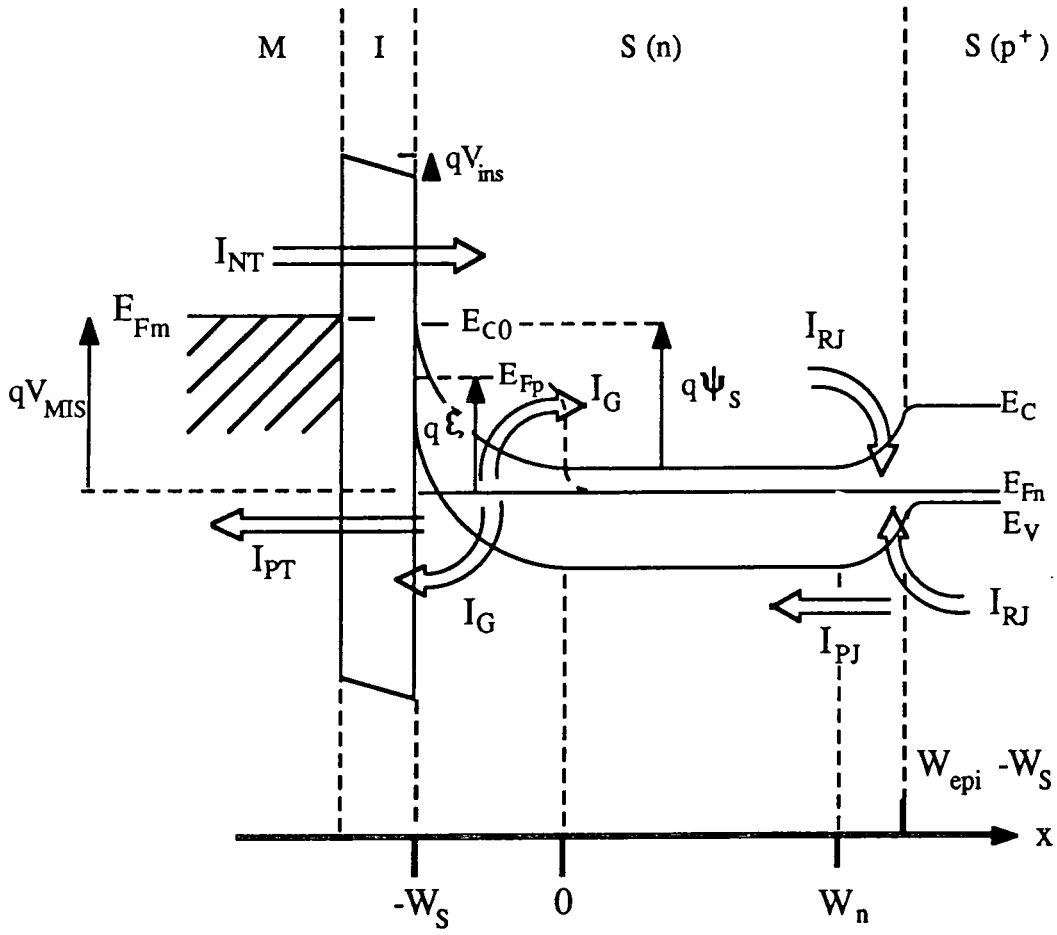


Figure 4.2; Band diagram of a M-I-S(n)-S(p) structure showing the principle potential differences and current components within the device.

The energy band diagram for the m-i-n-p$^+$ structure is shown in figure 4.2. Here the main current components which are to be considered and the various energy drops are also indicated.

In the normal mode of operation of a MISS device, the pn junction is forward biased and the MIS is biased in reverse. Thus, as the potential across the structure is increased, the depletion region at the semiconductor surface, of width W_S, will grow deeper according to the relation;

$$W_S = \left(\frac{2.\epsilon_s.\epsilon_o}{q.N_D.}.\psi_S\right)^{1/2} \tag{4.1}$$

which is derived from Poisson's equation as applied to the space charge region and assumes the depletion approximation. Here, ψ_S is the surface potential, N_D is the donor concentration in the n-type material and ϵ_s and ϵ_o the relative permittivity of silicon and the permittivity of free space respectively.

Now, considering that the device consists of a pn and a MIS junction in intimate contact, interfaces between these two elements may be defined and each considered separately. Then the two elements can be brought together within the constraints of current continuity and charge neutrality and a model for the whole system established.

The first interface is defined as the plane x=0 in figure 4.2 whose position corresponds to the edge of the MIS depletion region which, from (4.1), is dependent on the surface potential, ψ_S. The second interface is defined as the insulator-semiconductor (I-S) interface at $x = -W_S$.

The two diodes are in communication with each other via the surface depletion region ($-W_S < x < 0$) in which the field sweeps electrons towards the pn junction and holes to the I-S interface.

4.3 The pn Junction

By analogy with the bipolar transistor, the p$^+$ and n-type regions of the forward biased junction correspond to an emitter and base respectively. Where the emitter is highly doped in comparison with the base, the depletion region at the junction will exist mainly in the lower doped side and have a width, W_J which is dependent on the applied

potential, V_J;

$$W_J = \left(\frac{2\epsilon_s.\epsilon_0}{q.N_D}\right)^{1/2}.(\phi_{bi} - V_J)^{1/2} \tag{4.2}$$

This represents a one-sided junction approximation where V_J is applied such that the emitter-base junction is forward biased. The volume of semiconductor between the pn and MIS depletion regions is considered to remain neutral since the field in this region is for most purposes negligible. The width of the neutral region is given by;

$$W_n = W_{epi} - W_S - W_J \tag{4.3}$$

4.3.1 **Recombination in the pn Junction**

If the p^+-n junction is again treated as one-sided, the recombination current at the junction is due to recombination in the depleted n-type material. The standard Sah, Noyce and Schockley[8] expression for symmetrical junctions (where $N_A \approx N_D$) becomes in this case;

$$I_{RJ} = \frac{q.n_i}{\tau_p}.\frac{W_J}{2}.\exp\left(\frac{V_J}{2.V_T}\right) \tag{4.4}$$

4.3.2 **Hole Diffusion Current.**

The forward potential, V_J, existing across the pn junction gives rise to an excess density of holes (minority carriers) at the edge of its depletion region ($x = W_n$). The excess density, p_j, is a function of both V_J and the equilibrium concentration of minority carriers in the n-type base;

$$p_j = p(W_n) - p_{n0} = \frac{n_i^2}{N_D}.\left[\exp\left(\frac{V_J}{V_T}\right) - 1\right] \tag{4.5}$$

The density of holes at the edge of the MIS depletion region, p_s, is given by a similar expression;

$$p_s = p(0) = \frac{n_i^2}{N_D}.\exp\left(\frac{-\xi}{V_T}\right) \tag{4.6}$$

and will, at least at low current levels, be lower than the equilibrium concentration.

Diffusion of the excess of minority carriers away from the pn junction and towards the MIS depletion layer constitutes a hole current, I_P, which can be described as a function of distance, x by;

$$I_P(x) = q.D_p.\frac{\partial p(x)}{\partial x} \tag{4.7}$$

A solution for $I_p(x)$ may be obtained given the limits $p(W_n) = p_j$ and $p(0) = p_s$;

$$I_P(x) = \frac{q.D_p}{L_p.\sinh(\frac{W_n}{L_p})}.\left[p_s.\cosh\left(\frac{W_n - x}{L_p}\right) - p_j.\cosh\left(\frac{-x}{L_p}\right)\right] \tag{4.8}$$

Now, the hole current injected at the pn junction where $x = W_n$ becomes;

$$I_{PJ} = \frac{q.D_p}{L_p.\sinh(\frac{W_n}{L_p})}.\left[p_s - p_j.\cosh\left(\frac{W_n}{L_p}\right)\right] \tag{4.9}$$

and the hole current collected at the MIS depletion edge (x=0) is given by;

$$I_{PJS} = \frac{q.D_p}{L_p.\sinh(\frac{W_n}{L_p})}.\left[p_s.\cosh\left(\frac{W_n}{L_p}\right) - p_j\right] \tag{4.10}$$

However, in a real MISS device, the width of the base region (5 to 10 μm) will be considerably smaller than a hole diffusion length (100 to 1000 μm). As such $W_n \rightarrow 0$, $\cosh(W_n/L_p) \rightarrow 1$ and $\sinh(W_n/L_p) \rightarrow W_n/L_p$ which allows the above expressions to be reduced to;

$$I_{PJ} = I_{PJS} = q.\frac{D_p}{W_n}.(p_s - p_j) = q.\frac{D_p}{W_n}.\frac{n_i^2}{N_D}\left[\exp\left(\frac{V_J}{V_T}\right) - 1 - \exp\left(\frac{-\xi}{V_T}\right)\right] \tag{4.11}$$

The assumption implicit in this expression is that the base transport coefficient, β (the ratio of the collected to the injected current) is unity. This is equivalent to neglecting recombination in the neutral region.

A further simplification of the expression is possible if it may be assumed that the minority carrier density at the edge of the surface depletion region, p_s, is negligible. This approximation is valid in the off state where the degree of disequilibrium (and thus

the magnitude of ξ) is large but ceases to be the case as the on-state of the MISS is approached[3] . Thus, in the low current 'off' state at least;

$$I_{PJ} = q.\frac{D_p}{W_n}.\frac{n_i^2}{N_D}.\left[\exp\left(\frac{V_J}{V_T}\right) - 1\right] \tag{4.12}$$

Having made these assumptions, the hole current through the bipolar transistor is essentially only a function of;

(a) the level of injection which is proportional to $\exp(V_J/V_T)$ and

(b) the rate of diffusion through the neutral base layer which is proportional to ${W_n}^{-1}$.

4.3.3 Electron Diffusion Current.

The current of electrons diffusing away from the base-emitter junction into the p^+ emitter can also be obtained by solution of the diffusion equation. If the p^+material is taken as being semi-infinite, the limits for integration become, $n(x = \infty) = n_{p0}$ and n (depletion edge) $= n_{p0}.\left(\exp\left(\frac{V_J}{V_T}\right) - 1\right)$ giving;

$$I_{NJ} = q.\frac{D_n}{L_n}.\frac{{n_i}^2}{N_A}.\left[\exp\left(\frac{V_J}{V_T}\right) - 1\right] \tag{4.13}$$

However, for the low current levels corresponding to the off-state of the MISS, the injected current will be insignificant compared to the recombination current in the junction depletion region. Thus for the analysis of the off-state and probably most of the NDR state, the electron current at the pn junction will be described by equation (4.4). Equation (4.13) will only apply in the on-state of the MISS where the level of injection is high[3] .

The above equations describe an ideal pn junction where the doping profile is assumed to be a step function. In general, realistic junctions, whether epitaxial or diffused, will have graded doping profiles. In addition, in most realisations of the MISS device, there must exist an interface which cuts the plane of the junction and thus provides a possible leakage path. This might be represented electrically by a non-linear shunt resistance in parallel with the diode.

4.4 The MIS Junction

The theory of MIS junctions has been outlined in chapter 2 of this thesis. It has been shown that where there is sufficient communication between the metal and the semiconductor, the diode may be in a state of thermal disequilibrium when reverse biased. The concentration of minority carriers (holes in a m-i-s(n) diode) in the surface depletion region may be lower than the thermal equilibrium value because their rate of supply is semiconductor limited. This disequilibrium allows the depletion region to penetrate deeply into the semiconductor beyond the limit normally set by strong inversion under equilibrium conditions.

The ability of the deep depletion region to grow and support most of the applied voltage but then to contract again when the rate of supply of minority carriers is increased (forcing the condition back towards thermal equilibrium) is fundamental to the operation of the MISS device.

4.4.1 Direct Tunnel Currents.

Approximate expressions for the electron and hole tunnel currents were given in chapter 2. For values of reverse bias sufficient to cause $\psi_S \gg V_T$, a good approximation for I_{NT} was, from equation (2.23);

$$I_{NT} = -I_{NT_0}.\exp\left(\frac{V_{ins} - \phi_{Fn}}{V_T}\right) \tag{4.14}$$

where $I_{NT_0} = A_e.T^2.\exp(-\chi_e{}^{1/2}.\delta_{ins})$

It was also shown that the hole tunnel current may be treated in a first order approximation as;

$$I_{PT} = I_{PT_0}.\frac{p(0)}{N_V} \tag{4.15}$$

where $I_{PT_0} = A_h.T^2.\exp(-\chi_h{}^{1/2}.\delta_{ins})$ The hole tunnel current is thus proportional to the surface hole concentration which is given by equation (2.14);

$$p(0) = \frac{n_i{}^2}{N_D}.\exp\left(\frac{\psi_S - \xi}{kT}\right) \tag{4.16}$$

Here, ξ represents the separation of the minority and majority carrier quasi Fermi levels and as such is a measure of the degree of thermodynamic disequilibrium at the surface.

4.4.2 Generation in the Surface Depletion Region.

A full expression for the current, I_G, generated in the deep depletion region of a non-equilibrium MIS diode is given in equation (2.19). This expression describes equally the flux of electrons which drift out of the depletion region into the neutral n region and the flux of holes which drift to the IS interface.

4.5 Formulation of the Steady State Model

When brought into intimate contact to form a MISS device, the pn and MIS diodes are at once spatially and electrically connected in series. As such, two constraints are placed on the steady state;

(i) The hole current must be continuous. Thus, neglecting for interface assisted tunneling (I_{PT}^{ss}) for simplicity and assuming a unity transport coefficient for the neutral (base) region such that $I_{PJ}(x=0) = I_{PJ}(x=W_n)$;

$$I_{PT}(\xi, \psi_S) = I_G(\xi, \psi_S) + I_{PJ}(\xi, \psi_S, V_J) \tag{4.17}$$

(ii) The electron current must be continuous, which, again neglecting interface state current (I_{NT}^{ss}), implies;

$$I_{NT}(V_{ins}) = -I_G(\xi, \psi_S) + I_{RJ}(V_J) \tag{4.18}$$

These two equations, together with the expression for the surface field in the MIS, (B6);

$$\mathcal{E}_S = -\left(\frac{2.q.V_T.N_D}{\epsilon_s.\epsilon_o}\right)^{1/2}\left[\frac{\psi_S}{V_T} + \frac{p(0)}{N_D}\right]^{1/2} \tag{4.19}$$

and the charge neutrality condition, equation (2.8);

$$\epsilon_{ins}\epsilon_o.\frac{V_{ins}}{d_{ins}} = \epsilon_s\epsilon_o\mathcal{E}_S - Q_{ss} \tag{4.20}$$

constitute a complete description of the MISS system.

It was noted in section 4.2.2 that the 'S-type' I-V curve is current-controlled which implies that a unique solution for V_{TOT} must exist for any given current. In order to solve this set of equations, the voltage across the pn junction, V_J may be taken as the independent variable. Since the total device current is a monotonic function of V_J, this approach is tantamount to driving the MISS with a constant current source and calculating the voltage as the current is incremented.

Taking a particular value for V_J then, the four equations (4.17) to (4.20) contain three unknowns ξ, ψ_S and V_{ins} and can be solved numerically[3] . As V_J is incremented, the total voltage across the switch, given by the sum;

$$V_{TOT} = \psi_S + V_{ins} + V_J + V_{FB} \tag{4.21}$$

is obtained point by point as a continuous function of the total current through the device which is given by;

$$I_{TOT} = I_{PT} + I_{NT} + I_{PT}^{ss} + I_{NT}^{ss} \tag{4.22}$$

Moreover, the MISS voltage is obtained as a continuous function of the current throughout the switching characteristic, including the NDR region.

4.6 A General outline of the Switching Mechanism.

There has been a great tendency in the past for authors to attempt to obtain closed form analytical expressions for the principal characteristics of the MISS; I_S, V_S, I_H and V_H. However, in order to derive tractable expressions, such attempts necessarily require further approximations. As an example, for many years Simmons and Habib consistently only considered two modes of device operation corresponding to the limiting cases of punch-through of the surface depletion layer to the PN junction[1] and avalanche multiplication in the surface depletion field[9] . Experimental work has shown however that these limiting cases very often do not apply in practice. The switching process is far more general and can only be properly described by a numerical solution of the set of equations (4.17) to (4.22). Full numerical solutions have been obtained for tunnel oxide MISS devices by Habib (1981)[10] Fiore de Mattos (1986)[11] and more recently by Lavelle (1989)[3] at Durham. Each of these authors has been able to investigate the effects of important parameters such as (i) epilayer doping density, (ii) oxide thickness, (iii) metal work function and (iv) thickness of epilayer on the I-V characteristic.

Lavelle[3] has also enhanced the model to account for two dimensional effects which are important in real devices and will be discussed in some detail in chapter 5.

4.6.1 **The Off State.**

The off state of the device corresponds to a deeply depleted state of the MIS junction. Most of the applied bias is dropped across the deep depletion region and $V_{TOT} \approx \psi_S$. Inversion of the silicon surface is prevented by the leakage of minority carriers through the semi-insulating layer. As such, the minority carriers in the depletion region are not in thermal equilibrium. Their flux is controlled by their rate of supply from the semiconductor and the dynamics of the MIS system may be considered 'semiconductor limited'. The majority carrier current meanwhile is determined by the strength of the electric field at the semiconductor surface. This field is a function of both the surface potential and the equilibrium density of minority carriers at the surface. The proportion of the total device current due respectively to hole or electron tunnelling is strongly dependent on the work function of the metal and the doping of the epilayer[3] . The first determines the thermionic barrier faced by tunnelling electrons and the latter how strongly the surface field is affected by the reverse bias.

4.6.2 **The Switching Point.**

The switching point corresponds to the maximum depth of the depletion layer and also, as pointed out by Fiore de Mattos[2] , the maximum disequilibrium at the surface. It may be envisaged as the point where the detailed balance between the rates of supply and loss of holes from the surface is tipped in favour of the growth of an inversion layer. The cause of this shifting balance must be an increase in the surface hole flux above a certain threshold value. It is the physical source of this increased hole flux which determines the mode of operation of the switch. Possible sources include avalanche multiplication, punch through or illumination which gives rise to enhanced generation current. However, in the more general sense, the rise in hole flux above threshold need not be related to a definable 'event' such as these. It is just as possible that a gradual rise in hole injection from the pn junction should lead to the switching condition.

The switching condition is clearly closely linked to the ability of the minority carriers to escape through the semi-insulating layer. In general, it might be expected that a layer

with a higher impedance to minority current flow will allow the critical level of inversion charge to form at a lower current. Although the switching point does not in general coincide with the onset of strong inversion, it is recognised[12] that the existence of an inversion layer is a necessary pre-condition for switching to occur.

4.6.3 The NDR Region.

As the current is increased beyond the threshold value, the density of the inversion charge continues to increase and the deep surface depletion layer contracts. The MIS system starts to return to a state of thermodynamic equilibrium as the minority carrier Fermi level moves closer to that of the majority carrier.

Thus the total device potential decreases, yet the surface field continues to increase as a result of the growing inversion layer. These are the conditions that give rise to the negative differential resistance behaviour. Although the total voltage is falling, I_{TN} is able to increase due to the rise in $\mathcal{E}_S$ and I_{PT} increases due to the increase in $p(0)$. Thus;

$$\frac{\Delta I_{TOT}}{\Delta V_{TOT}} = \frac{\Delta I_{PT} + \Delta I_{NT}}{\Delta V_{ins} + \Delta V_j + \Delta \psi_S} < 0 \tag{4.23}$$

It is important to note at this point that the NDR effect in the MISS is not related to any filamentary conduction processes which are typical of NDR behaviour in amorphous materials[13-15] In the MISS, the phenomenon is purely the result of a redistribution of charge and potential within the device which occurs in a controlled manner. Solutions to the steady state set of equations which describe the MISS behaviour are obtained throughout the NDR region.

The device current is obtained as a continuous function of device voltage throughout the S-type NDR curve.

In addition, there is no suggestion in the formulation of the first order model that either the charge density or the current density should be anything other than homogeneous in the plane of the junctions. That is not to say, of course, that in real devices inevitable non-uniformities in material properties may not give rise to a degree of localisation of carriers in certain regions.

4.6.4 The Holding Point.

As the device current in the NDR regime is increased further, a point is eventually reached where either;

(a) the rise in current cannot be accomodated if the depletion region continues to contract or

(b) complete equlibrium is restored at the surface and the strong inversion condition, equation (2.10) sets the lower limit to ψ_S.

At this point, the surface depletion region can contract no further and the differential resistance given by (4.23) again becomes positive. This marks the holding point.

4.6.5 The On State.

The on state corresponds to a high level of inversion of the silicon surface. In this state, any increment in applied bias will fall almost wholly across the semi-insulating layer. The increased field in the layer will be terminated by an increased density of inversion charge, Q_{inv}, as discussed in Appendix B. As such, only a small increase in bias is required to cause a large increase in total current in the on state.

4.7 The Modes of Operation of the MISS Device

4.7.1 The 'Avalanche' Mode.

Conventional theory[9] suggests that switching in devices with relatively heavily doped ($> 10^{16}\text{cm}^{-3}$) epilayers corresponds to the onset of avalanche multiplication in the high field at the surface of the MIS depletion region. In this respect, the switching phenomenon is assumed to be very similar to 'first breakdown' in bipolar power transistors (avalanche breakdown of the collector-base junction). The increasing hole flux to the I-S interface which arises from avalanche multiplication can not be sunk by the tunnel diode and an inversion layer is able to develop, forcing the semiconductor surface back towards equilibrium.

However, in the theoretical study of Lavelle[3] it is shown that the influence of the avalanche effect is in fact quite insignificant compared to the effect of the high

surface field itself. From (B6) in Appendix B, $\mathcal{E}_S$ is proportional to $(N_D)^{1/2}$. For high doping densities, V_{ins} and hence I_{NT}, may consequentially be high for moderate applied voltage. A large electron current entering the pn junction gives rise to an even larger hole current in the direction of the MIS and the threshold condition for switching may be reached before the threshold field for avalanche multiplication. Thus the switching mechanism in MISS devices with highly doped epilayers is more accurately described as high field-induced.

4.7.2 The 'Punch-Through' Mode.

The switching mechanism in MISS devices with lightly doped epilayers has generally been attributed to punch-through (or reach-through) of the surface depletion layer to the pn junction depletion layer. Certainly, such an event would lead to a large increase in the minority carrier flux to the surface. The same mechanism is one of the principle breakdown modes in bipolar transistors. In this case, the switching voltage is approximately equal to the surface potential required to completely deplete the epilayer. This may be obtained from equation (4.1)for a M-i-n-p$^+$structure;

$$V_S(\mathrm{p-t}) \approx \frac{q.N_D.}{2.\epsilon_s.\epsilon_o}.W_{epi}^2 \tag{4.24}$$

In principle then, the switching voltage of these devices should be a function of the epilayer doping and thickness and quite independent of the semi-insulator properties.

However, Lavelle[3] shows that in low doped devices, switching occurs before the depletion region completely reaches through to the pn junction. The switching mechanism is rather one of enhancement of the minority carrier injection due to the contraction of the neutral base region between the depletion layers. In bipolar transistors, this phenomenon is called the 'Early' effect, and is due to the inverse relationship between I_{PJ} and W_n in equation (4.12);

$$I_{PJ} \propto \frac{1}{W_n} \tag{4.25}$$

4.7.3 The Generation-Controlled Mode.

As the applied bias on a MISS in it's off state is increased, the surface depletion region deepens and the amount of generation current arising from within it must increase.

This contribution to the hole flux at the I-S interface may be significant enough for it to exceed the threshold value before any of the other switching mechanisms becomes active. The generation current is likely to be most effective when the semi-insulating layer is highly resistive to minority carrier leakage. Also, it is obviously highly dependent on temperature and illumination. The strong effect of the latter on the switching point is the basis of the proposed application of the MISS as a latching photodetector for optical communications.

4.7.4 **The 'High Current' Mode.**

A further mode of switching, which arises particularly when the impedance of the semi-insulator to majority carrier flow is low, has been identified by Lavelle[3] . If a large flux of majority carriers enters the neutral base from the MIS junction, the pn junction is forced on and a larger flux of minority carriers (determined by the emitter efficiency) is returned, via the neutral base, to the I-S interface. Thus, although the threshold current for switching may be high due to the low impedance of the insulating layer, it may still be attained if the minority carrier injection from the pn junction is large.

4.7.5 **A Synopsis of the pn-MIS Junction Interaction.**

In discussing the various switching modes, the importance of the interaction between the MIS and pn junctions becomes clear. In a metal-tunnel oxide-n-p^+structure, the amount of electron tunnelling through the oxide determines the level of injection from the pn junction. This in turn determines the equilibrium hole density in the inversion layer and consequently the field in the oxide. Completing this sequence of interaction, the field in the oxide controls the electron tunnel current.

Consideration of this coupling of the two bipolar elements leads inexorably to a view of the MISS as a current-controlled regenerative feedback system. Each junction feeds the other with a flux of injected carriers which controls the flux of carriers of opposite type that it receives in return. Before going on to explore regenerative feedback as a mechanism which can account for the switching effect, it is important to make a clear distinction between the above steady state analysis which, by definition, assumes carrier distributions which are constant with time and the feedback analysis which, as will be discussed, is based on the concept of incremental changes in carrier distribution.

4.8 Regenerative Switching in the MISS.

4.8.1 Introduction.

Many of the principal workers in the field of MISS devices have analysed the problem in the steady state along the lines of the previous section but then proceeded, often in the same paper, to describe the switching process in terms of a regenerative feedback loop[1] [2] . There is indeed a great temptation to invoke the concept of current controlled regenerative feedback as a mechanism which can account for the 'S-type' NDR characteristic. Such an approach is widely accepted in thyristor theory[16] and, more recently, in the study of planar-doped barrier switches[17] and will also be used in this thesis as a tool for the analysis of effects of different device configurations. Consideration of regenerative feedback and the importance of the current gain allows experimental results to be accounted for, at least qualitatively, without recourse to full numerical solution of the steady state equations which is both time consuming and requires a great deal of computation. However, it is first important to consider how such a regenerative feedback model (RFM) should be constructed and what limitations are inherent in this analysis.

4.8.1 The Regenerative Feedback Model of the MISS.

Habib and Simmons[1] Zolomy[4] and Majkusiak et. al.[6] were perhaps the first authors to apply the RFM to the MISS. Borrowing from thyristor theory, they consider the MISS to embody two current gain elements, a bipolar transistor and a MIS diode as indicated in figure 4.3. In an epitaxial MISS device, the transistor comprises (i) the heavily doped substrate which constitutes the emitter, (ii) the neutral epilayer which is the transistor base and (iii) the MIS depletion region which behaves as the collector.

Significantly, in these analyses, the MIS is not considered to represent a transistor, probably because it is not obvious where the base region exists in this device. Some confusion has arisen in the literature as to how the current gains of these two components should be defined. Zolomy[4] , for instance, defines the gain of the transistor, α as the ratio of the hole current collected at the surface depletion edge to the total current through the pn junction;

$$\alpha_1 = \frac{I_{PJS}}{I_{PJ} + I_{NJ}} \tag{4.26}$$

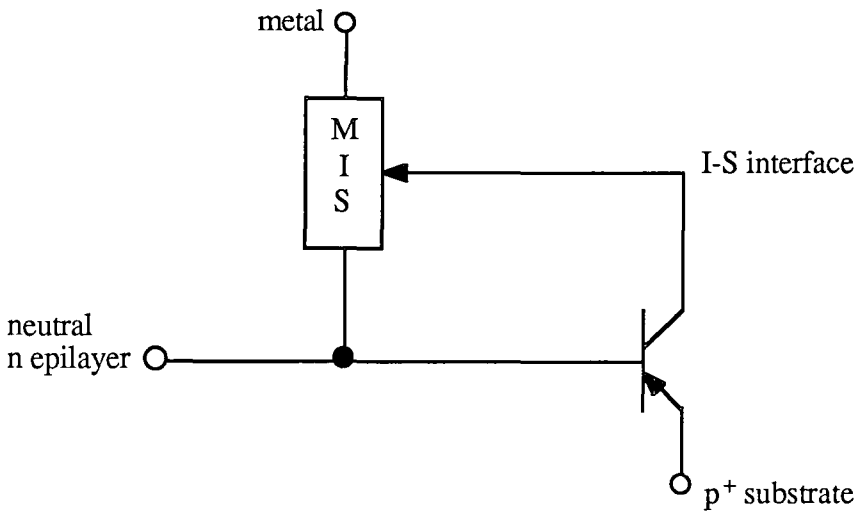


Figure 4.3; The 'thyristor' model of the MISS device as proposed by Zolomy [4] and Majkusiak et. al. [6]. Here the gain elements are configured as common base.

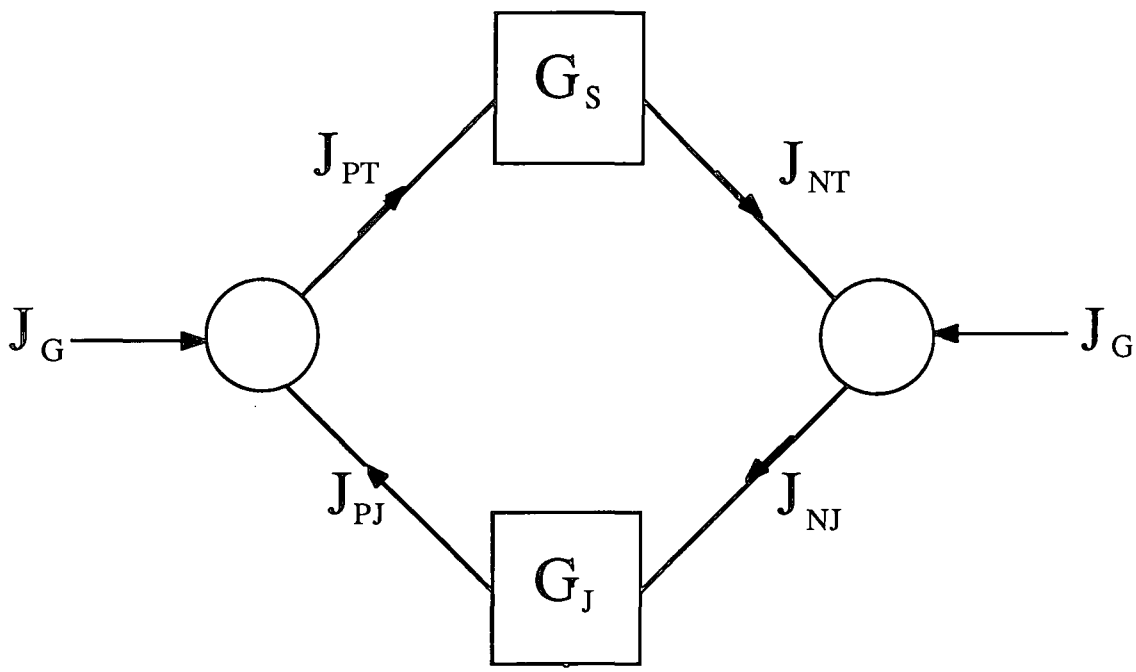


Figure 4.4; Definition of the MISS regenerative feedback loop using common-emitter gain elements (after Fiore de Mattos [11])

which is the common base gain of the bipolar transistor. The amplification factor for the MIS is defined similarly;

$$\alpha_2 = \frac{I_{NT}}{I_{NT} + I_{PT}} \tag{4.27}$$

The total current through the MISS is then given in terms of these amplification factors as;

$$I_{TOT} = \alpha_1 . I_{TOT} + I_G + \alpha_2 . I_{TOT} \tag{4.28}$$

which leads to;

$$I_{TOT} = \frac{I_G}{1 - (\alpha_1 + \alpha_2)} \tag{4.29}$$

where I_G is the current generated in the MIS depletion region and is taken to constitute the independent input current to the feedback circuit. Switching is then understood to occur when the sum of α_1 and α_2 becomes equal to unity, causing the device current to increase to infinity and thus become limited only by the external circuit.

More recently, Fiore de Mattos[2] has defined an equivalent loop gain expression based on his own definitions of the gains of the two elements;

$$G_J = \frac{I_{PJS}}{I_{NJ}} \qquad \text{and} \qquad G_S = \frac{I_{TN}}{I_{TP}} \tag{4.30}$$

It may be noted that these are equivalent to common-emitter expressions for the current gain of a bipolar transistor, often called h_{FE}, for the case of unity base transport coefficient. He also considers the generation current I_G to be the input to this loop as represented in figure 4.4 and then obtains the total current as;

$$I_{TOT} = \frac{(1 + G_J).(1 + G_S)}{1 - G_J.G_S} . I_G \tag{4.31}$$

This expression can be seen to be equivalent to (4.29) when it is noted that the common-emitter gains of Fiore de Mattos are related to the common-base gains of Zolomy by;

$$\alpha_1 = \frac{G_J}{G_J + 1} \qquad \text{and} \qquad \alpha_2 = \frac{G_S}{G_S + 1} \tag{4.32}$$

Although these analyses may be shown to be equivalent, they are defective in one important respect. They both consider direct current values of the gain (or multiplication) factors. It was recognized early in the development of thyristor theory that it is

not the D.C. but rather the small signal current gains[18] that determine the switching point. Unity small-signal loop gain is consistent with zero output impedance;

$$\left.\frac{\partial V}{\partial I}\right]_{V=V_S} \rightarrow 0 \tag{4.33}$$

which is the proper definition of the switching condition[18] . Unity D.C. gain, by contrast, is associated with infinite total current through the device[16] which is not a valid condition.

In general, the small signal common base gains α' and the common-emitter gains G' are defined as;

$$\alpha' = \lim_{\Delta I_E \rightarrow 0} \frac{\Delta I_C}{\Delta I_E} \qquad \text{and} \qquad G' = \lim_{\Delta I_B \rightarrow 0} \frac{\Delta I_C}{\Delta I_B} \tag{4.34}$$

As they are generally appreciably greater than the large-signal or D.C. gains, switching will be initiated before $\alpha_1 + \alpha_2 \rightarrow 1$ or $G_S.G_J \rightarrow 1$. It is critical to the switching process that the gain of one or both of the elements should be enhanced as the total current through the device is increased. Otherwise the unity loop gain condition will never be attained. Whereas the gain of a BJT will in general increase with the total current (due mainly to the improving emitter efficiency), the same trend has yet to be shown for the MIS gain.

4.9 A Proposed Model of the MISS Consistent with Thyristor Theory.

All the regenerative feedback theories so far proposed for the MISS consider the device to consist of a bipolar junction transistor (BJT) with floating base and a two terminal MIS diode. In respect of the latter, they fall short of a full thyristor analogue because this requires the coupling of two three-terminal gain elements. Where three terminals may be identified, a rigorous definition of current gain is possible. Where only two terminals exist, as in the MIS diode, the theoretical treatment of the current gain becomes more vague.

Board[7] in a recent review of regenerative feedback devices takes the course of treating the MIS as a charge storage element. He explicitly avoids considering it as a

gain element and excludes it from any activity in the regenerative feedback mechanism. Its function, he asserts, is to provide the possibility of both a stable low current, high voltage (off state) regime and a stable high current, low voltage (on state) regime for the MISS. These two regimes are dependent on the amount of minority carrier charge stored at the IS interface.

Zolomy[19] in another persuasive review of 'new bipolar amplifying and negative-resistance devices' also has problems dealing with the MIS diode as a current gain element. In his discussion, he quite rightly states that a bipolar current amplifier must consist of (a) a forward biased emitter-base junction which provides current amplification and (b) a reverse biased collector-base junction whose depletion region performs the function of extracting the minority (output) current from the majority (input) current in the base. He also recognizes that the surface depletion region of the MIS performs the dual role of minority carrier collection for both the bipolar transistor and the MIS diode. He is vague, however in his definition of the MIS gain, again because no base is apparent in the MIS structure.

4.9.1 **The MIS as an 'Inversion Base Bipolar Transistor'.**

It is now proposed that the MIS structure itself becomes a transistor as the switching point is approached. Although the MIS is simply a diode in most of the off-state of the MISS, it is transformed into a transistor when an inversion layer starts to form at the semiconductor surface. The inversion layer behaves like the base region in a bipolar transistor and what was a two-layer MIS diode becomes in effect a three-layer MIS emitter transistor.

The concept of a bipolar transistor whose base region consists of an inversion layer is not new. Such a transistor, shown in figure 4.5 was invented by Taylor and Simmons[20] at Bell Labs in 1985 under the acronym BICFET, meaning bipolar inversion channel field-effect transistor. This device represents the most significant recent advance in transistor technology, providing for very fast performance through virtual elimination of base transit time and charge storage delays. The same transistor was subsequently re-invented in a slightly different guise by a group at MITI, Japan[21] under the acronym IBT for inversion-base bipolar transistor. The Bell group proposed both Si- and GaAs-based variants of the BICFET and have since realised several III-V forms of this device.

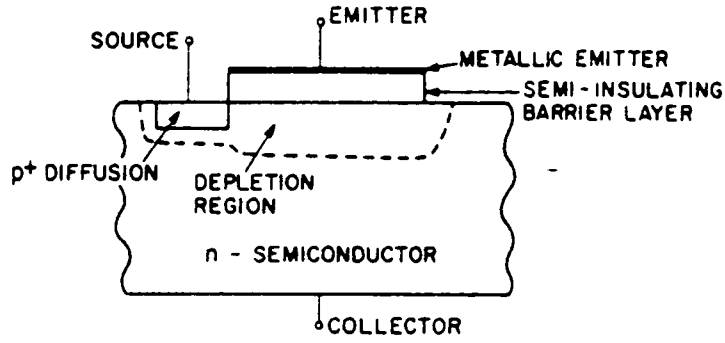


Figure 4.5; Schematic cross-section of a BICFET after Taylor and Simmons [20]

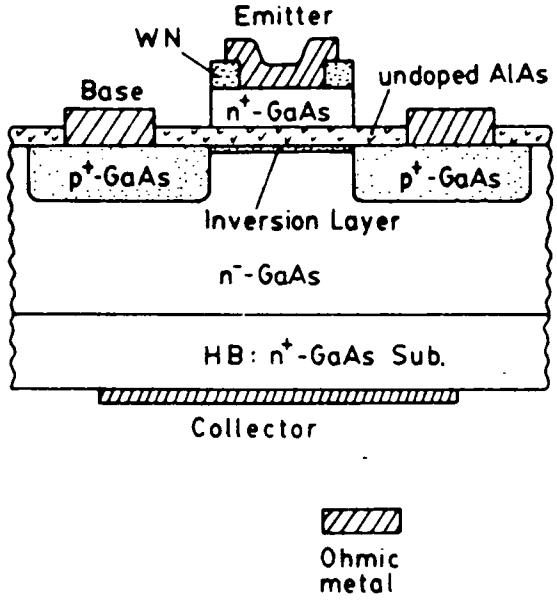


Figure 4.6; Cross section of an IBT after Matsumoto et. al.[21]

The MITI group formed the device on GaAs, using undoped AlAs as the semi-insulating layer.[21]

For the purposes of the following discussion, the term IBT is preferred because it implies the role of the inversion layer as the base in the bipolar transistor. The term BICFET has connotations of field effect transistor which implies unipolar conduction in the inversion layer and as such may be misleading. Indeed, Simmons and Taylor in their theoretical treatment of the BICFET refrain from calling the inversion layer a base, preferring instead to refer to it as a source.

Fundamental to the operation of the IBT is the provision of electrical contact to the inversion layer to provide a third input terminal. This is achieved in all the reported work by self-aligned implantation of doped regions around the semi-insulator semiconductor interface, as indicated in figures 4.5 and 4.6. No such access to the inversion base is available in the usual MISS structure and it must therefore be considered as floating. Its supply of minority carriers (the input current of the IBT) cannot be controlled by an external source and is dependent on the internal current flows of the MISS.

The small signal current gain of an IBT in the common-emitter arrangement is simply;

$$G'_{IBT} = \frac{\partial I_C}{\partial I_B} \tag{4.35}$$

For the present case, where the IBT is in the form of a metal tunnel-oxide silicon structure, the collector current I_C is composed wholly of the electron tunnel current I_{NT} (under the reasonable approximation of a unity base transport coefficient).

Because the inversion base is floating, the only provision of base current is from (a) generation in the surface depletion layer or (b), as the switching point is approached, injection of minority carriers from the p^+-n emitter junction. Whatever the source of holes, they all leak away through the thin oxide by tunnelling (neglecting surface recombination) and it is possible to state to a first approximation $I_B = I_{PT}$.

For the sake of completeness, the emitter current is then $I_E = I_C + I_B = I_{NT} + I_{PT}$ which is consistent with equation (4.28) for the total current through the MISS. The gain factor of the tunnel oxide IBT is then;

$$G'_{IBT} = \frac{\partial I_{NT}}{\partial I_{PT}} \tag{4.36}$$

which is comparable to the expression for the D.C. gain of the MIS diode, equation (4.30), given by Fiore de Mattos[11] in his formulation of the regenerative model. The result is the same but the new approach suggested here, which recognizes the existence of a three layer transistor structure in the MIS section, is conceptually more satisfying and leads to a better description of the MISS as an analogue of the thyristor.

4.9.2 The Small-Signal Gain of a Tunnel Oxide IBT

A simple **low frequency** small-signal gain expression may be obtained for the IBT using the steady state equations for the MISS given in sections 4.3 and 4.4. The expression for G'_{IBT}, equation (4.36), may be divided into four partial derivatives;

$$G'_{IBT} = \frac{\partial I_{NT}}{\partial I_{PT}} = \frac{\partial I_{NT}}{\partial V_{ox}} . \frac{\partial V_{ox}}{\partial \mathcal{E}_S} . \frac{\partial \mathcal{E}_S}{\partial p(0)} . \frac{\partial p(0)}{\partial I_{PT}} \tag{4.37}$$

These partials may then respectively be obtained by differentiation of expressions (4.14), (4.20), (4.19)and (4.15);

$$\begin{aligned} \frac{\partial I_{NT}}{\partial V_{ox}} &= \frac{I_{NT}}{V_T} \\ \frac{\partial V_{ox}}{\partial \mathcal{E}_S} &= \frac{\epsilon_s}{\epsilon_{ox}} . d_{ox} \\ \frac{\partial \mathcal{E}_S}{\partial p(0)} &= \frac{q.V_T}{\epsilon_s . \epsilon_o} . \frac{1}{\mathcal{E}_S} \\ \frac{\partial p(0)}{\partial I_{PT}} &= \frac{N_V}{I_{PT_0}} \end{aligned} \tag{4.38}$$

which leads to;

$$G'_{IBT} = \frac{q.N_V}{I_{PT_0}.C_{ox}} . \frac{I_{NT}}{\mathcal{E}_S} \tag{4.39}$$

The main features of this expression are;

(i) The gain increases with increasing I_{NT} and therefore with increasing total current.

(ii) The gain improves for lower oxide capacitance and lower hole leakage by tunnelling. Both of these factors suggest gain should be enhanced by using a thicker oxide layer or an alternative metal-semiinsulator combination that presents a larger transmission barrier to holes than to electrons.

(iii) To maximise the gain, the surface field should be minimised. Expression (4.19) suggests this is best achieved by using a low-doped epilayer and reducing the inversion layer hole concentration, which is inherent in (ii).

4.9.3 The Regenerative Feedback Loop Gain.

In order to establish a loop gain expression, the small-signal versions of the common emitter current gains should be used, as discussed in section 4.8.1 The feedback mechanism of Fiore de Mattos[2] may be restated as the change in the total current ΔI_{TOT} which results from a small incremental change in the independent input current, ΔI_G;

$$\Delta I_T = \frac{(1 + G'_{IBT}).(1 + G'_{BJT})}{1 - G'_{IBT}.G'_{BJT}}.\Delta I_G \tag{4.40}$$

The switching condition may then be stated as;

$$G'_{loop} = G'_{IBT}.G'_{BJT} = 1 \tag{4.41}$$

It is interesting to note that there is some agreement between this expression and the steady state description of the device. Switching occurs in the steady state when the surface potential, ψ_S reaches a maximum. Since I_G is directly related to the surface potential through equation (2.17), it thus follows that I_G is also a maximum and as such, $\Delta I_G = 0$ at the switching point. By evaluating the loop gain at V_S, it can be seen that it does indeed attain unity at V_S. Substituting for I_{NT} and I_{PT} from equations (4.17) and (4.18), G'_{IBT} becomes;

$$G'_{IBT} = \frac{\Delta I_{RJ} - \Delta I_G}{\Delta I_{PJ} + \Delta I_G} \tag{4.42}$$

giving for the loop gain;

$$G'_{IBT}.G'_{BJT}\Big]_{V=V_S} = \frac{\Delta I_{RJ} - \Delta I_G}{\Delta I_{PJ} + \Delta I_G}.\frac{\Delta I_{PJ}}{\Delta I_{RJ}} \tag{4.43}$$

which equals one at the switching point where $\Delta I_G = 0$.

4.9.4 **Limitations of the Regenerative Feedback Model.**

Although the regenerative feedback model (RFM) can account for the switching action and may be used to define the switching point, it is rather limited in its range of application for several reasons;

(i) The RFM gives little insight into the behaviour of a switching device after it has turned on. The only information it provides concerning the NDR part of the switching characteristic is that the loop gain exceeds unity in that region, making it inherently unstable. The gain is then thought to return to a value less than unity as the on-state is entered, at which point stability is restored.

(ii) The RFM is defined independently of the physical processes which are occurring in the switching device. It must be recognized therefore that although this model may be applied to various forms of regenerative switching device (*e.g.*the thyristor, triangular barrier switch or any variant of the MISS), there is no implication that the physical mechanisms which give rise to the switching effect are at all related. In the thyristor, the transition from the off to the on state corresponds to the saturation of a central reverse biased pn junction with high level injection of electrons and holes from neighbouring forward biased pn junctions. In the MISS, this transition corresponds to a decline in the potential dropped across the surface depletion region as the MIS moves towards thermal equilibrium and the inversion charge increases at its interface. In a triangular barrier switch, the NDR region corresponds to the reduction in size of a double depletion region due to the presence of an increasing density of compensating free-carrier charge[7] . What all these mechanisms do have in common however, is the phenomenon of a large increase in the free-carrier (stored) charge internal to the device as the on-state is approached.

(iii) As described in this chapter, the RFM is, like the steady state model, only a one-dimensional approximation. As such, no consideration has been given to the efficiency of interaction between the two gain elements. which has been implicitly treated as unity. In a real epitaxial MISS device, the area of the pn junction will be somewhat larger than that of the MIS junction. Thus not all the minority carriers injected by the p^+-n emitter will be collected by the MIS surface depletion region. The question of collector efficiency will be addressed later in the next chapter on experimental two-dimensional MISS devices.

(iv) Although the switching phenomenon in MISS devices has been explained in terms of a small-signal regenerative feedback model, no account has been taken of frequency effects. In particular, the reduction in gain that must be expected in any amplifying element at high frequencies has not been considered. The inability of the feedback mechanism to react above a certain 'cut-off' frequency will become apparent in some surprising experimental results to be presented in chapter 7.

4.10 Conclusions

The theory of operation of the MISS device has been introduced in this chapter in terms of both steady state and regenerative feedback analyses. The MISS has been shown to be directly analogous to a p-n-p-n thyristor in that it consists of two gain elements, a bipolar junction transistor and an inversion base transistor. In each of these elements, it is possible to identify;

(a) A forward biased emitter-base junction which provides current gain proportional to the efficiency with which it injects minority carriers into the base layer. This takes the form of a p^+-n junction in the BJT and a metal-insulator-inversion layer junction in the IBT.

(b) A reverse biased base-collector junction which separates the minority carrier (output) current from the majority carrier (input) current in the base layer. In the MISS, the same depletion region acts as the collector for both the BJT and the IBT.

Furthermore it has been shown that the small signal gains of both these elements may be expected to increase with increasing device current which is a necessary condition for switching.

As a final point, it is interesting to note that a failure to recognize the existence of a four-layer structure in the conventional MISS has led some authors to propose unnecessarily complex alternatives. For example, Zólomy[19] suggests that a M-I-n-p-n^+ structure is required to properly imitate the thyristor and such a device has been successfully produced[22] . The theory advanced in this chapter shows the extra layer to be superfluous.

REFERENCES

1. S.E-D.Habib and J.G.Simmons, Solid State Electronics **22** 181 (1979)
2. A.C.Fiore de Mattos and G.Sarrabayrouse, phys. stat. sol. (a) **87** 699 (1985)
3. S.Lavelle, Ph.D thesis, University of Durham, U.K. (1989)
4. A.Adan and I.Zólomy, Solid State Electronics **23** 449 (1980)
5. J.G.Simmons and G.W.Taylor, Solid State Electronics **29** 287 (1986)
6. B.Majkusiak, A.Jakubowski and J.Ruzyłło, phys. stat. sol. (a) **60** K171 (1980)
7. K.Board, Phys. Rev D. **12** 1595 (1985)
8. C.T.Sah, R.N.Noyce and W.Schockley, Proc. IRE **45** 1228 (1957).
9. S.E-D.Habib and J.G.Simmons, Solid State Electronics **23** 497 (1980)
10. S.E-D.Habib, PhD Thesis, University of Toronto, Canada (1981)
11. A.C.F.Fiore de Mattos, These d'Etat No. 1158, Université Paul Sabatier, Toulouse, France (1986)
12. G.Sarrabayrouse, J.Buxo, A.E.Owen, A.Munoz Yague and J.P.Sebaa, Proc. IEEE, Part I **127** 119 (1980)
13. S.R.Ovshinsky, Phys. Rev. Lett. **21** 1450 (1968)
14. A.D.Pearson, IBM J. Res. Develop. **13** 510 (1969)
15. A.E.Owen, P.G.LeComber, W.E.Spear and J.Hajto, J. Non-cryst. Sol. **60** 1273 (1983)
16. J.L.Moll, M.Tanenbaum, J.M.Godley and N.Holonyak, Proc. IRE, **44** 1174 (1956)
17. S.E-D.Habib and K.Board, IEE Proc.**I 130** 292 (1983)
18. W.Fulop, IEEE Trans. El. Dev. **ED-10** 120 (1963)
19. I.Zolomy, Solid State Electronics **28** 537 (1985)
20. G.W.Taylor and J.G.Simmons, IEEE Trans. El. Dev. **ED-32** 2345 (1985)
21. K.Matsumoto, Y.Hayashi, N.Hashizume, T.Yao, M.Kato, T.Miyashita, N.Fukuhara H.Hirashima and T.Kinosada, IEEE El. Dev. Letters **EDL-7** 627 (1986)
22. D.C.Y.Chang, Chung.L.L and Tan.F.L, Solid State Electronics **32** 179 (1989)

CHAPTER FIVE

Epitaxial Tunnel Oxide MISS Devices.

5.1 Introduction

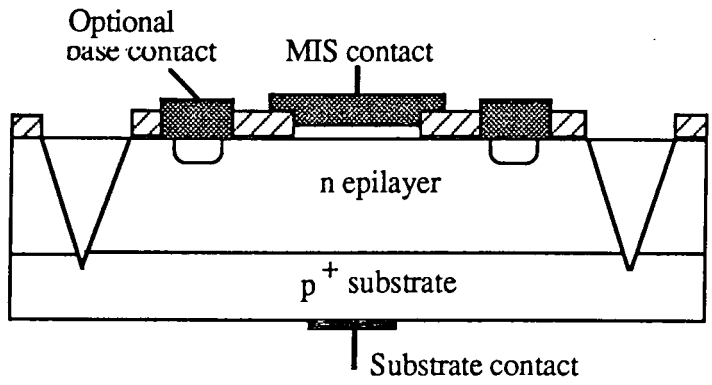
Fundamental to all the proposed applications of the MISS in the field of microelectronics, for example as a memory element[1] or as an active load in digital gates[2] , is the ability to scale the device to ever smaller dimensions. To be useful in the high density, low power integrated circuits of the present day, this implies dimensions of the order of 1μm unless the device is limited to strictly peripheral functions, for example as a bond pad driver or high tension input protection device.

Despite this implicit requirement, very few studies of MISS devices have been concerned with the consequences of reducing the MISS area. A notable exception is the work of Faraone et. al.[3] who made a systematic study of the effects of varying the pn junction area for a given MIS diode area. They established that the switching characteristics were very strongly influenced by area effects to such an extent that for a given n-p$^+$ epitaxial substrate and tunnel oxide, the characteristic could vary from the case $V_H > V_S$ (the device exhibits no NDR or on state) for a small MIS:pn junction area ratio to the case $V_S < V_H$ (no off state) for a large area ratio. As such, geometrical considerations generally override the effects of other device parameters.

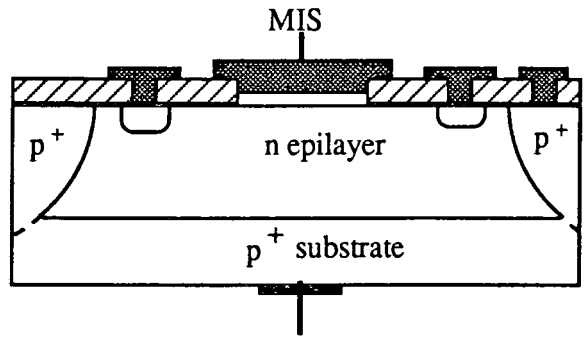
In order to investigate the effect of the junction area ratio in the present work, a means had to be established to define the extent of the epitaxial pn junction. Apart from this additional requirement, the processing required for the fabrication of epitaxial MISS devices is identical to that described for the MIS diodes in the previous chapter. There are several options for achieving electrical isolation of an epitaxial junction, three of which are shown diagramatically in figure 5.1;

(a) Etching of V-grooves to a depth greater than the epilayer thickness provides well defined, physically isolated mesas on which the MISS may be formed. This technique has been applied by several authors[3] but has the great disadvantage of creating a highly undesirable morphology for subsequent processing.

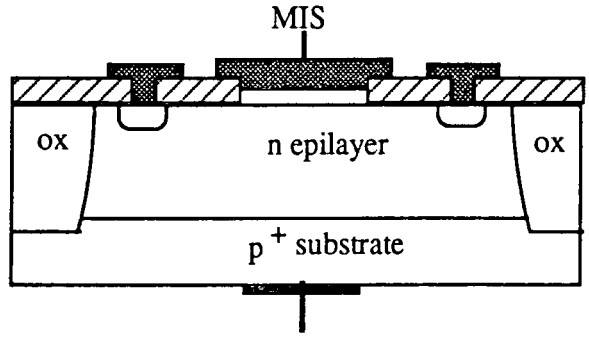
(b) Deep implantation of acceptor ions to create an isolated n-well, surrounded by p$^+$ material. This method has been applied by Chang et. al.[4] and has the advantage



(a)



(b)



(c)

Figure 5.1; Possible methods of isolating an epitaxial junction;
(a) deeply etched V-grooves
(b) deep implantation of acceptors to create an 'n-well'
(c) Implantation of oxygen ions to form an oxide wall.

of maintaining a planar surface.

(c) Implantation of oxygen ions to form an n-well surrounded by an insulating barrier of oxide. To the authors knowledge, this approach has not been implemented in the study of MISS devices but is certainly under development for application in integrated circuit fabrication.

Since the second two approaches require high energy ion implantation, a technique which was not readily available at the time of these studies, the first method which makes use of etched grooves in the substrate was developed.

5.2 Etching of V-grooves in {100} Silicon

The principal requirements of an etch to provide controlled etching of V-grooves in {100}-oriented silicon substrates are;

(i) A high selectivity for removal of {100} planes with respect to {111}-oriented crystal planes. Such selectivity allows the desired 'V' shaped groove profile to be defined in which the sloping side walls correspond to {111} planes which are inclined at 54.74° to the {100} as indicated in figure 5.2.

(ii)A high selectivity between the etch rates for the {100}-oriented silicon and a suitable masking material such as native oxide.

Several chemical etches have been developed to provide preferential etching of the {100} orientation together with a very low rate of attack of the masking layer and have been reviewed by Bean[5] . In general, anisotropic etches are solutions of a base (NaOH, KOH, ethylenediamine or hydrazine) and a compexing agent (IPA or pyrocatechol) in water[6] . Two standard etches are 65% hydrazine in water (which does not require a complexing agent)[6] and 5 : 4 : 16 KOH : IPA : H_2O [7] which offers excellent selectivity (etch rate ratio 100 : 1 : 0 for {100} Si : {111} Si : Si_3N_4).

However, the first presents an undesirable safety hazard and the second makes use of a potassium compound, the presence of which is best avoided in any silicon process where an oxide is critical to the functioning of the device. Therefore an alternative etch first used by Finne and Kleine[8] has been adopted.

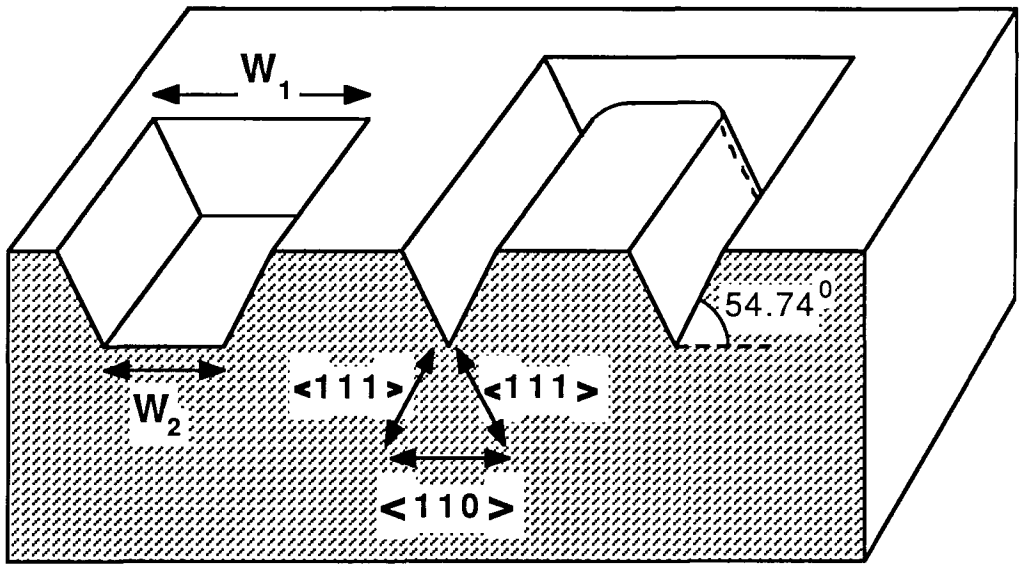


Figure 5.2; Definition of an etched V-groove and isolated mesa in a {100}-oriented silicon substrate.

The etch used consists of ethylenediamine (EDA) ($NH_2(CH_2.CH_2)NH_2$), pyrocatechol ($C_6H_4(OH)_2$) and water and is found to give the best selectivity for the following mixture;

35 mole%	Ethylenediamine	(1,2-Diaminoethane hydrate)
3.7 mole%	Pyrocatechol	(o-dihydroxybenzene)
61 mole%	Water	

This etchant also has the advantage that it almost ceases etching when it reaches highly doped p^+ silicon. As such, the p^+ substrate of the epitaxial material used in this study should provide a natural etch stop.

There are several important aspects of the use of any of these anisotropic etches for defining rectilinear V-grooves or mesas in {100} silicon;

(i) The desired pattern must first be defined in a suitable masking layer. In the case of the EDA etch, SiO_2 offers very good resistance, being removed at only about 8Å/min[8] .

(ii) The edges of features to be etched must be well aligned to the orthogonal $< 110 >$ axes on the {100} surface, as indicated in figure 5.2. Otherwise, a unique {111} atomic plane will not be etched on each side wall and a stepped morphology[9] will result with undercutting of the mask in proportion to the degree of misalignment.

(iii) Where etched features have convex corners, quite severe undercutting of these corners will result due to the relatively fast removal of exposed {331} planes[5] and the corners will be rounded. This effect may be offset by adding 'lobes' to any convex corners in the desired pattern which then behave as 'sacrificial' etch material.

(iv) The etching action may be inhibited by localised surface or bulk defects and by residual hydrous silica where this is not removed rapidly enough. Where this occurs, square pyramids of unetched silicon will remain. It is therefore necessary to ensure the surface is well cleaned and dry prior to etching.

5.2.1 **Characterisation of the V-groove Etching Process**

Several etch tests were conducted to establish the efficacy of the EDA mixture in defining the desired isolation grooves. A {100} silicon wafer (n-type) was scribed parallel and perpendicular to the flat edge which corresponds to a $< 110 >$ axis and then broken into quarters. These were cleaned using the standard process described in chapter 3 and then oxidised for 40 mins in dry oxygen which provided a 325Å SiO_2 layer as determined by ellipsometry. A suitable test mask which included a variety of slot and mesa features was carefully aligned to the cleaved edges ($< 110 >$ axes) and printed in AZ1350 resist. Then the pattern was transferred to the SiO_2 masking layer using 4:1 buffered HF and the resist stripped off. The EDA etch solution was prepared in a flask fitted with a condenser to prevent vapour loss and heated to 100° C by immersion in boiling water. The samples were then given a final dip in 6% HF to ensure complete removal of surface oxide in exposed areas and blown dry. After a 10 min etch in the EDA, they were rinsed well in deionised (d.i.) H_2O and dried from IPA. The mask was later removed by a short etch in buffered HF.

Inspection of the etched samples in a scanning electron microscope revealed that;

(i) Etched areas were littered with small square pyramids of unetched silicon, despite the precautions taken,

(ii) Features with concave corners had well-defined 90° angles, figure 5.5 , despite the relatively poor definition of sharp angles achieved by photolithography (diffraction effects at the mask-making and contact printing stages tend to round off square features). In this respect the anisotropic etch makes a marked improvement in the final definition of rectilinear V-grooves.

(iii) Convex corners were considerably rounded to the extent that the mesas ceased to appear rectangular as shown in figure 5.5. Although this result is somewhat displeasing, it was felt that so long as the receding corners did not encroach on the MIS windows which were to be defined at the mesa centres, the situation was tolerable. The intent here is merely to define the extent of the epitaxial junction in MISS devices. It was therefore concluded that the option of including sacrificial lobes at mesa corners would not be taken. This measure would introduce the additional complication of ensuring their complete removal.

5.2.2 Measurement of Groove Depth

The depth of etched V-grooves, whether they be flat-bottomed or not, may be easily determined by measurement of their lateral dimensions. Using an optical microscope with a calibrated eyepiece graticule, the widths of the tops and bottoms of several grooves (W_1 and W_2 in figure 5.2) were measured by judicious re-focussing and their depths, D, obtained from the geometrical relation;

$$\frac{D}{(W_1 - W_2)/2} = \tan 54.74^\circ = 1.414 \tag{5.1}$$

By this means the etch rate was established to be $0.79 \pm 0.02\mu$m/min which is in fair agreement with Bean[5] who obtained a value of 1.1μm/min.

5.2.3 Photolithography on a Grooved Topology.

In initial experiments to test the standard photolithography process on surfaces with deep V-grooves, it soon became apparent that their presence poses severe problems. The Shipley Microposit 1350 resist used in the Durham Microelectronics laboratory, which spins on to a thickness of only 0.5μm, provided quite unsatisfactory coverage as can be seen from the micrograph in figure 5.3. Here the resist has withdrawn from the groove edges, presumably under the influence of surface tension, causing discontinuities in the pattern defined in it. In addition, the grooves have been filled with resist which has then failed to be fully developed and thus remains where it should have been removed.

An attempt was made to improve the coverage by dip-coating the resist rather than spinning it on. Although this approach did provide complete coverage right up to the groove edges, it introduced the additional problem of poor thickness uniformity which is evident in the photomicrograph of figure 5.4. Also, handling difficulties arose at the next process step, the resist soft bake, due to the reverse faces of the wafers also being coated.

The coating problem was finally resolved by procuring a different photoresist known as TF20 from the same manufacturer. This product has a much greater solids content and is much more viscous than the Microposit 1350. It therefore spin coats to a considerably greater depth, around 4-5μm and proved to provide quite satisfactory coverage.

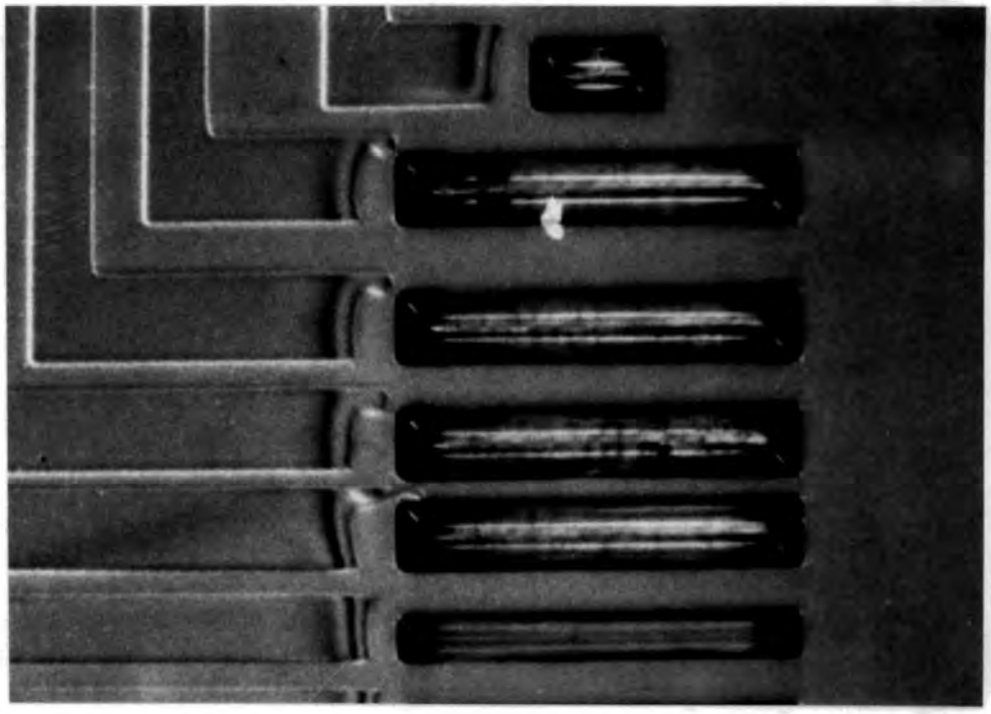


Figure 5.3; Scanning electron micrograph of a pattern defined in Shipley 'Microposit' 1350 photoresist showing track discontinuities caused by withdrawal of resist from the V-groove edges.

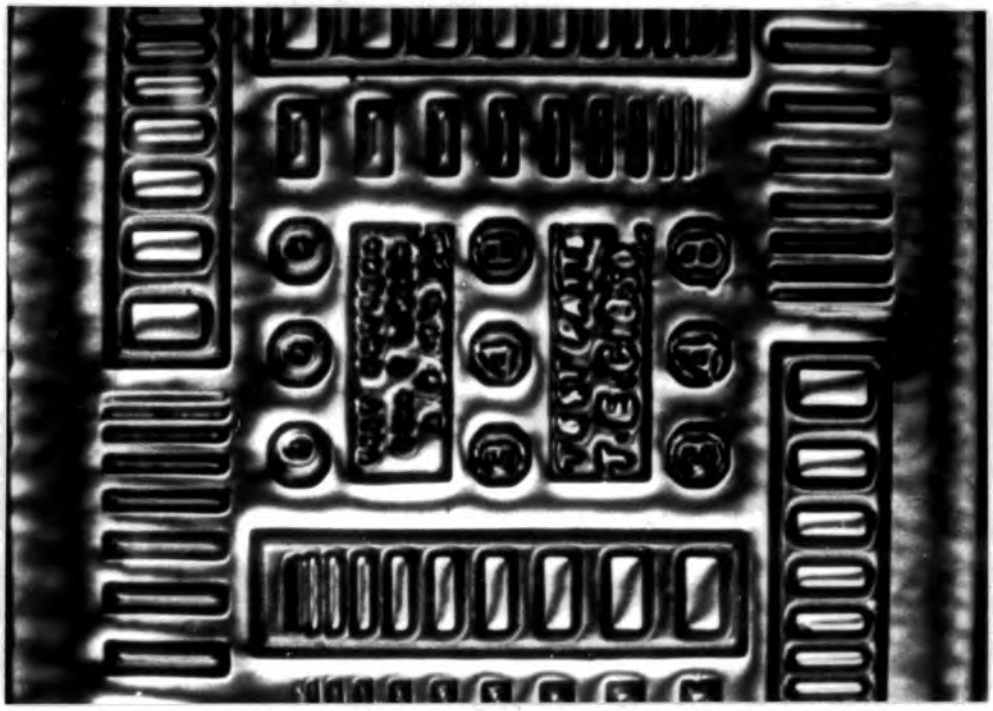


Figure 5.4; Nomarski photomicrograph of a grooved surface dip coated with 1350 resist, showing the poor thickness uniformity achieved.

However there did remain the problem of poor uniformity, with thicker resist in the centre of mesas thinning towards the groove edges. This effect has two consequences;

(i) When a metal track is required to pass over an isolation groove, the track width is reduced at the groove edge due to over exposure of the resist which is thinner at that point. In severe cases, there is a possibility of a break in the track.

(ii) Conversely, the size of a MIS window defined in the centre of a mesa depends on the resist thickness in its locality. In particular, where the mesa is small, as for switch no.12 on the MISS test chip (see appendix C), the resist tends to be thinner and the MIS window becomes enlarged.

In spite of these problems, this V-groove process was thought to be adequate for the purposes of this study where yield was not a crucial factor.

5.2.4 **Fabrication of Isolated Epitaxial MISS Devices**

The process used for fabrication of V-groove isolated MISS devices, summarised in table 5.1, was basically an extension of the one established for the processing of MIS diodes. The only additional steps required were (a) definition of the isolation groove pattern in a masking layer of oxide using mask #1 (Appendix C) and (b) etching of the grooves to the desired depth and removal of the mask. Subsequent to etching the grooves, a thick layer of 'field' oxide was grown to passivate the surface and provide the necessary isolation for the metal interconnects and bond pads.

The starting material for these devices was epitaxial {100} n-on-p^+ silicon with a nominal epilayer thickness of 4.5μm and resistivity at room temperature of 5 to 8 Ω.cm (corresponding to a doping of about 1.5 $10^{15}cm^{-3}$). The substrate resistivity was of the order of 0.01Ω.cm. On both wafers a 4×8 array of chips was defined with each chip containing 8 isolated and 1 unisolated discrete epitaxial switches.

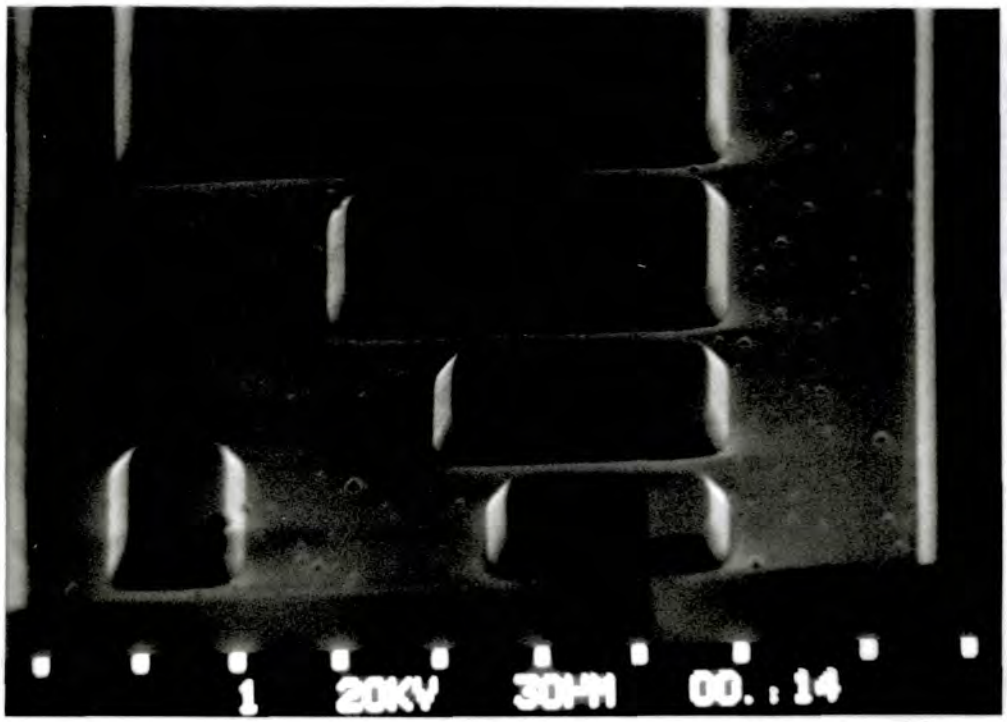
The scanning electron micrographs in figure 5.5 (a) and (b) show examples of completed isolated devices corresponding to the series 9-12 (varying pn junction areas) and 14-17 (varying MIS areas) as defined in figure C.1.

Step	Description	Details	Assessment
1	Scribe {100}Si wafer into quarters	‖ and ⊥ to flat	
2	Degrease in 1,1,1-trichloroethane	10 mins boiling; repeat	
3	Etch in sulphuric 'bomb'	20 mins in boiling 1:1 H_2SO_4:H_2O_2	
4	Rinse in re-circulating d.i.H_2O	Until resistivity of effluent > 10MΩ	
5	**Grow 3-400Åmasking oxide**	40 mins dry ox. @ 1000°C	d_{ox} by ellipsometry
6	**Print mask #1** (isolation pattern)	AZ1350 resist; expose 12 secs (align ‖ to a scribed edge)	
7	Hard bake resist	20 mins on hotplate @ 120°C	check linewidths
8	**Etch masking oxide**	Immerse in BHF† until hydrophobic	
9	Strip resist	Proprietary stripper and acetone.	
10	Rinse in re-circulating d.i.H_2O	Until resistivity of effluent > 10MΩ	
11	Dip in 6% dilute HF	3-4 secs then quick rinse	
12	**Etch V-grooves in silicon surface**	12 min EDA etch @ 100°C (see text)	Check through to p^+; thermoelectric probe
13	Rinse well in re-circ. d.i.H_2O	Until resistivity of effluent > 10MΩ	
14	Remove masking oxide	BHF etch until hydrophobic	
15	Rinse well in re-circ. d.i.H_2O	Until resistivity of effluent > 10MΩ	
16	**Grow field oxide**	240 mins wet ox. @ 1000°C	check colour
17	**Print mask #4**	TF20 resist; expose 50 secs	
18	Hard bake resist	20 mins on hotplate @ 120°C	check linewidths
19	**Etch MIS windows**	Immerse in BHF† until hydrophobic	
20	Strip resist	Proprietary stripper and acetone.	
21	Rinse well in re-circ. d.i.H_2O	Until resistivity of effluent > 10MΩ	
22	Dip in 6% dilute HF just prior to next step	3-4 secs then quick rinse in H_2O and blow dry from I.P.A	
23	**Grow tunnel oxide**	Dry ox. @~750°C for required time	d_{ox} by ellipsometry
24	**Evaporate aluminium**	Load samples direct from furnace When base pressure < 10^{-7}Torr 5 × 8 sec bursts, e-beam @ 400mA	
25	**Print mask #6**	TF20 resist; expose 60 secs	
26	Soft bake resist	20 mins on hotplate @ 85°C	align & linewidths
27	**Etch aluminium**	HPO_4 until visibly cleared	check linewidths
28	Strip resist	Proprietary stripper and acetone.	
29	Coat front face	Spin resist, soft bake.	
30	Lap back face.	0.25μm diamond paste.	
31	Etch back face.	BHF etch to remove oxide.	
32	Evaporate aluminium for back contact	One 30-40 sec evaporation. -shadow mask two stripes.	Check resistance between stripes.

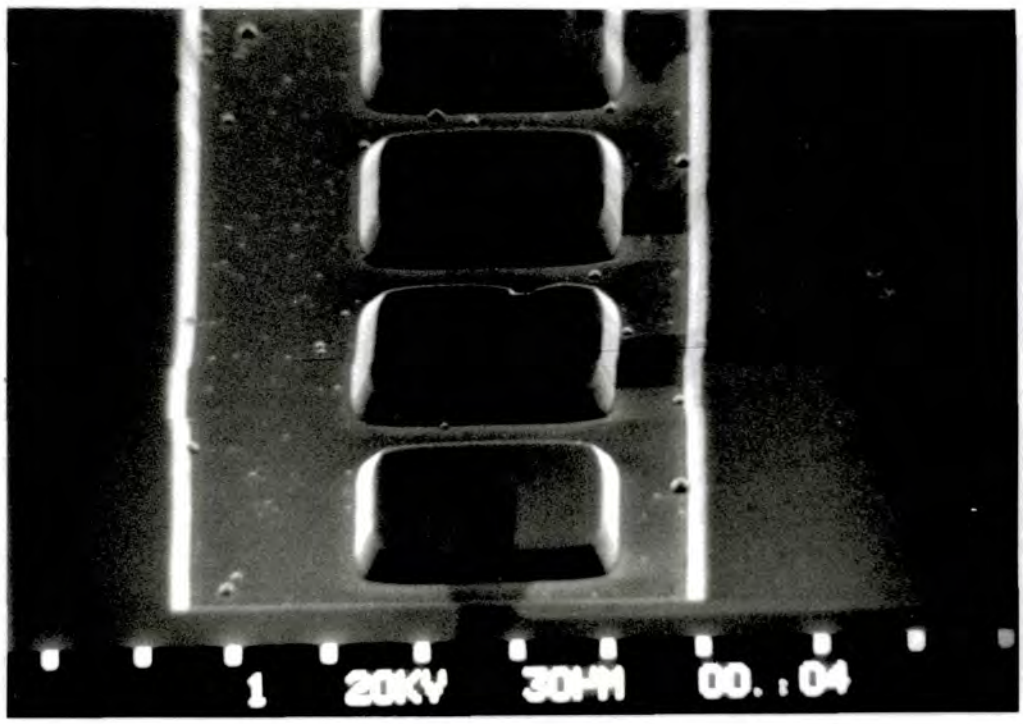
† BHF; 4:1 buffered hydrofluoric acid (NH_4OH : HF)

Table 5.1

Process for fabrication of V-groove isolated epitaxial MISS devices.



(a)



(b)

Figure 5.5; Scanning electron micrographs of completed V-groove isolated epitaxial MISS devices corresponding to the series;
(a) 9-13 (varying pn junction areas) and
(b) 14-16 (varying MIS areas).

5.3 Switching Characteristics of Isolated Epitaxial MISS Devices.

A number of wafers of isolated epitaxial MIS switches have been fabricated using the above process. Two of these, referred to as 'A' and 'B', have been electrically evaluated in some detail. They were processed identically except for the growth of the tunnel oxide layers which correspond to oxidation periods of 8 and 6 minutes respectively.

5.3.1 Measurement of MISS Switching Characteristics

Initially, switching characteristics of the isolated devices were obtained using an 'X-Y' pen plotter in the circuit shown in figure 5.6. Here, the load resistor R_L serves to limit the current in the highly conductive 'on' state and develops a voltage across it which is measured on the Y- (current) axis of the pen plotter.

At this point it is important to give some consideration to the superposition of the load line due to R_L on the MISS switching characteristic, bearing in mind that this includes an 'S'-type negative differential resistance (NDR) region.

Figure 5.7 demonstrates how a steep load line (low value of R_L) can intersect an S-shaped curve at more than one point while a sufficiently shallow load line (high value of R_L) may cut the curve at a unique point for any applied bias. With no prior knowledge of the nature of the I-V curve between the switching and holding points, it is only possible to state that the load resistance required to achieve the latter condition is given by;

$$R_L > |\mathrm{NDR}(I)| = \left.\frac{\Delta V(I)}{\Delta I}\right]_{I_S<I<I_H} \tag{5.2}$$

where NDR represents the value of the negative differential resistance at a current I. This condition may be assured, of course, if a constant current source (for which $R_L \rightarrow \infty$) is used in place of the voltage supply and load resistor.

It may further be stated that to measure both the switching and the holding point of the MISS with a single sweep of the applied bias, a requirement for R_L is;

$$R_L > \frac{(V_S - V_H)}{(I_H - I_S)} \tag{5.3}$$

If a lower value of R_L than this is used, then on increasing the applied bias above

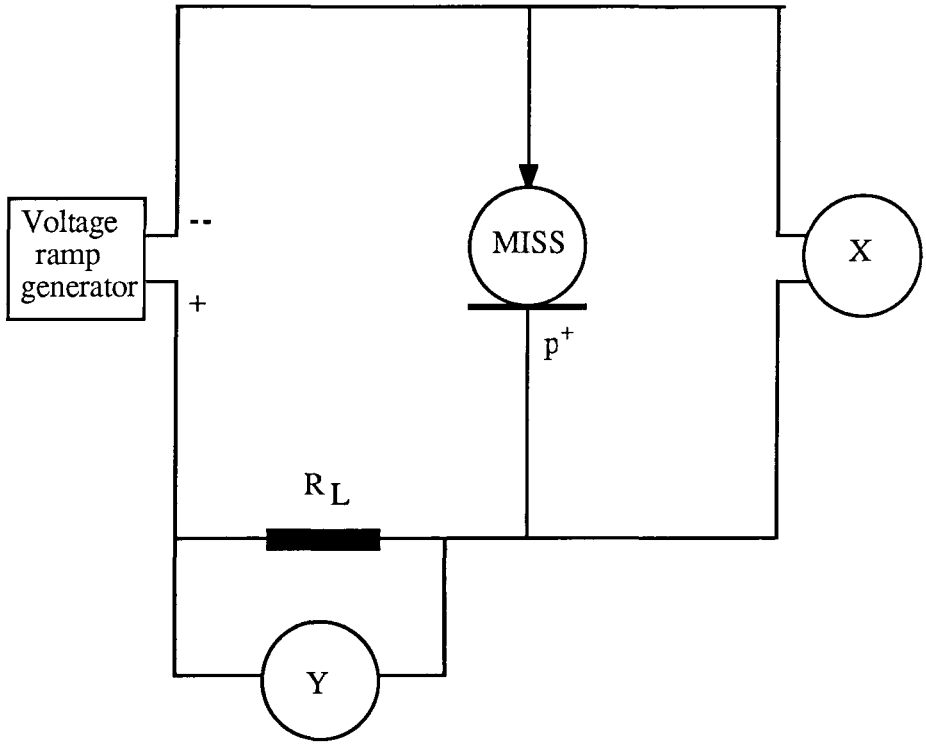


Fig 5.6; Measurement circuit used to plot MISS I-V characteristics using an X-Y pen plotter.

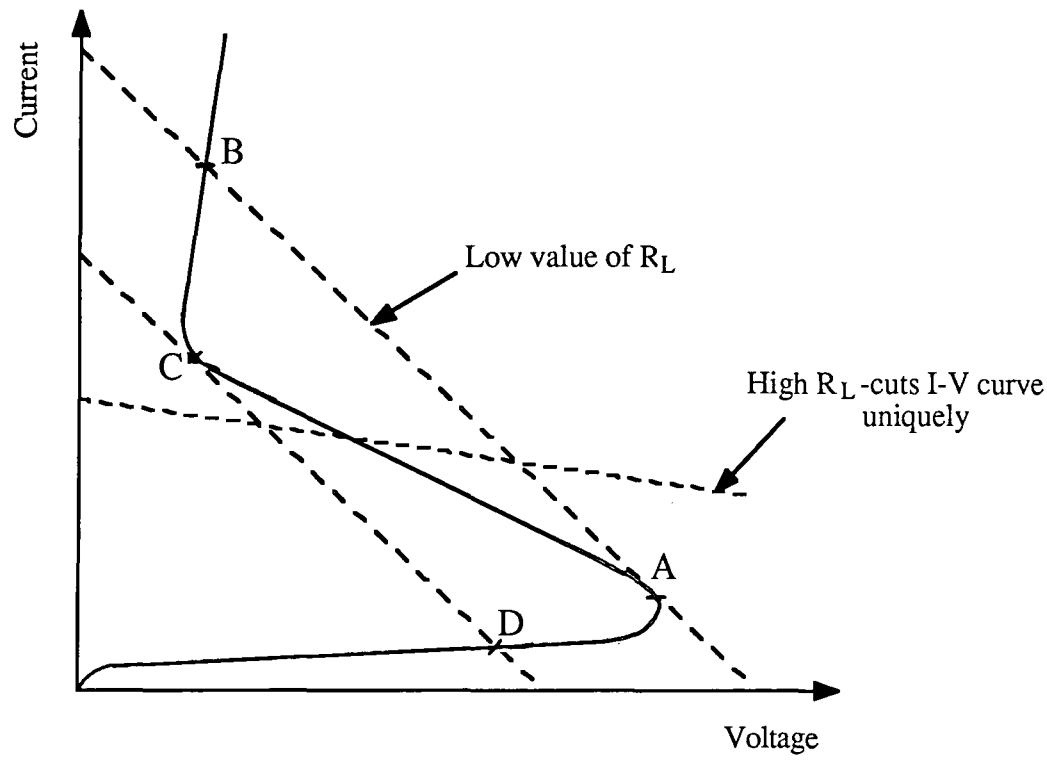


Figure 5.7; Superposition of load lines on an S-type NDR current-voltage characteristic.

the switching point, the quiescent point will move directly to an intercept with the 'on' state curve, somewhere above the holding point (transition A→B in figure 5.7). Thus a correct value for the switching point will be obtained but the true holding point will not be determined. If however, the applied bias is subsequently decreased, the holding point will be detected but the swiching point will not (transition C→D).

5.3.2 The Myth of the Third State

The effects of the position of the load line may be seen in the 'X-Y' pen plots of figure 5.8and 5.9 which were obtained with a low value of R_L (1kΩ) and an 'infinite' value of R_L (constant current source) respectively. Another consequence of using a pen plotter is evident in figure 5.8where the curve describing the switching transition is a result of the slow response of the pen plotter to changes in voltage at its inputs. This curve has no meaning in terms of the MISS characteristic. The curves drawn using the constant current source, figure 5.9, demonstrate another effect of the measurement circuit which might easily be misinterpreted as a property of the MISS device (and indeed often is in the literature). Whenever a constant current source or a high value of load resistance is used, a 'third state', intermediate between the stable 'on' and 'off' states, appears to exist , as indicated in figure 5.9. It can quite easily be shown, of course, that this apparent third stable state is in fact an artefact of the oscillatory behaviour which normally arises when a MISS device is biassed into its NDR region.

When the oscillations are detected using an oscilloscope they are found to be highly assymetric as indicated in the diagram inset in figure 5.9. This assymetry gives rise to the perceived I-V curve which results from measuring time-averaged values of current and voltage. The same phenomenon arises when a transistor curve tracer is used to measure a MISS switching characteristic and a high value of series resistance is selected as shown in figure 5.10. Here, the temptation to interpret the artefact of the unstable response as a third stable state is even greater.

Although this discussion of the proper determination of the switching characteristic may appear trivial, it must be pointed out that a good number of authors have reported the existence of three states in MISS characteristics.[10–12] Indeed, the third state has even been explained theoretically[13] . Unfortunately, an element of doubt must be cast on any such results.

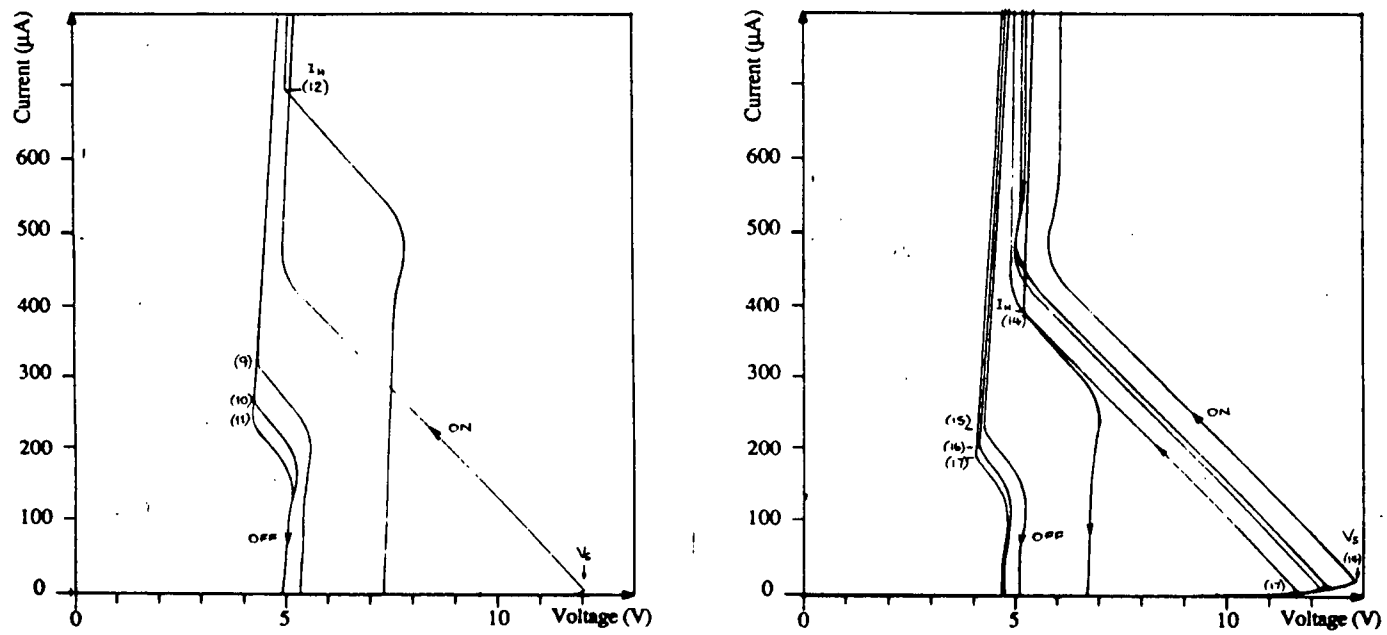


Figure 5.8; Typical MISS switching characteristics obtained using the X-Y plotter with a load resistance of 1kΩ.
(a) series 9-13
(b) series 14-17

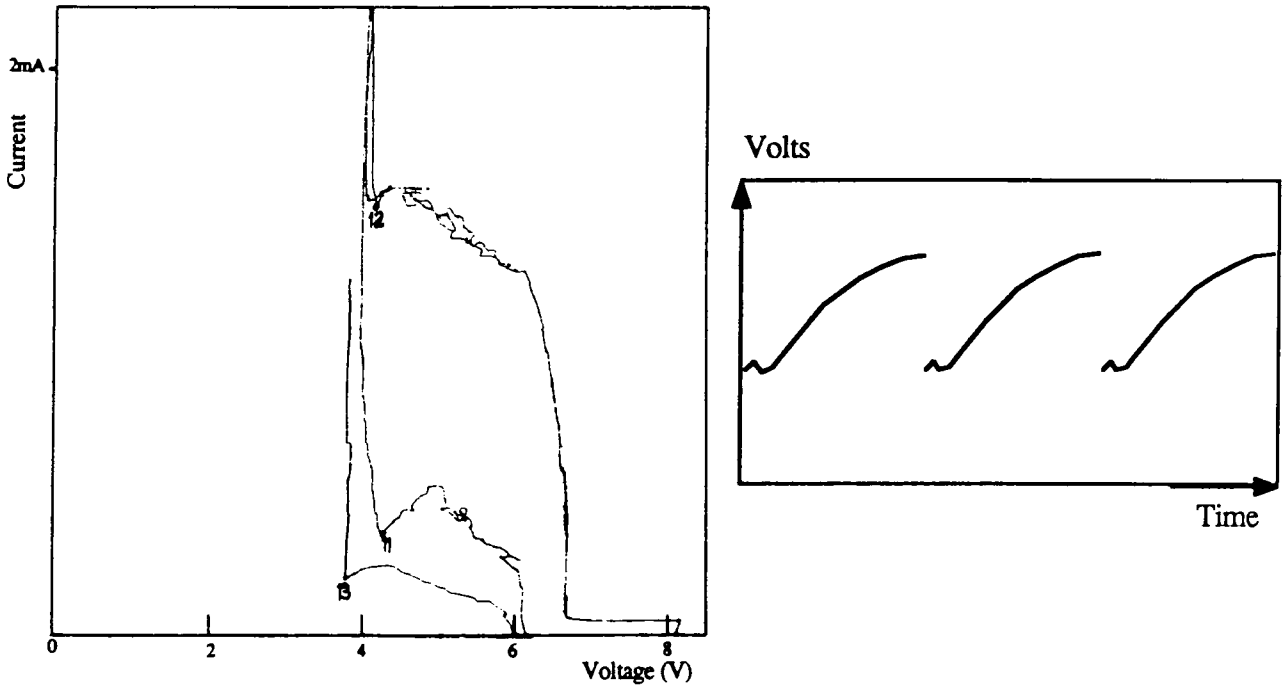


Figure 5.9; Typical switching characteristics obtained on a pen plotter using a constant current source (R=∞). Note the apparent stable state in the NDR region which actually corresponds to oscillatory behaviour (as indicated in the inset diagram). The devices are numbers 11, 12 and 13 off wafer A.

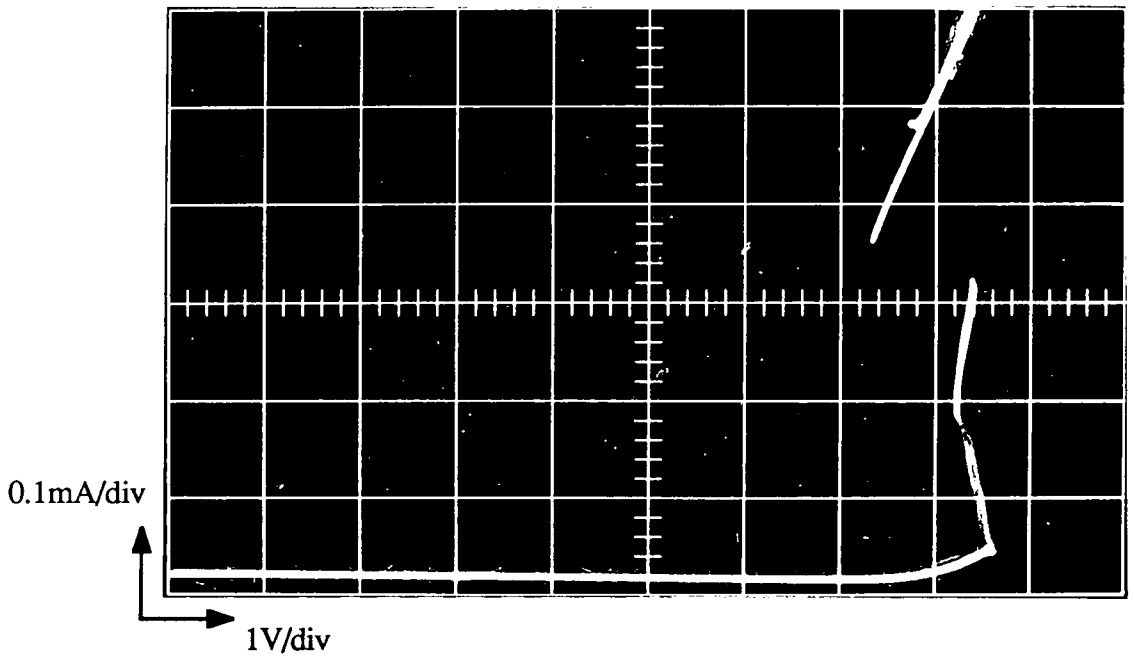


Figure 5.10; The I-V characteristic of a MISS device as it appears on a transistor curve tracer display when a large value of series resistance is selected. The 'intermediate state' is again a product of unstable behaviour and the relatively slow response of the measuring instrument.

Interestingly, the same phenomenon has recently been causing some consternation amongst workers in the field of resonant tunnelling heterostructure diodes[14] These diodes exhibit 'N-type' NDR and are currently attracting great interest as memory devices. They may also appear to exhibit a third state due to oscillatory behaviour in the NDR region. Such a naïve observation became the subject of intense debate in 'Physical Review Letters' recently,[15,16] until the misinterpretation was eventually pointed out.[17]

The question of instability in the NDR region will be treated in much greater detail in chapter 7. It is sufficient for the purposes of measuring the principal switching parameters V_S, I_S, V_H, and I_H to be aware of the possibility of falsely determining the stable states.

5.3.3 Automated Measurement of MISS Switching Characteristics

A large number of devices were initially measured using both the pen plotter and the transistor curve tracer. It became clear that reproducibilty of the switching parameters for any given size of device was not good enough to be able to accurately determine the effects of area variations. In addition, these measurement techniques were limited in range and were insufficiently sensitive for measuring the important off-state currents.

Therefore, subsequent I-V measurements were made using a picoammeter in an automated system similar to that described in chapter 3 for the study of MIS diodes. The only additions required to the circuit of figure 3.5 are a load resistor R_L and a D.C. voltage amplifier to increase the bias range of the D/A converter. The complete measurement circuit for MISS switches is shown in figure 5.11. An advantage in using a microcomputer for data capture is that it provides a means for automatically determining the principle switching parameters. The author has written simple algorithms to detect the turning points (V_S, I_S) and (V_H, I_H) and also the off-state current at a nominal off-state voltage of 3.0V during the measurement routine. Then, by capturing a sufficient amount of data, it becomes possible to determine variations in device characteristics on a statistical basis.

Figure 5.12 shows a sample of switching characteristics from devices numbers 10 and 17 on wafer 6, measured using the automated system. An important feature of these figures is the 'forming' effect which is manifested by these devices. The off state current

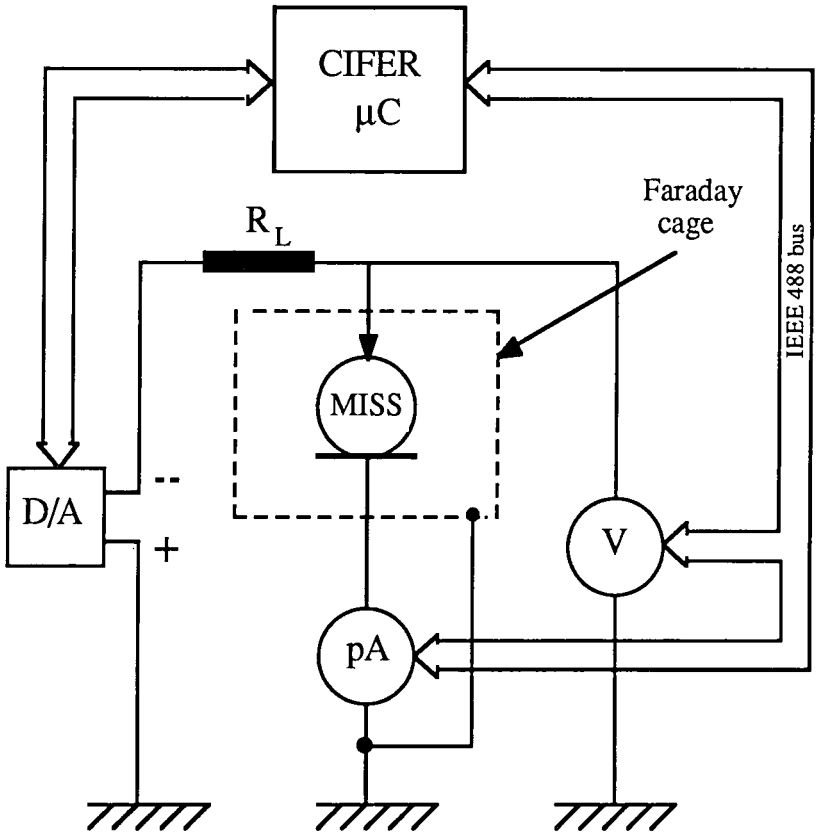
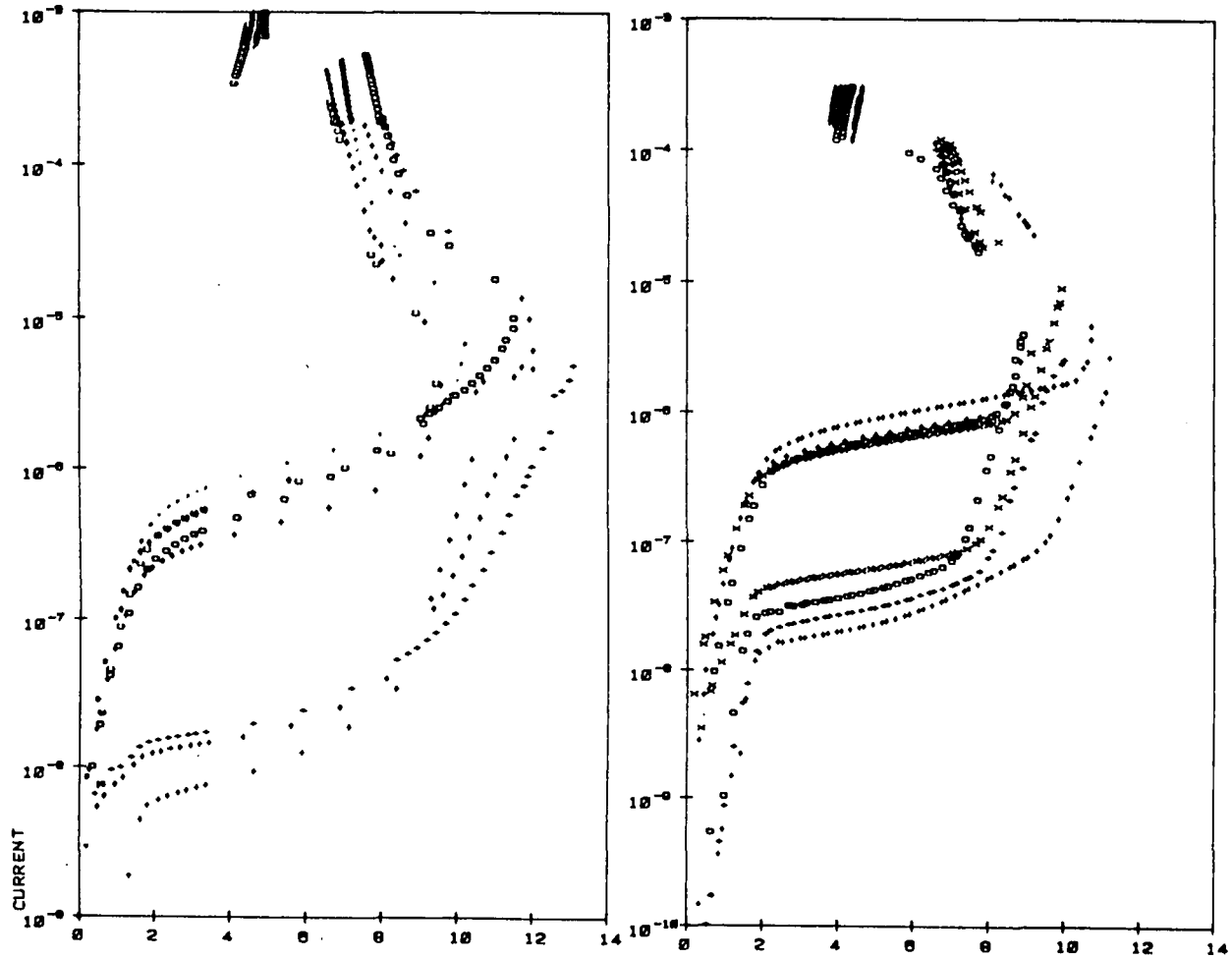


Figure 5.11; Computer-controlled measurement system for the study of MISS switching characteristics and extraction of the principle switching parameters.



(a) Switch 10

(b) Switch 17

Figure 5.12; Switching characteristics obtained point-by-point from several MISS devices of types 10 (a) and 17 (b). These plots demonstrate the spread in the measured characteristics and the effect of forming. (the lower curves correspond to unformed (virgin) devices).

Sw	V_S(virgin)	V_S(formed)	J_S(virgin)	J_S(formed)	J_0(virgin)	J_0(formed)
14	13.72 ± 0.82	13.29± 1.64	7.63± 8.65	61± 20	4.53 ± 3.38	135 ± 73
15	11.42 ± 1.35	10.43± 1.27	0.79± 0.67	1.52± 1.60	1.35 ± 0.89	43 ± 10
16	10.71 ± 1.00	10.17± 0.90	0.22± 0.11	0.42± 0.24	1.09 ± 0.36	50 ± 23
17	10.03 ± 0.69	9.79± 0.84	0.18± 0.06	0.22± 0.09	1.58 ± 0.67	30 ± 6

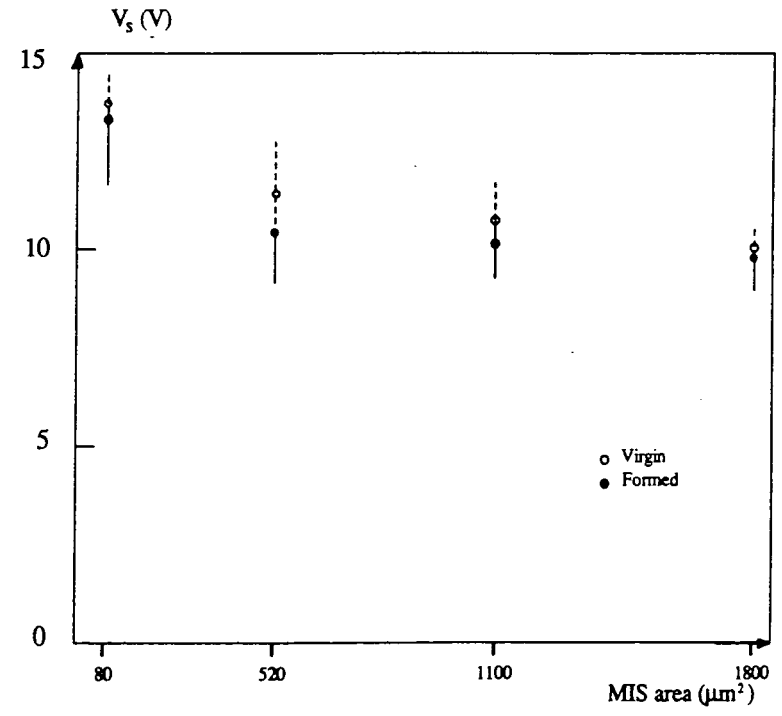
Table 5.2

Switching parameters (in volts, A/cm^2 and mA/cm^2 respectively) of virgin and formed MISS devices

of virgin devices (those that have not been previously biased beyond the switching point), is of the order of a few pA but this increases by at least an order of magnitude after the first switching cycle. In view of the possible importance of off-state leakage in real applications, this effect will first be examined in greater detail, before going on to discuss the effects of area.

5.4 Initial Forming of MISS Switching Characteristics.

A 'forming' effect in MISS devices was first reported by Kroger and Wegener[10] who found that the switching point of tunnel oxide MISS devices changed after the very first switching cycle but subsequently remained constant. To illustrate the effect, mean values of V_S, J_S and J_0 (with standard deviations) have been obtained from devices 14 to 17 both before and after forming and are given in Table 5.2. These results, plotted in figure 5.13, show that it is not only the switching point that is modified but the whole off-state curve moves upwards markedly. Indeed the off-state 'leakage' current increases by a factor of about 30! In contrast, no detectable change in the on-state curve is measured. An important point to note in these measurements is that the device must be switched into a highly conducting state for forming to occur. No change is detected if a MISS is merely driven along its off-state characteristic. However, no systematic study of the relation between the total current passed and the degree of forming has been undertaken. This is a possible avenue for further study if commercial application of tunnel oxides in this role is desired.



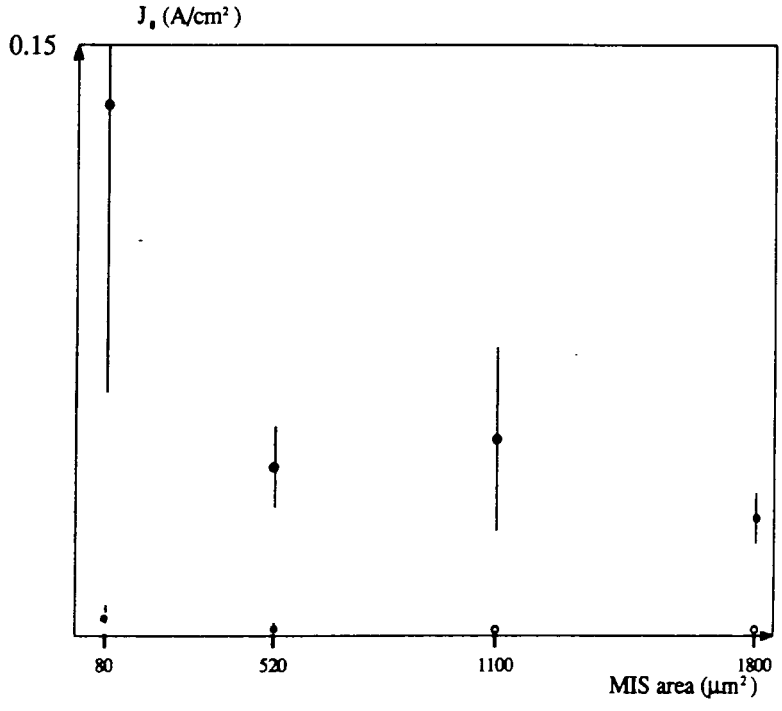


Figure 5.13; The switching voltage and off-state leakage current of formed (●) and unformed (virgin) (O) MISS devices. The data are mean values with standard deviation error bars, as given in table 5.3..

The observations of forming provide further evidence that the off current (the reverse saturation current of the MIS diode) is not semiconductor controlled, as established in chapter 3. It would be hard to account for the large increase on the basis of enhanced generation of carriers in the silicon.

The changes that occur due to forming have been further investigated by measuring the high frequency C-V curve (Boonton 1MHz) of a MISS before and after the first switching cycle. Figure 5.14(a) shows a typical set of 'virgin' (unformed) curves obtained from devices #14 to #17 while figure 5.14(b) shows the consequences of forming on the shape and height of the capacitance peak. Although the results from only two devices are shown for clarity (one of which is offset by the bond pad capacitance), these are typical of a reproducible trend.

5.4.1 **Possible Causes of Forming**

Following the discussion of reverse saturation currents in subsection 3.7.3 and of MIS C-V curves in 3.7.4, there would in fact appear to be two possible causes of the significant increase in both the peak capacitance and the leakage current:

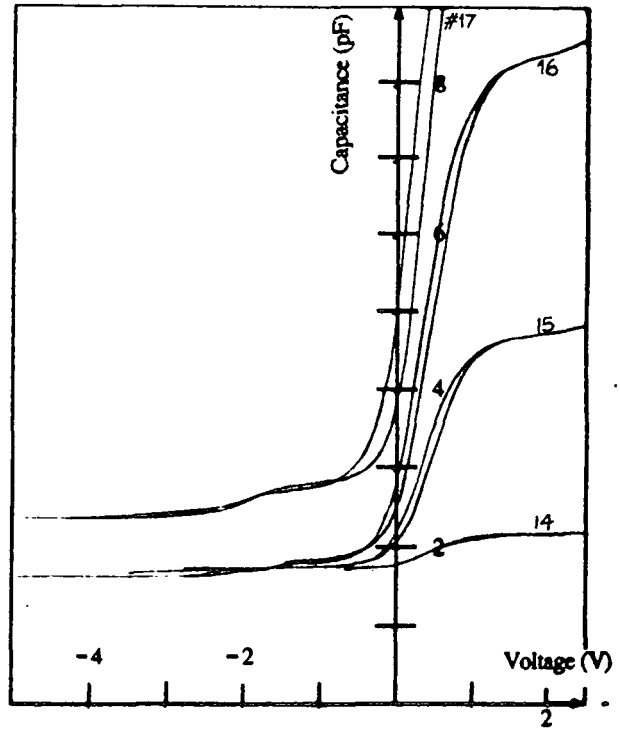
(a) A decrease in the effective thickness of the tunnel oxide.

This might arise from penetration of the contact metal (aluminium) deeper into the oxide layer, in the manner discussed in section 3.2.2. The heat energy released by the large flux of carriers passing through the oxide layer in the on-state may be sufficient to cause the reaction (3.1). The density of the power dissipated at the MIS junction may be estimated as J_H W/cm^2, assuming a 1 volt drop across the oxide, which corresponds to the very high value of $\sim$ 2000 W/cm^2 for the smallest device #14 ! Such an input of power to the junction might be expected to be rather destructive.

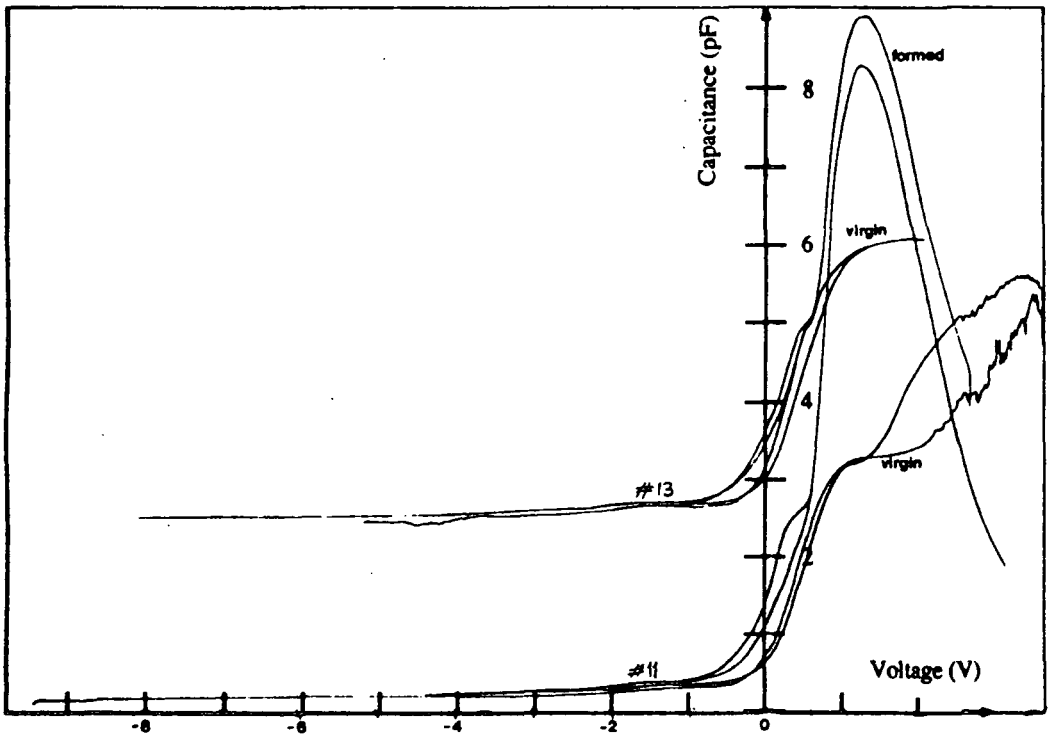
(b) An increase in the density of interface states.

It was suggested in chapter 3 that a large proportion of the total current passing through a MIS diode is assisted by interface states. An increase in their density, arising from the conditions to which the MIS diode is subjected in the on-state of the MISS, might therefore account for the increased leakage current.

However, at this point it is important to mention that the effect of forming was not permanent. Although after forming no further change would be apparent during the



(a)



(b)

Figure 5.14; 1MHz C-V plots from MISS devices on wafer 5;
(a) series 14-17 (all virgin)
(b) switches 11 and 13 (with bond pad connected) both before and after the first switching cycle.

same measurement session, a device re-measured several days or weeks later would be found to have returned to its 'virgin' state!

Again, no systematic study has been made to discover what time constants are involved in this 'mending' process or whether it may be assisted by thermal annealling. There is obviously some scope for further research into this effect. However, one conclusion that may be drawn from this apparent 'metastability' of the formed state is that it is unlikely to be related to aluminium penetration. The reaction (3.1) between Al and SiO_2 would not be expected to be reversible. Thus it seems more likely that generation of interface states is responsible after all.

5.4.2 **Generation of Interface States in MIS Diodes**

Generation of interface states in the metal-oxide-semiconductor system due to high field injection of carriers is not a new phenomenon. In fact, the effect has been extensively studied in relation to metal-nitride-oxide-silicon (MNOS) electrically alterable memory devices (EAROMs)[18] which also make use of tunnelling through a thin oxide layer. On the basis of this work, the conditions for their creation are now quite well established and can be seen to be identical to those existing in the MISS when it is turned on.

It has been shown that both electrons and holes may generate interface states upon injection in MOS structures[19] . Whereas electrons give rise to a typical broad U-shaped distribution (peaking close to E_V and E_C) which is attributed to strained and dangling bonds,[20] holes cause an additional peak 0.7eV above E_V[20] [21] Thus the presence of energetic holes at or near the interface is a necessary condition for creating the peak. In particular the 0.7eV peak has been attributed to the interaction of holes with Si-OH centres[20] . Water-related species such as SiOH groups in the oxide are also thought to accelerate the generation process[20] [22]

It has also been suggested that holes passing through the oxide may break strained Si-O bonds both at the interface (resulting in a trivalent Si and a non-bridging O) and in the oxide bulk. Indeed, in the case of dry oxides, this seems the more likely mechanism. Subsequently those broken bonds generated in the strained region of oxide near the interface diffuse in the strain gradient and pile up at the interface. Any Si-OH groups would probably assist the diffusion process by making the oxide lattice more labile.

Since the better part of a tunnelling thickness oxide layer may be expected to remain strained, as discussed in section 3.2.3, then it might be suspected that all broken bonds generated in the bulk SiO_2 will migrate to the interface. However, the enhancement of the MISS leakage current appears to be temporary, suggesting that the interface states created are metastable. It is not apparent why this should be the case, although it is known that traps created in SiO_2 layers by high energy radiation may be annealed out at low temperatures (around 100° C)[23] , so their stability is low.

5.5 Area Effects in the Isolated MISS.

In view of the considerable spread in the measured MISS characteristics, a number of devices of each size (typically ten) have been measured on wafer B and mean values of the five key parameters, together with standard deviations have been calculated. Wafer B was chosen for this study because the initial measurements using the plotter and transistor curve tracer indicated a better yield and greater reproducibility of devices on this wafer. The results are summarised in Table 5.3 where the device dimensions, as determined by use of an optical microscope with a calibrated eyepiece, are also given. The measured feature sizes do not correspond precisely with those on the masks, mainly due to the photolithographic problems discussed in section 5.2. Also included in table 5.3 are values of current density, J (A/cm^2) corresponding to I_S, I_H and I_O which were calculated using the measured MIS contact sizes.

5.5.1 Measured Effects on Device Characteristics

Several effects of area variations are particularly notable and will be examined;

The off-state

The mean values of off-state leakage current density, J_0 (@ 3V) (calculated by dividing the measured currents I_0 by the measured MIS areas) are not very much larger than their standard deviations, as might be expected in the light of the results obtained for the reverse biassed MIS diodes in chapter 3. There it was shown that just a ± 2Åvariation in d_{ox} may give rise to a six-fold variation in the leakage current density. However, neglecting switch 10 for which the mean of J_0 has an unusually high deviation, the measured values would appear to be reasonably constant, implying that the off-state is controlled by the MIS diode characteristics. Therefore, it might be concluded that

Switch No.	MIS area μm^2	pn area μm^2	V_S V	V_H V	I_S μA	I_H μA	I_0 nA	J_S A/cm^2	J_H A/cm^2	J_0 A/cm^2
9	510	52500	10.36 (1.07)	5.71 (0.46)	8.12 (3.64)	678 (434)	523 (324)	1.59 (0.71)	133 (85)	0.103 (0.06)
10	540	22800	10.28 (0.68)	4.58 (0.36)	5.76 (1.84)	523 (221)	1570 (3390)	1.066 (0.34)	97 (41)	0.291 (0.63)
11	520	12400	10.40 (0.97)	4.69 (0.99)	7.48 (3.91)	520 (234)	523 (290)	1.438 (0.75)	100 (45)	0.100 (0.055)
13	530	8305	10.27 (0.96)	4.65 (1.04)	8.97 (9.17)	705 (340)	420 (207)	1.69 (1.73)	133 (64)	0.079 (0.039)
14	80	12400	13.29 (1.64)	7.26 (1.36)	48.8 (16.5)	1630 (361)	108 (59)	61 (20.62)	2037 (451)	0.135 (0.073)
15	520	12400	10.43 (1.27)	5.25 (0.85)	7.89 (8.32)	825 (400)	223 (56)	1.52 (1.600)	158.7 (76.9)	0.043 (0.01)
16	1100	12400	10.17 (0.90)	4.45 (1.06)	4.60 (2.67)	488 (44)	476 (250)	0.418 (0.243)	44.4 (39.7)	0.050 (0.023)
17	1800	12400	9.79 (0.84)	4.11 (0.29)	3.88 (1.65)	183 (54)	531 (110)	0.216 (0.092)	10.17 (2.98)	0.030 (0.006)

Table 5.3

Mean values of switching parameters obtained from MISS devices of type B (with standard deviations in brackets). MIS and pn junction areas vary as indicated.

any effect of pn or MIS junction dimensions is insignificant. It must be noted however that these results do not take into account the effect of forming. All the readings taken have been from 'non-virgin' devices i.e. from devices that have been switched at least once prior to the measurement. It is possible though that the initial forming process was more effective for switch 10 or that a recurrent defect has arisen in the fabrication of this particular device. However, given that the other switching parameters of #10 are consistent with those of the other devices, this is not considered significant.

The Switching Point (V_S,I_S);

Whereas the switching voltage appears to be insensitive to variations in the pn junction area A_{pn} (switches 9-13), a definite trend is evident for changes in MIS junction area A_{MIS} (14-17). To be more precise, V_S clearly bears an inverse relationship to the size of the MIS contact. This effect is particularly pronounced for the smallest device, #14 which switches at a markedly higher voltage than the other devices.

The switching current, I_S varies with A_{MIS} in a similar manner to V_S, which leads to the apparent paradox that I_S actually increases with decreasing MIS area.

It is important to note however, that although small differences exist between devices with different MIS area, the switching voltage is fairly consistent for all the devices measured. Certainly, the large variation observed by Faraone et. al.[3] using a similar range of MIS and pn junction areas is not apparent. The implication of this consistency is that the MISS devices in this study are operating in the so-called 'punch-through' mode whereas those of Faraone et. al. were clearly not. This despite the specifications of the two processes being very similar. Following the discussion of section 4.7.4, it may be postulated that their devices were working in the more general feedback gain-controlled regime, suggesting the presence of a less conductive oxide layer. The near-punch-through mode of switching is characterised in both theoretical[24-26] and experimental[34] studies by;

(a) relative insensitivity to area effects compared to devices operating in the more general generation-controlled or high emission modes described in chapter 4 and

(b) a distinct rise in device current as the switching point is approached.

However, the fact remains that the area of the MIS diode does have some influence over the switching condition even in these punch through mode devices. In subsection

4.7.2, it was made clear that this mode is not as clearly defined as the term 'punch-through' implies. Theoretical studies by Lavelle[26] indicate that the actual switching criterion in these devices is not the meeting of the depletion regions but rather the rapidly improving efficiency of the BJT as the neutral base width contracts. In terms of the feedback analysis, this may be interpreted as an enhancement of the current gain which leads to switching before the full punch-through condition is reached.

The holding point, (V_H, I_H):

In common with the results for the switching voltage and current, the general trend is for V_H and I_H to increase with decreasing A_{MIS} but remain relatively insensitive to changes in A_{pn}. It is not clear, however, why the junction isolation should be so ineffective. Faraone et. al. found a reduction in I_H as the pn junction area was decreased (the pn:MIS area ratio decreased) and again recognised that it was due to an improvement in the efficiency of the regenerative feedback loop.

However, there has been as yet no theoretical explanation for holding voltages of the magnitude measured in this work. As such it is difficult to discuss the holding point in terms of the model presented in the previous chapter. Therefore the following discussion will concentrate on the effect of the MIS and pn junction areas on the switching point only.

5.5.2 Effects of Area on the Feedback Gain

Faraone et. al.[3] recognized that the MIS:pn junction area ratio has an effect on the switching characteristics through its direct influence on the gain of the regenerative feedback loop. They found experimentally that the switching current was insensitive to variations in A_{pn} and was controlled by the MIS area, A_{MIS} and oxide thickness[3] . In that work it was suggested that three conditions at the switching point are;

(i) The total device current is due almost excusively to hole conduction;

$$I_S = I_{PT}\rceil_{V=V_S} \tag{5.4}$$

(ii) In view of the high emitter efficiency of the p^+ -n junction, hole recombination is negligible;

$$I_{PT} \approx I_{PJ} \tag{5.5}$$

(iii) For any given d_{ox}, a critical inversion layer hole density, p_{crit}, is required to initiate the switching process. Using the same expression for I_{PT} as used in this thesis (equation 2.21) this was stated as;

$$I_{PT}]_{V=V_S} = A_{MIS}.J_{PT0}.\frac{p_{crit}}{N_V} \tag{5.6}$$

The reduction in V_S and I_S due to constriction of the pn junction was explained by the following argument. Isolation of the pn junction enhances its injection efficiency, γ, thus 'strengthening the regenerative feedback mechanism'. A higher loop gain in a pre-punch through mode MISS will give a lower value of V_S for two-terminal operation. In other words, the pn junction current density is increased if its area is reduced and the higher the current density, the higher the minority carrier injection efficiency γ as given by;

$$\gamma_{pn} = \frac{I_{PJ}}{I_{PJ} + I_{NJ}} \tag{5.7}$$

Before going on to offer an alternative explanation for the area effects in MISS devices, it is informative to consider other semiconductor devices in which current spreading is a factor, the bipolar junction transistor and the heterostructure laser. The performance of both of these devices is also affected by the relative areas of a minority carrier emitter junction and a top contact. To maximise gain, BJTs are always designed such that the collector-base junction has a greater area than the emitter base junction. When operated in reverse, the gain of a BJT is significantly reduced.

For instance, in I^2L devices the E-B junction of the 'vertical' transistors is inevitably larger than their C-B junctions, as shown in figure 5.15. The importance of the relative dimensions of the junctions was apparent early in their development[27-29] . Klaassen[29] noted that the common emitter gain, $\beta = I_C/I_B$ will be proportional to the ratio of the effective collector-base and base-emitter junction areas, A'_{CB} and A'_{BE} respectively;

$$\beta \propto \frac{A'_{CB}}{A'_{BE}} \tag{5.8}$$

He also suggested that, based on the results of earlier modelling of current spreading in epitaxial bipolar transistors[30] , the effective (electrical rather than physical) size of the base-emitter junction may be approximated by adding $3 \times W_{base}$ to the collector dimensions. This is equivalent to stating that the current spreading extends over a distance $1.5 \times W_{base}$ either side of the top (collector) contact.

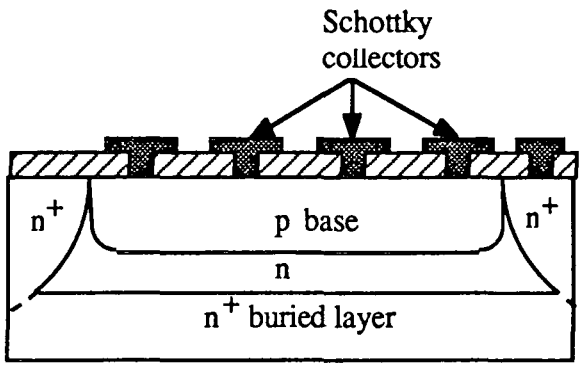


Figure 5.15; The principle features of an I^2L logic cell with Schottky collector contacts.

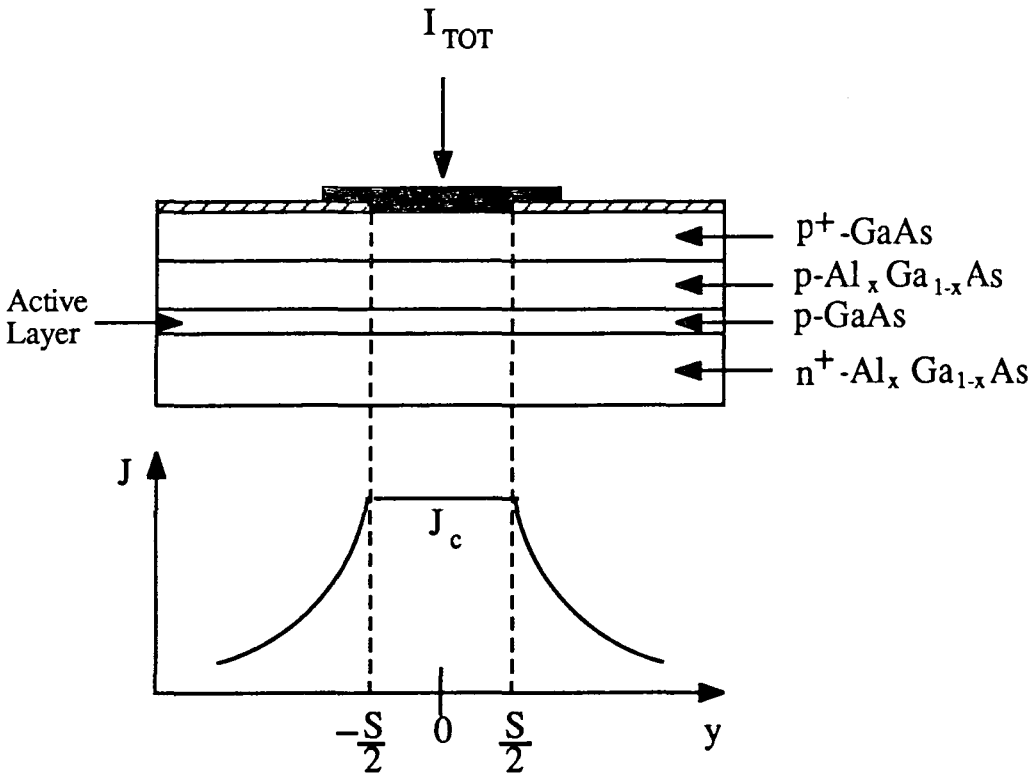


Figure 5.16; Schematic diagram of a stripe-contact hetero-structure laser, with a graph depicting the current spreading effect.

5.5.3 Current Spreading in the Stripe-Contact Laser

The lasing action in a GaAs/$Al_xGa_{1-x}As$ heterostructure laser[31] is dependent on the existence of a large concentration of minority carriers in a thin 'active' layer of GaAs. These minority carriers (usually electrons) are injected into the active layer (p-type) from a forward biased n^+p junction in which the n^+ emitter layer is formed of $Al_xGa_{1-x}As$.

In a simple 'stripe-contact' heterostructure laser, as depicted in figure 5.16(a), the lateral dimensions of the device are defined only by the width of the top (ohmic) contact, in very much the same way as in a non-isolated MISS device. The consequences of this arrangement are twofold;

(i) The majority carrier drift current in the p-layer spreads laterally as illustrated in figure 5.16(b).

(ii) Minority carriers which have been injected into the active layer diffuse laterally due to the existence of a concentration gradient.

The implication of the first factor is that the current flows through an area of the n^+-p junction that is larger than the stripe contact itself. Yonezu et. al.[32] first considered this problem and derived an approximate expression for the current distribution. Assuming a uniform current density, J_c, directly underneath the stripe contact, then for a contact stripe of width S centred at $y = 0$, figure 5.16, the current density some distance $|y| \geq S/2$ from the geometrical stripe edge is given[32] as;

$$J(y) = \frac{J_c}{\left(1 + \frac{|y|-S/2}{l_o}\right)^2} \tag{5.9}$$

where

$$J_c = \frac{I_0}{l_o.L} \tag{5.10}$$

Here, L is the stripe length, I_0 is the total spreading current on either side of the stripe region and l_o is the characteristic spreading length given by;

$$l_o = \frac{2.n.V_T.L}{\rho_s.I_0} \tag{5.11}$$

where, n is the non-ideality factor of the n^+-p junction and ρ_s is the sheet resistivity of the p-type layers above the active region.

Tsang[33] suggested that the spreading may alternatively be described approximately by an exponential decay type of relationship;

$$J(y) = J_c.\exp\left(-\frac{(|y| - S/2)}{l_o}\right) \tag{5.12}$$

The total majority current flow is then, neglecting recombination in the p-type;

$$I_{TOT} = I_c + 2.I_0 = S.L.J_c + 2L.\int_{S/2}^{\infty} J(y).dy \tag{5.13}$$

5.5.4 Current Spreading in Epitaxial MISS Devices

Following on from the discussion of area effects in the BJT and the stripe contact laser, it might be expected that two of these effects in particular will be important in the epitaxial MISS device;

(i) The loss of minority carrier current due to recombination in the neutral base layer surrounding the (MIS) collector depletion region will occur in the MISS as it does in the BJT with its emitter and collector terminals reversed. As the base region between the MIS and pn depletion regions is effectively neutral, there is no field acting on the minority carriers entering this region from the substrate. Hence their paths are determined only by diffusion/recombination. As such, if the area of the MIS collector depletion region is smaller than the area over which the minority carriers are being emitted, there will inevitably be a loss of efficiency. Some fraction of the holes injected by the forward biased n-p$^+$ junction will not be collected by the MIS depletion region but will recombine in the neutral bulk of the n-type epilayer. As a consequence, the collector efficiency of the tunnel oxide IBT (as defined in chapter 4) will be reduced. It then follows that the gain of the regenerative feedback loop formed of the IBT and BJT will be weakened by the same factor. In addition, the electrons that recombine with the holes in the outlying regions of the neutral epilayer have to be supplied by the MIS. As a result, the electron flux through the p-n junction is reduced with a consequent loss of emitter efficiency and hence a reduction in the gain G'_{BJT}

(ii) The majority current spreading effect that occurs in a stripe contact laser will similarly apply to the MISS. The electrons will diverge to some extent as they pass

through the neutral base layer between the MIS and pn junction depletion edges. However, the secondary effect of lateral diffusion of injected minority carriers may usually be neglected in the MISS since the residence time of minority carriers in the neutral base of the MISS will be short compared to that in the active layer of the laser (in which they are deliberately confined). In general, lateral diffusion may be neglected in the MISS as long as the aspect ratio $L'_{pn} : W_{neut}$ is large. (where $L'_{pn} = (A'_{pn})^{1/2}$ is the effective width of the pn junction over which the carriers are injected).

Since the MISS is a current-controlled device, the forward potential V_J across its pn junction at a distance y is determined by the current density $J(y)$ passing through the junction. Thus the distribution of the majority current $J_{NJ}(y)$ in the plane of the junction directly influences the distribution of the injected minority current $J_{PJ}(y)$.

5.5.5 Evaluation of the Current Spreading in MISS Devices.

At this point, it would be useful to estimate the extent of the current spreading for the MISS devices presently under study. A complete model of the spreading using a finite-element approach obviously requires considerable computation[26] and is beyond the scope of this thesis. Some measure of the extent of the current spreading is, however, provided by the notional characteristic length, l_o, as discussed in section 5.5.3. This parameter will now be estimated for the MISS devices in this study using equation (5.11).

Taking Tsangs expression for the spreading current, equation (5.12), the total majority current entering the top contact of an epitaxial MISS may be expressed, from (5.13), as ;

$$I_{TOT} = A_{MIS}.J_c + P_{MIS}.\int_{S/2}^{\infty} J_c.\exp\left(-\frac{(|y| - S/2)}{l_o}\right).dy \tag{5.14}$$

where A_{MIS} and P_{MIS} are the area and the perimeter of the MIS respectively. Solving (5.14) then;

$$I_{TOT} = J_c.[A_{MIS} + P_{MIS}.l_o] \tag{5.15}$$

The same result is achieved if Yonezu's expression, (5.9) is used instead. Now, from

(5.10)and (5.11), a substitution may be made for J_c in terms of l_o;

$$J_c = \frac{2.n.V_T}{\rho_s.l_o^2} \tag{5.16}$$

which results in a quadratic in $1/l_o$;

$$I_{TOT} = A_{MIS}.\frac{2nV_T}{\rho_s.l_o^2} + P_{MIS}.\frac{2nV_T}{\rho_s.l_o} \tag{5.17}$$

with the solution;

$$l_o = 2.\frac{A_{MIS}}{P_{MIS}}.\left[\sqrt{1 + 2.\frac{A_{MIS}.I_{TOT}}{P_{MIS}^2.n.V_T}.\rho_s} - 1\right]^{-1} \tag{5.18}$$

Thus the spreading length is proportional to the area to perimeter ratio! This leads to a rather surprising result for spreading in an unisolated MISS structure. Taking the MIS contact to be a circle of radius L and an 'effective' spreading length l_{eff}, the ratio of the effective pn and MIS junction areas is given by;[34]

$$\frac{A_{pn}'}{A_{MIS}} = \frac{\pi.(\frac{L}{2} + l_{eff})^2}{\frac{\pi}{4}.L^2} \tag{5.19}$$

giving;

$$\frac{A_{pn}'}{A_{MIS}} = 1 + 4.\frac{l_{eff}}{L} + 4.\frac{{l_{eff}}^2}{L^2} \tag{5.20}$$

A consequence of this expression is that if l_{eff} is considered to be equal to l_o then, substituting from (5.18), the ratio A_{pn}'/A_{MIS} is a constant, independent of the size of the contact! This however contradicts the experimental evidence of Duncan et. al.[34] and Majlis[35] that the size of the MIS contact does affect the switching characteristic of an unisolated MISS. Although this might imply the expression used for l_o (5.18) is invalid, it is more likely the case that l_o can not be simply treated as the same effective spreading length considered in equation (5.19).

ψ_S(V)	0	1	2	3	4	5	6	7	8	9	10	10.43
$[40\mu]^2$	2106	1458	1190	984	810	656	518	390	272	160	54	0
$[20\mu]^2$	1053	729	595	492	405	328	259	195	136	80	27	0
$[8\mu]^2$	421	292	238	197	162	131	104	78	54	32	11	0

Table 5.4

The spreading length l_o (μm) calculated as a function of the MIS surface potential for three sizes of contact (see text for parameters used).

5.5.6 The Spreading Length in 'Punch-Through' Mode Devices

The discussion of l_o will now be restricted to the near-punch-through mode devices which are of particular interest in the present study.

Before using equation (5.18) to estimate the spreading length, it is noted that in the case of a MISS device ρ_s is the sheet resistivity of the **neutral region**. As such, it is equal to the bulk resistivity of the epilayer, ρ_{epi} divided by the thickness of the neutral region, W_n;

$$\rho_s = \frac{\rho_{epi}}{W_n} \tag{5.21}$$

In this respect, the spreading phenomenon in a MISS device differs from that in a stripe contact laser for which the thickness of the spreading layer is fixed.

It follows that ρ_s is a function of the surface potential of the MIS. From (4.6) and (4.1);

$$\rho_s = \frac{\rho_{epi}}{W_{epi} - W_S} = \frac{\rho_{epi}}{W_{epi} - \left(\frac{2.\epsilon_s.\epsilon_o}{q.N_D}.\psi_S\right)^{1/2}} \tag{5.22}$$

In particular, when the doping of the MISS epilayer is low, as in the devices presented in this chapter, ρ_s and hence l_o will be highly bias dependent.

Estimations of l_o obtained using (5.18) and (5.22) are given in Table 5.4 for a range of values of surface potential, ψ_S. In these calculations, W_{epi} is taken as 3μm (allowing for up-diffusion of p^+ substrate dopants during high temperature processing[3]) and ρ_n

as 6Ω.cm. This choice of parameters would appear reasonable as it leads to a value of 10.43 for the punch-through voltage which is comparable to the measured values. The factor $n.V_T$ for a pn junction under forward bias at room temperature takes the value 50 mV at low current density (for which the junction current is dominated by recombination) and 25 mV at high current density (diffusion controlled). The contact dimensions L and S are equal for square geometries and for devices 14, 16 and 17 are 8μm, 20μm and 40μm respectively, as shown in the Table.

Clearly, as the switching point of a near punch-through mode MISS is approached, the spreading length reduces sharply. In the limit where the MIS depletion region reaches right through to the pn depletion region, the spreading actually goes to zero;

$$\lim_{V \to V_S} W_n = 0 \longrightarrow \lim_{V \to V_S} l_o = 0 \tag{5.23}$$

The reduction in l_o will arise not only due to the contraction of the neutral layer, as demonstrated in Table 5.4, but also due to the consequent increase in the total device current, which has not been taken into account in these calculations.

In essence then, near the switching point of a near punch-through MISS, the importance of current spreading is lost since l_o is likely to be small compared with the size of the pn junction. It is for this reason that the pn junction area has a negligible effect on the switching point of punch-through mode devices, as observed experimentally.

5.5.7 The MIS Aspect Ratio and Depletion Layer Spreading

It is necessary now to account for the effect of the MIS area on the switching point in the absence of current spreading. Near to punch-through, it would appear that the factor which most influences the ratio of the effective pn and MIS junction areas is the lateral extent of the depletion region. To a first order approximation, the depletion layer will extend sideways by the same amount as it extends downwards, W_S. Thus, the effective area of the pn junction becomes approximately;

$$A'_{pn} \sim (L_{MIS} + 2.W_S)^2 \tag{5.24}$$

The importance of the MISS aspect ratio L_{MIS}:W_{epi} becomes apparent. For example, if the diameter of a circular MIS contact is the same as the epilayer depth, then

near punch-through, the effective pn:MIS area ratio is about 9! Now, in a qualitative sense at least, the results of Table 5.3 may be explained.

The smallest device (# 14) suffers worst from what might be termed 'depletion layer spreading' and hence has the most reduced feedback gain and the highest switching voltage. The largest device (# 17) is the least affected because it has the largest aspect ratio.

To conclude then, whereas current spreading may be expected to affect the switching properties of a pre-punch-through MISS, the main effect on a punch-through device is likely to be depletion layer spreading. A consequence of this is that if the size of the MIS contact is decreased to achieve higher packing density and reduced power consumption, the depth of the epilayer must be scaled accordingly. The next generation of integrated circuits will work to 1 micron or even sub-micron design rules. In practice though, control of epilayer depth from growth through subsequent processing to the end product is very difficult, as discussed earlier. Certainly, epilayers thinner than a few microns would be severely affected by up-diffusion from the substrate and removal by oxidation. Thus the implication is clear; there is a basic lower limit to the practical size of an epitaxial MISS device. Alternative MISS structures which may overcome this geometrical limitation will be discussed in the concluding chapter.

5.6 Conclusions

Epitaxial MISS devices with V-groove junction isolation have been successfully fabricated using an 'I'-layer of tunnel oxide. It has been shown that under certain conditions of the measurement circuit, the switching characteristics of these devices may exhibit an apparent third state, intermediate between the ON and OFF curves. However, whereas other workers have attributed this feature to a second switching event[24] , it has been shown here that it may simply be dismissed as an artefact of the experiment. In particular, the intermediate curve is a consequence of making a D.C. measurement of a skewed time-varying (oscillating) signal.

Detailed measurements of the switching characteristics of a large number of MISS devices have revealed both 'forming' and geometrical effects. Forming is manifested as a large increase in the OFF-state current subsequent to the first switching event. This has been attributed to extra interface state generation under the conditions of high current and oxide field which prevail in the ON-state. The principle effect of geometry in these

'near punch-through' mode devices has been found to be an increase in the magnitude of the switching point (V_S, I_S) and holding point (V_H, I_H) parameters as the size of the MIS contact is reduced. Such trends are consistent with a reduction in feedback loop efficiency as the effective pn:MIS junction area ratio is increased.

It has been suggested by some authors[34] that the punch through mode is the ideal mode for practical 2-terminal MISS devices since $V_S(p-t)$ is (a) quite insensitive to temperature and oxide thickness and (b) determined by the epilayer doping and thickness. However, in the light of the results and discussion presented here, such reasoning is evidently not valid. Even in a near punch-through MISS, the switching characteristic is considerably influenced by geometric effects which become more severe as the device is scaled to smaller dimensions.

The assertion that the punch-through regime in epitaxial MISS devices is inherently better controlled than other switching modes is also disputable. Control of doping during growth of a low-doped n (or p) epilayer on a highly doped substrate p^+ (or n^+) is in practice very difficult. At the high temperatures required for epitaxy, 'autodoping', the incorporation into the growing epilayer of evaporated acceptor (or donor) species from the substrate, may be considerable. In addition, the abruptness of the p^+-n junction cannot possibly be maintained and the epitaxial junction will be graded. Furthermore, fabrication of microelectronic devices necessarily involves high temperature processes such as oxidation and activation of implants which will further up-diffuse and degrade the epitaxial junction.

However, the fact remains that the other modes of operation of the device are highly sensitive to variations in the tunnel oxide thickness and interface properties. Thus, if a non-punch through mode is desired, the uniformity and reproducibilty of the semi-insulator layer must be improved.

REFERENCES

1. J.G.Simmons and A.El-Badry, *The Radio and Electronic Engineer***48** 215 (1978)
2. G.Sarrabayrouse, A.Essaid and J.Buxo, **17** 681 (1982)
3. J.G.Simmons, L.Faraone, U.K.Mishra and F-L.Hsueh, IEEE Electron Dev. Lett. **EDL-2** 109 (1981)
4. C.Y.Chang, F.C.Tzeng, C.T.Chen, S.J.Wang and Y.D.Wang, IEEE Electron Dev. Lett. **EDL-6** 545 (1985)
5. K.E.Bean, IEEE Trans. El. Dev.,**ED-25** 1185 (1978)
6. M.J.Declerq, L.Gerzberg and J.D.Meindl, J. Electrochem. Soc. **122** 545 (1975)
7. J.B.Price, *Semiconductor Silicon 1973* (editors H.R.Huff and R.R.Burgess, Electrochemical Society)
8. R.M.Finne and D.L.Kleine, J. Electrochem. Soc. **114** 965 (1967)
9. D.L.Kendall, Appl. Phys. Letters **26** 195 (1975)
10. H.Kroger and H.A.Wegener, Solid State Electronics **21** 643 (1978)
11. A.Adan and K.Dobos, Solid State Electronics **23** 17 (1980)
12. D.C.Y.Chang, Chung L.L. and Tan F.L, Solid State Electronics **32** 179 (1989)
13. S.E-D.Habib and J.G.Simmons, Solid State Electronics **23** (1980)
14. T.C.L.G.Sollner, W.D.Goodhue, P.E.Tannenwald, C.D.Parker and D.D.Peck, Appl. Phys. Letters **43** 588 (1983)
15. V.J.Goldman, D.C.Tsui and J.E.Cunningham, Phys. Rev. Lett. **58** 1256 (1987)
16. T.C.L.G.Sollner, Phys. Rev. Lett. **59** 1622 (1987)
17. J.F.Young, B.M.Wood, H.C.Liu, M.Buchanan, D.Landheer, A.J.SpringThorpe and P.Mandeville, Appl. Phys. Letters **52** 1398 (1988)
18. K.Jeppson and C.M.Svensson, J. Appl. Phys. **48** 2004 (1977)
19. S.K.Lai, Appl. Phys. Letters **39** 58 (1981)
20. C.T.Sah, J.Y.C.Sun and J.J.T.Tzou, J. Appl. Phys. **53** 8886 (1982)
21. L.DoThanh and P.Balk, in *Insulating Films on Semiconductors* editors J.F.Verweij and D.R.Wolters (North-Holland, 1983)

22. C.M.Svensson, in *The Physics of SiO_2 and its Interfaces* editor S.T.Pantelides (Pergamon, 1978)

23. G.Sarrabayrouse, private communication.

24. S.E.D.Habib and J.G.Simmons, Solid State Electronics **22** 181 (1979)

25. A.F.C.Fiore de Mattos and G.Sarrabayrouse, (1986)

26. S.Lavelle, *PhD thesis*, University of Durham (1989)

27. F.W.Hewlett, IEEE J. Solid-State Circuits **SC-12** 206 (1977)

28. H.E.J.Wulms, IEEE J. Solid-State Circuits **SC-12** 143 (1977)

29. F.M.Klaassen, IEEE Trans. Electron Devices **ED-22** 145 (1975)

30. J.W.Slotboom, IEEE Trans. Electron Devices **ED-20** 669 (1973)

31. H.C.Casey and M.B..Pannish, *Heterostructure Lasers (Part B)* (Academic Press)

32. H.Yonezu, I.Sakuma, K.Kobayashi, T.Kamejima, M.Ueno and Y.Nannichi, Jap. J. App. Phys. **12** 1585 (1973)

33. W.T.Tsang, J. Appl. Phys. **49** 1031 (1978)

34. K.A.Duncan, P.D.Tonner, J.G.Simmons and L.Faraone, Solid State Electronics **24** 941 (1981)

35. B.Y.Majlis, *PhD Thesis*, University of Durham, UK (1988)

CHAPTER SIX

Deposition and Characterisation of Silicon Rich Oxide

6.1 Introduction

Semi-insulating polysilicon (SIPOS), otherwise known as oxygen-rich polysilicon (ORPS) or silicon-rich oxide (SRO) has received considerable attention in recent years due to its potential as a passivation layer for integrated circuits[1] [2] and for high voltage devices[3] . The stoichiometry of a silicon-rich oxide layer may be represented by the value of x in SiO_x . When the excess silicon content is low ($x \rightarrow 2$), the small but significant conductivity of the layer has the effect of smoothing-out high electric field concentrations at the silicon surface and of providing a barrier to external fields.

More recently, the same material has shown great promise as a charge injection layer for floating-gate electrically-alterable memory devices (EAROMs)[4] [5] where it replaces tunnelling-thickness oxide layers.

In a highly doped form, SIPOS has also been used as a wide gap semiconductor for improved injection efficiency in heterojunction bipolar transistors[6] and has been proposed as a possible insulating emitter layer for a silicon BICFET[7] which was discussed in chapter 4.

6.2 Application of SRO in MISS Devices

It is expected that SRO will be a highly suitable replacement for the tunnel oxide layer in MIS switching devices. This material would appear to offer several advantages;

(i) In tunnel oxides, a change in thickness of only a few Å may cause a large change in the tunnel currents, as discussed in chapter 3. This sensitivity leads to variations in the I-V characteristics of different devices and may also be expected to give rise to non-uniform conduction through the I-layer of any particular device. As SRO films should provide comparable conductance to tunnel oxide for thicknesses an order of magnitude greater, it is hoped that uniformity of conduction will be improved.

(ii) One of the limitations on the switching speed of MISS devices is the delay required for charging of the insulator capacitance, C_{ins}. Although in terms of capacitance, the greater thickness of an SRO layer is partly offset by its larger dielectric constant (in the range 6 to 8), a five-fold reduction in C_{ins} may be expected.

(iii) Whereas the electrical properties of a tunnel oxide layer may only be controlled through a single parameter, its thickness, the conduction through SRO layers may also be adjusted through the composition (x) and structure (by annealing). Thus it should be possible to tailor the SRO layer to obtain the required switching properties for a given application. This may be of particular importance when the device is to be implemented alongside MOS transistors in integrated circuits. In this case, the MISS will have to be fabricated within the constraints of the existing MOS technology and the switching characteristic will only be adjustable by modification of the insulating layer parameters.

6.3 The Structure of SiO_x

Although the two terms SRO and SIPOS seem to be used interchangeably in the literature, it might be suggested that the former ought to refer to layers which exhibit properties typical of an amorphous material and the latter to those which are polycrystalline in nature.

The distinction between amorphous and polycrystalline states is somewhat vague and depends whether the structural or electrical properties of the material are being considered.

In general, the structure of a material may be considered amorphous if no order is apparent at a range of about 50Å and polycrystalline if microcrystallites of this size or greater are present. Reflected high energy electron diffraction (RHEED) provides a sensitive means for determining the degree of order in the structure.

In terms of the electrical properties, the SiO_x may be considered amorphous if its conductivity is described by either (a) a model based on extended band tails in the energy gap (as for amorphous silicon)[8] or (b) by the Poole-Frenkel model of field-enhanced emission from Coulomb-well type trapping centres[9] . Conduction in a polycrystalline material is mainly attributable to the presence of grain boundaries between the microcrystals where a potential barrier exists[10] . In the case of SiO_x , a model which considers the small silicon crystals to be surrounded by thin sheets of non-stoichiometic

oxide with a large density of interface traps was proposed by Tarng[11] . In this model, the potential barriers at the grain boundaries were treated as symmetrical Schottky barriers.

It is now quite well established that the structure of an SiO_x layer is principally determined by post-deposition annealing[12] . Although silicon 'clusters' about 10Å in diameter are observed by TEM in as-deposited films, electron diffraction studies indicate that the whole layer is structurally amorphous[12] . After annealing at 1100°C, however, diffuse electron diffraction rings indicative of a polycrystalline component in the layer are seen[12] and the grain size is found to increase with increasing silicon content (decreasing x) in the film. The smallest crystallites observed are about 25Å which is close to the theoretical minimum for silicon[13] . The critical temperature for the growth of silicon crystallites also appears to depend on the Si concentration. Although for $x < 1$, annealing at 850°C seems to suffice, a temperature greater than about 1000°C appears to be necessary to obtain crystalline Si clusters for $2 > x > 1$[14] . Where these conditions are not met, clusters may still be enlarged by annealing but will remain amorphous[12] .

6.4 Deposition of SRO

In the present study, a horizontal silica-walled atmospheric pressure chemical vapour deposition (APCVD) reactor has been used to deposit SRO films at a temperature of 650°C . In this system, the substrates, usually quarters of 2 inch silicon wafers, lie on a graphite susceptor which is heated to the required temperature by an array of heating lamps. The temperature is monitored by an infra-red pyrometer aimed at the susceptor through the top wall of the reactor. The reactants used are silane (SiH_4) and nitrous oxide (N_2O) with nitrogen as the carrier gas and the film composition is controlled through the ratio γ of the gas flows of these two components;

$$\gamma = \frac{N_2O\text{gasflow}}{SiH_4\text{ gasflow}} \tag{6.1}$$

The silane enters the reactor tube at one end where it is mixed with the carrier flow but the N_2O is admitted into the gas stream much closer to the susceptor in order to avoid excessive prereaction in the gas phase. This arrangement initially gave rise to poor uniformity of deposition because the N_2O was being injected through a series of

holes in the side of a stainless steel tube. This had the effect of streaming the N_2O flow and resulted in a streaked appearance of the deposited layer due to lateral variations in thickness and composition. However, a marked improvement in uniformity was achieved by using a fine stainless steel gauze in place of the holes in the injector nozzle which provided a finer dispersal of the injected N_2O into the gas stream.

Preparation of silicon substrates prior to deposition was by the method described for the growth of tunnel oxides in chapter 3. In this process the final treatment of the silicon surface was a short free etch in 10% HF in water to remove any nascent (atmospheric) oxide and this was done immediately before loading a sample into the CVD reactor. After loading, the reactor was flushed for a period of about 20 minutes before starting to heat the susceptor. The temperature was allowed to stabilise at the set point for a short time before starting the deposition by admitting the reactant gases. The growth period was controlled by an automatic timer and was typically in the range 20 to 120 seconds. However, as was pointed out in chapter 3 for the growth of thin oxides, the true deposition time is not defined precisely by this method. In the case of SRO deposition, the delays for passing the gases from their respective cylinders to the reactor was minimised by operating a 'vent-run' system; the gases were passed to the exhaust vent for a period before being switched into the reactor. However, it must be noted that because the N_2O was injected downstream, it must have arrived at the point of deposition in advance of the SiH_4.

6.5 Determination of SRO Thickness and Composition

The thickness and refractive index of freshly deposited SRO layers have been routinely measured by ellipsometry. For SRO layers with a mole fraction of Si in the range 0.42 to 0.45, less than a 10% error arises if absorption is neglected[15] (i.e. if the imaginary part of the refractive index is treated as zero). The viability of ellipsometry for measuring SRO layers has been confirmed by comparing measurements made using this technique with those obtained using two other methods;

(a) step height measurement ('Alpha-step') of features defined by plasma etching using a (CF_4+ 8% O_2) plasma and

(b) reflectance interferometry ('Nanospec'), assuming a refractive index of 3.05 from the ellipsometry readings.

Each instrument gave a reading of 250Å ± 10Å for the same SRO layer which is of the same order as the uniformity across a sample as determined by any one technique.

Figure 6.1 shows the variation in the measured SRO film thickness with deposition time for several values of γ. The main trends are for the deposition rate to increase due to either;

(i) an increase in N_2O concentration in the vapour (higher γ) or

(ii) an increase in both reactant flows (with constant γ)

However, the enhanced rate arising from (i) is always accompanied by a fall in the silicon content (an increased n_r) of the layer. Even for a constant γ, though a fall in n_r with decreasing layer thickness is consistently observed. This may be accounted for by the presence of a thin oxide layer both at the SRO-silicon interface and on the top SRO surface. The interfacial oxide would arise, in part, from oxidation in the laboratory ambient prior to loading the samples. It might further be suggested that this nascent oxide grows thicker because the N_2O reaches the sample in advance of the SiH_4 at the start of the deposition process (at which stage the substrate is at a temperature of 650° C). However, no experiments have been conducted to clarify this point. The presence of a surface oxide is also quite well established. It has been observed that the refractive index of SRO may decrease over a period of time, indicating that the material is quite unstable and may oxidise slowly in air.

Having established the correctness of the ellipsometry data, the measured refractive index may be further utilised to obtain an estimate of the film composition. If it is assumed that the relative permittivity ϵ_{SRO} is equal to the refractive index squared, then the Maxwell-Garnet relation gives;

$$\frac{\epsilon_{SRO} - \epsilon_{ox}}{\epsilon_{SRO} + 2.\epsilon_{ox}} = f.\frac{\epsilon_{Si} - \epsilon_{ox}}{\epsilon_{Si} + 2.\epsilon_{ox}} \tag{6.2}$$

where f is the mole fraction of silicon in the layer. This expression reduces to;

$$f = 2.4625.\left(\frac{\epsilon_{SRO} - 3.9}{\epsilon_{SRO} + 7.8}\right) \tag{6.3}$$

for the silicon/silicon dioxide system. Finally, the 'composition factor', x is given by;

$$x = 2 - 2.f \tag{6.4}$$

UNIV. DUNELM.

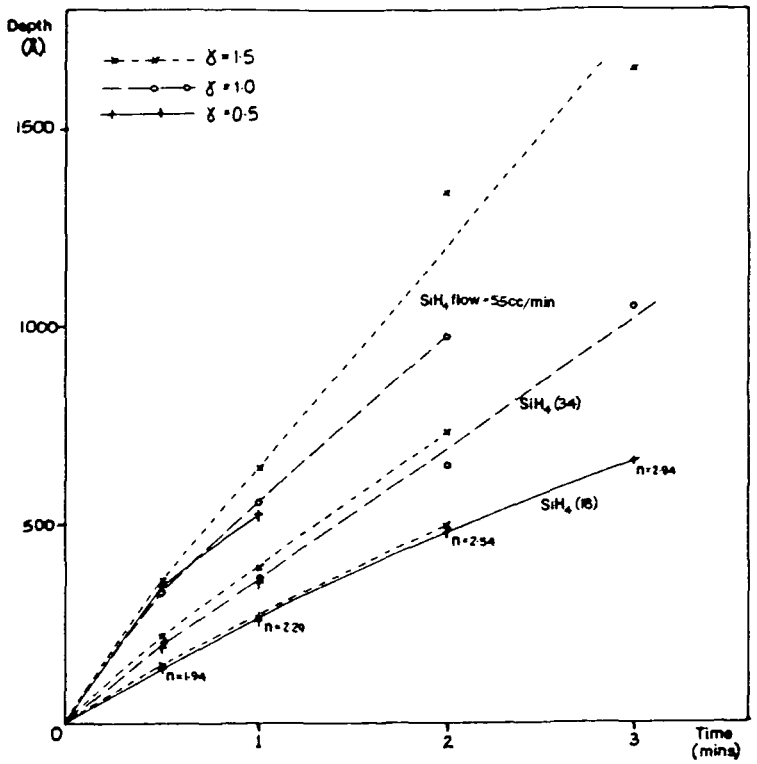


Figure 6.1 ; Thickness of SRO deposit against time for a variety of silane flows and γ.

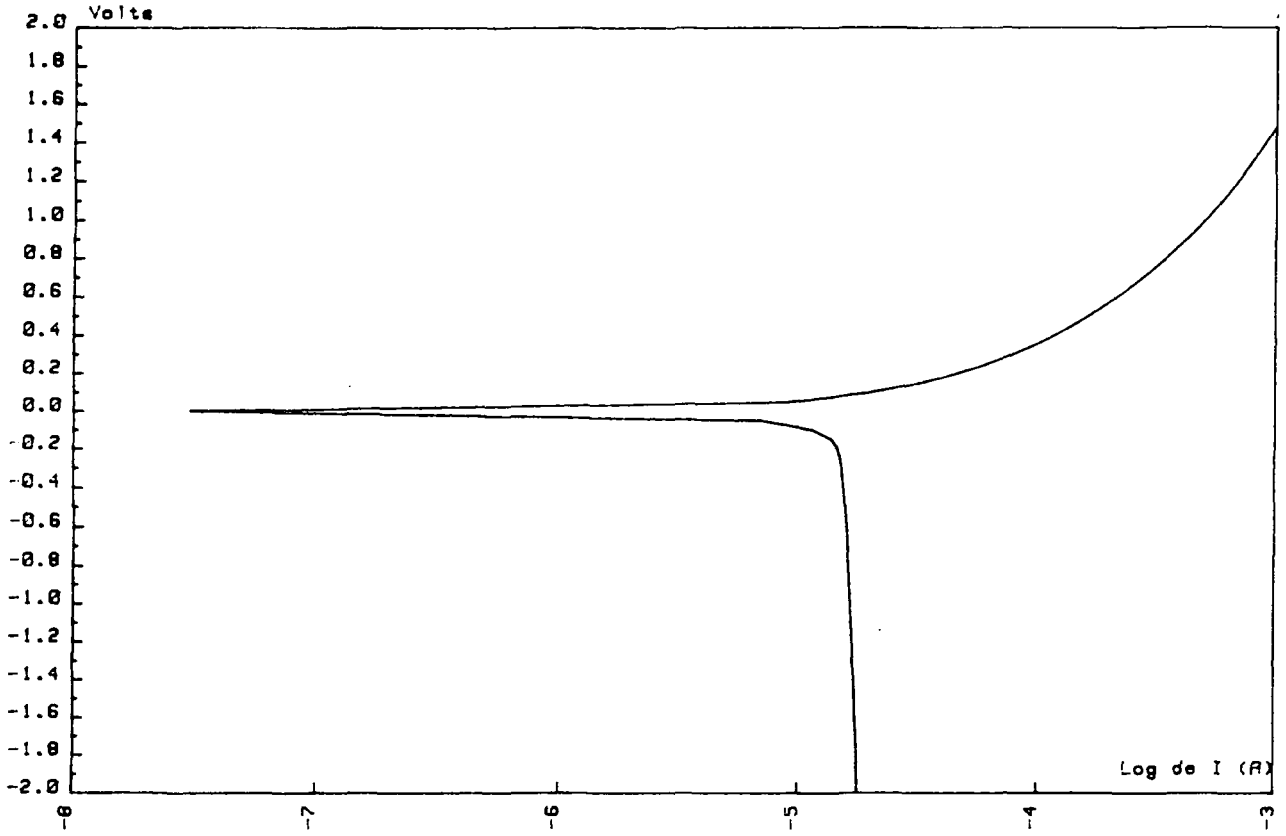


Figure 6.2 ; The I-V characteristic in forward (+ve) and reverse (-ve) bias of a 0.88 mm ^{2} SRO MIS diode

6.6 Annealing of SRO Films

A large number of deposited films have been annealed in order to investigate how this effects the switching properties of MISS devices and to confirm the structural effects which are reported to occur[12] . As-deposited samples were routinely scribed in half and one of the halves annealed in dry nitrogen in the wet oxidation furnace for 1 hour. Direct comparison was therefore possible between annealed and unannealed material.

Although the structure of unannealled SRO has been confirmed as amorphous.[16] Electron diffraction (RHEED) studies have confirmed that a degree of crystallinity does result from the anneal.[16] However, ellipsometry readings have consistently indicated that the main effect of annealling is to reduce n_r. This is probably accounted for by accelerated oxidation of the SRO surface which may arise due to either (a) the presence of residual oxidising species in the wet oxidation tube used or (b) adsorbed moisture on the SRO surface which subsequently reacted in the furnace.

Whatever the cause, oxidation of the SRO is undesirable and most of the layers used in the present work have been left unannealed for these reasons.

It is therefore asssumed at this point that the polycrystalline models will not be applicable and the conduction in these layers will best be described by an amorphous model.

6.7 Electrical Characterisation of SRO MIS Diodes

Both the current-voltage and capacitance-voltage characteristics obtained from MIS diodes with SRO I-layers are remarkably similar to those obtained from tunnel oxide devices.

6.7.1 Current-Voltage Characteristics

The I-V curve shown in figure 6.2 is typical of the conduction properties exhibited at room temperature. In reverse, there is a distinct saturation of the current which is again attributable to semiconductor limited minority carrier supply. The saturation current may be raised by illuminating the sample in the same manner as for the tunnel oxide MIS. The form of the I-V curve under forward bias and in reverse bias prior to saturation is not exponential, indicating that the conduction mechanism is different to that in

tunnel oxide. However, the two curves are symmetrical in the low current range, as shown in figure 6.3 in which the forward and reverse characteristics are superimposed for comparison. Such symmetry is indicative of a bulk conduction mechanism in the SRO, rather than a contact-limited one. The latter would generally give rise to assymetry due to the different work functions of the contacting materials (aluminium and silicon).

6.7.2 Capacitance-Voltage Characteristics

The C-V curves of figure 6.4 were obtained from a MIS diode with a 250Å layer of SRO with $x = 0.68$ (estimated from the measured $n_r = 2.86$). Strong frequency dispersion is evident over five decades of frequency. Again, this can not be accounted for by the series resistance correction and, following the discussion of dispersion in tunnel oxide capacitance, it must be concluded that interface states are also of great importance in the SRO MIS. This is perhaps not surprising given that SRO is a deposited rather than a grown material and must therefore be expected to leave a high density of dangling Si- bonds at the interface.

6.7.3 SRO MISS Devices

Not surprisingly, epitaxial MISS devices with SRO layers exhibit switching characteristics which are very similar to those of tunnel oxide devices. The area effects discussed in the previous chapter are easily reproduced. One important way in which SRO devices differ, however, is that they do not suffer from the same forming effect. It is probable that the forming process does not occur because very high insulator fields do not arise in the SRO. The SRO layers are about ten times the thickness of the tunnel oxides, yet the potential difference they have to support is little different (as evidenced by comparable holding voltages). It may be concluded then that SRO is a highly promising candidate for fabrication of practical MIS switching devices.

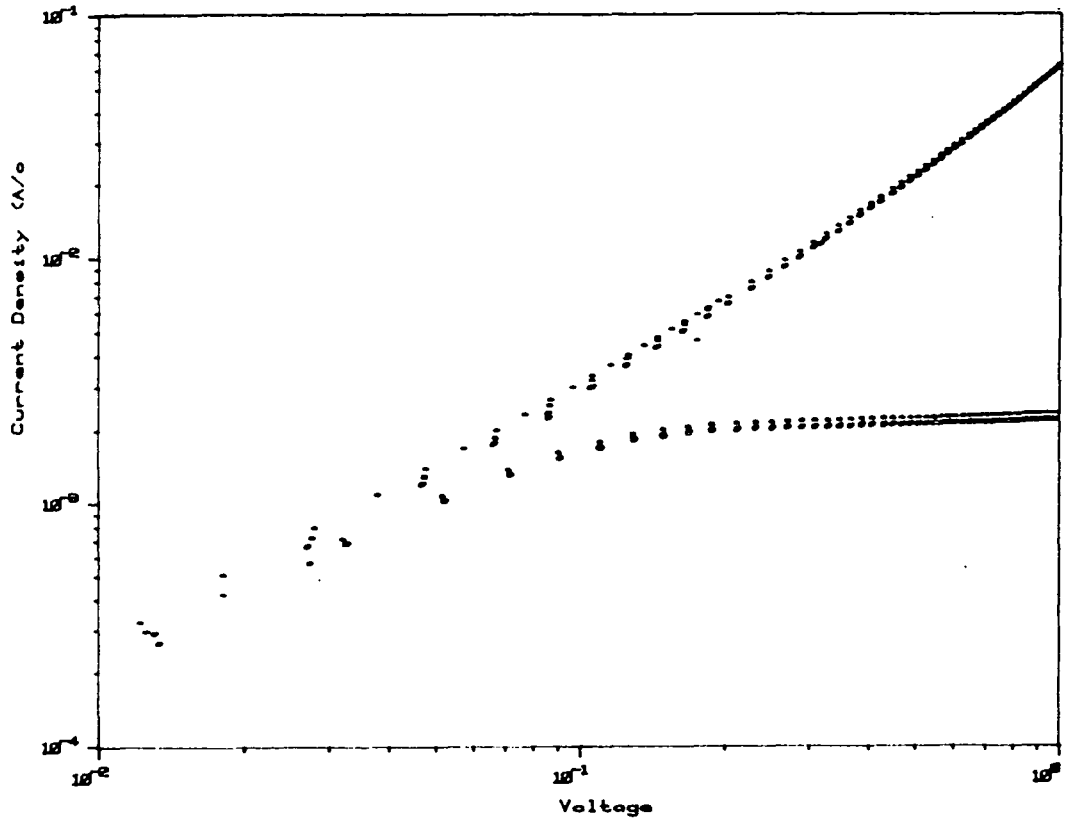


Figure 6.3; Forward and reverse I-V characteristics of a SRO MIS diode superimposed

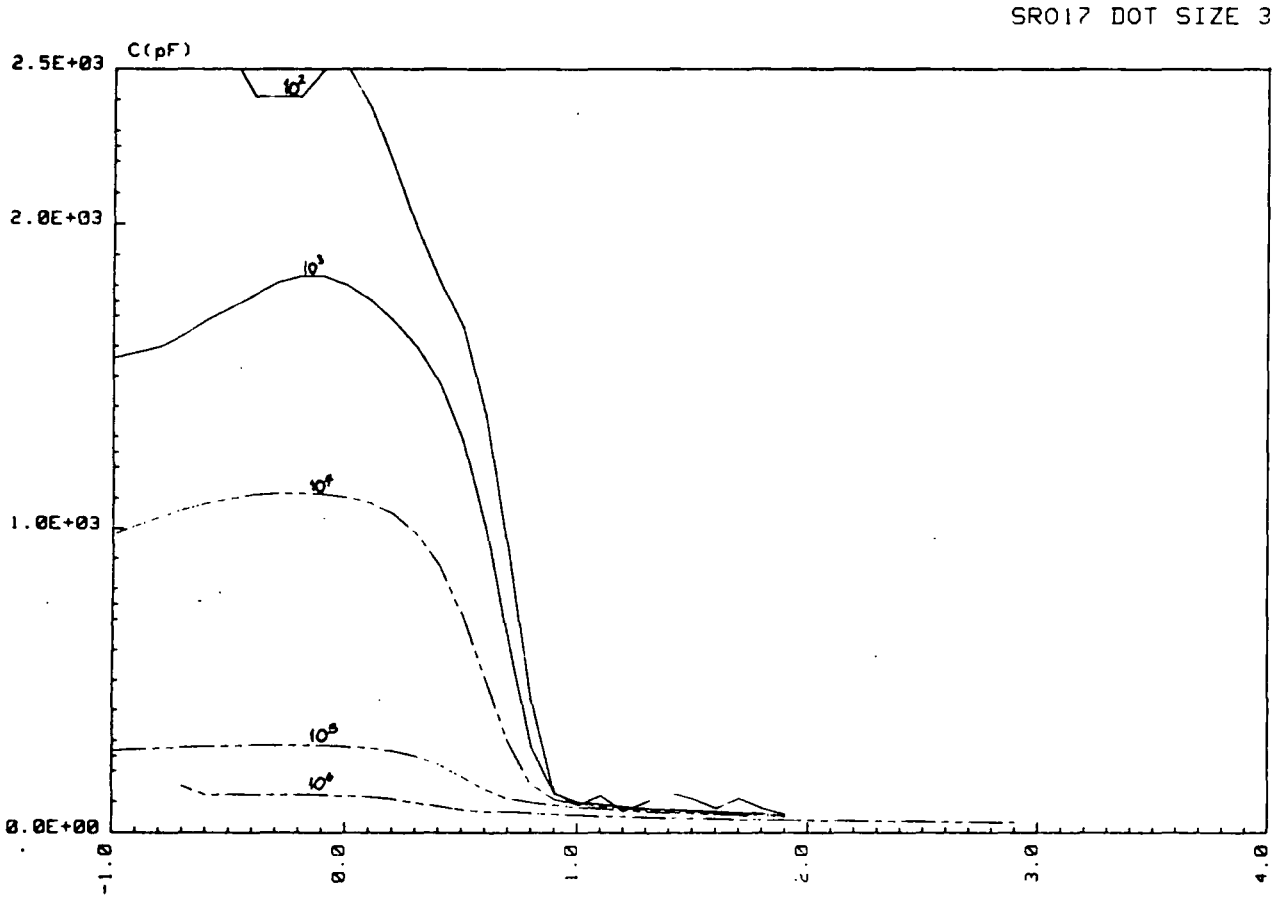


Figure 6.4; C-V characteristics obtained from a SRO MIS diode over five decades of measurement frequency.

REFERENCES

1. T.Matsushita, T.Aoki, T.Otsu, H.Yamoto, H.Hayashi, M.Okayama and Y.Kawana, Jpn. J. Appl. Phys. **15** 35 (1976)
2. H.Mochizuki, T.Aoki, H.Yamoto, M.Okayama, M.Abe and T.Ando, Jpn. J. Appl. Phys. **15** 41 (1976)
3. J.F.Verway, W.Ruis and I.Sens, Revue de Physique Appl. **13** 821 (1978)
4. D.J.DiMaria and D.W.Dong, J. Appl. Phys. **51** 2722 (1980)
5. D.J.DiMaria, K.M.DeMeyer, C.M.Serrano and D.W.Dong, J. Appl. Phys. **52** 4825 (1981)
6. T.Matsushita, N.Oh-uchi, H.Hayashi and H.Yamoto, Appl. Phys. Letters **35** 549 (1979)
7. G.W.Taylor and J.G.Simmons, IEEE Trans. El. Dev. **ED-32** 2345 (1985)
8. M.Hamasaki, T.Adachi, S.Wakayama and M.Kikuchi, Sol. St. Communications **21** 591 (1977)
9. J.R.Yeargan and H.L.Taylor, J. Appl. Phys. **39** 5600 (1968)
10. R.L.Petritz, Phys. Rev. **104** 1508 (1956)
11. M.L.Tarng, J. Appl. Phys. **49** 4069 (1978)
12. L.A.Nesbit, Appl. Phys. Letters **46** 38 (1985)
13. S.Verprek, Z.Iqbal and F.A.Sarrot, Philos. Mag. B**45** 137 (1982)
14. A.V.Dvurechensky, F.L.Edelman and I.A.Ryazanstev, Thin Solid Films **91** L55 (1982)
15. D.W.Dong and E.A.Irene, J. Electrochem. Soc. **125** (1978)
16. D.Buchanan, PhD. Thesis, University of Durham (1986)

CHAPTER SEVEN

Stable Negative Differential Resistance

7.1 Introduction; large area MISS devices

Although this thesis is mainly concerned with the properties of small geometry MIS switching devices, it has become clear during the course of this work that a great deal more can be learned about the switching mechanism itself by studying devices of larger area.

During the development of the CVD process for SRO, as described in chapter 6, there was a need to assess the suitability of new layers for MISS devices in a routine manner with minimal delay. Processing of small area devices required a minimum of two photolithography and etch steps (three for isolated devices) which would have introduced considerable lag in obtaining results to feed back into the SRO CVD development programme. For this reason, large area MISS devices were fabricated using a more straightforward sequence which simply involved evaporation of metal dots onto the SRO layers after they were deposited on epitaxial n-p^+ silicon substrates.

Many such large area MISS devices have been fabricated using SRO as the semi-insulating layer and the great majority of these have exhibited switching characteristics. However, poor reproducibility of the SRO layers has inevitably led to a lack of consistency in the electrical characteristics measured.

In addition, one wafer of large area MIS switches with a thermally grown tunnel oxide layer has been produced. However, only a few of the devices on this wafer exhibited any switching behaviour which is not surprising given the poor integrity of the thin oxide layers. Any attempt to produce a device which incorporates a sizeable area of tunnel oxide places a severe test on the uniformity of oxide growth, as discussed in chapter 3.

7.2 Measurement of MISS characteristics

Electrical contacts to large area dot devices have been made using the mechanical prober described in chapter 3 but with a gold ball rather than a tungsten needle probe because contact is direct to the cathode metallisation. The circuit used initially to plot their I-V characteristics, shown diagrammatically in figure 7.1, is similar to the one described in chapter 5 for measuring switching characteristics. The only addition is a high input impedance ($10^{12}\Omega$) unity gain buffer amplifier, connected as indicated to avoid current loading due to the $1\mathrm{M}\Omega$ input impedance of the x-axis input to the plotter.

The importance of the slope of the load line when measuring S-type (current-controlled) switching characteristics was discussed in chapter 5. It was demonstrated how a steep load line (low R_L) will intersect the S-type curve at three points, while a shallow load line (high R_L) cuts the curve uniquely at a single point whatever the applied bias. With no knowledge of the nature of the I-V characteristic between the switching and holding points of a MISS, it was only possible to state that the latter condition sets a minimum value for the load resistance given by;

$$R_L > -\left.\frac{\Delta V}{\Delta I}\right]_{I_S<I<I_H} \gg \frac{(V_S - V_H)}{(I_H - I_S)} \tag{7.1}$$

where $\left.\frac{\Delta V}{\Delta I}\right]_{I_S<I<I_H}$ is the negative differential resistance (NDR). If a lower value of series resistance were used, then on increasing the bias voltage above the switching point, the quiescent point would move suddenly to the intercept with the 'on' state curve, somewhere above the holding point. In this biasing mode, the MISS is forced to display a switching response with the device current appearing as a discontinuous function of voltage and there is no possibility of detecting any part of the curve between the low and high impedance portions. In the case where a large value of series resistance is used, it is possible to bias the device into the NDR section of its characteristic. This approach is useful if the switching and holding points are to be determined in a single sweep of the voltage source and has been used widely by other workers to measure MISS characteristics. Otherwise the holding point is missed on the increasing current sweep and the switching point is not observed when the supply voltage is decreased again back to zero. The upper limit on R_L, $R_L = \infty$, corresponds to an ideal current source and

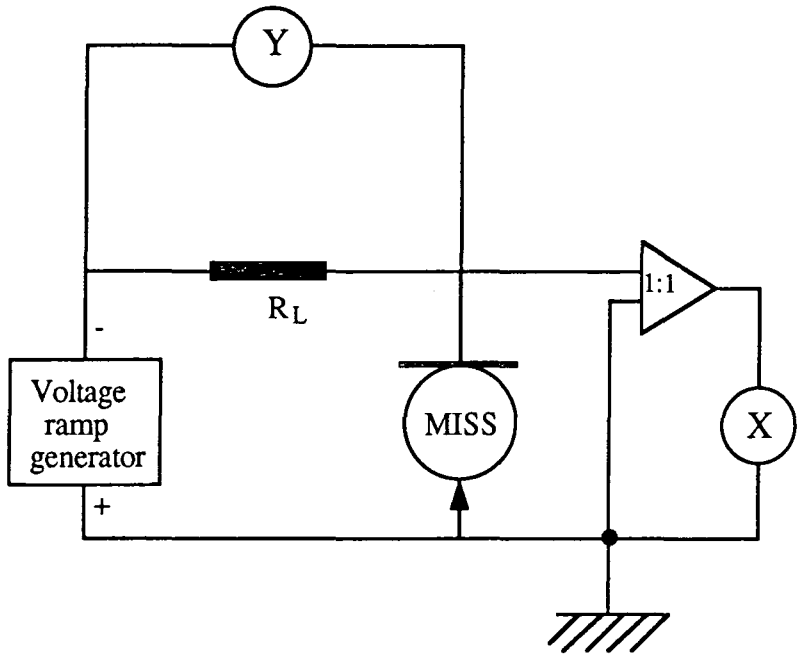


Figure 7.1; Biasing circuit used for plotting the switching characteristics of large area MISS devices.

it would seem that such a supply would constitute an ideal instrument for measuring current-controlled NDR characteristics. However, as will be discussed presently, such an approach has serious consequences for the investigation of the NDR region.

7.2.1 The First Observation of Stable Negative Differential Resistance

Following the argument of the previous section, a measurement circuit containing a large load resistance of 50kΩ was used to measure the I-V characteristics of a variety of large dot MISS devices. When measuring one of the largest examples, of area 2.56 mm^2, a quite unexpected result was observed. As the supply voltage was slowly ramped up from zero, the 'off' state curve was traced out smoothly as expected. However, on exceeding the switching voltage and passing through the turning point, the curve continued to be traced perfectly smoothly throughout the NDR region and into the 'on' state. Thus the current was plotted as an absolutely **continuous** function of voltage. The same measurement was repeated for a range of dot size as shown in figure 7.2(a). Furthermore, this S-shaped I-V characteristic could be retraced precisely in the reverse direction with no hysteresis and the device could be biassed in the steady state at any point on the curve. A similar curve has also been obtained from a tunnel oxide MISS device of large area and is shown in figure 7.2(b). Thus it appeared that the stability did not relate to a particular material property, but that it was a real effect in the MISS as a physical system.

When first observed, these results were remarkable in several respects:

(i) All current theories concerning the switching mechanism in MISS devices consider it to be an example of a current controlled regenerative feedback system. That is, the generally accepted approach is founded on thyristor theory. Clearly, the observation of stable negative differential resistance would appear to conflict with such a model. As discussed in chapter 4, switching in a regenerative feedback device is expected to occur as a single event at the point where the gain of the small-signal current feedback loop becomes equal to unity. Beyond the switching point, there should be no prospect of the device existing in a steady state condition until its quiescent point moves either on to the on-state curve (for which the loop gain returns to a value less than unity) or back onto the off-state curve. This principle of inherent instability in MISS devices forms the basis of their proposed application in high frequency oscillator circuits.

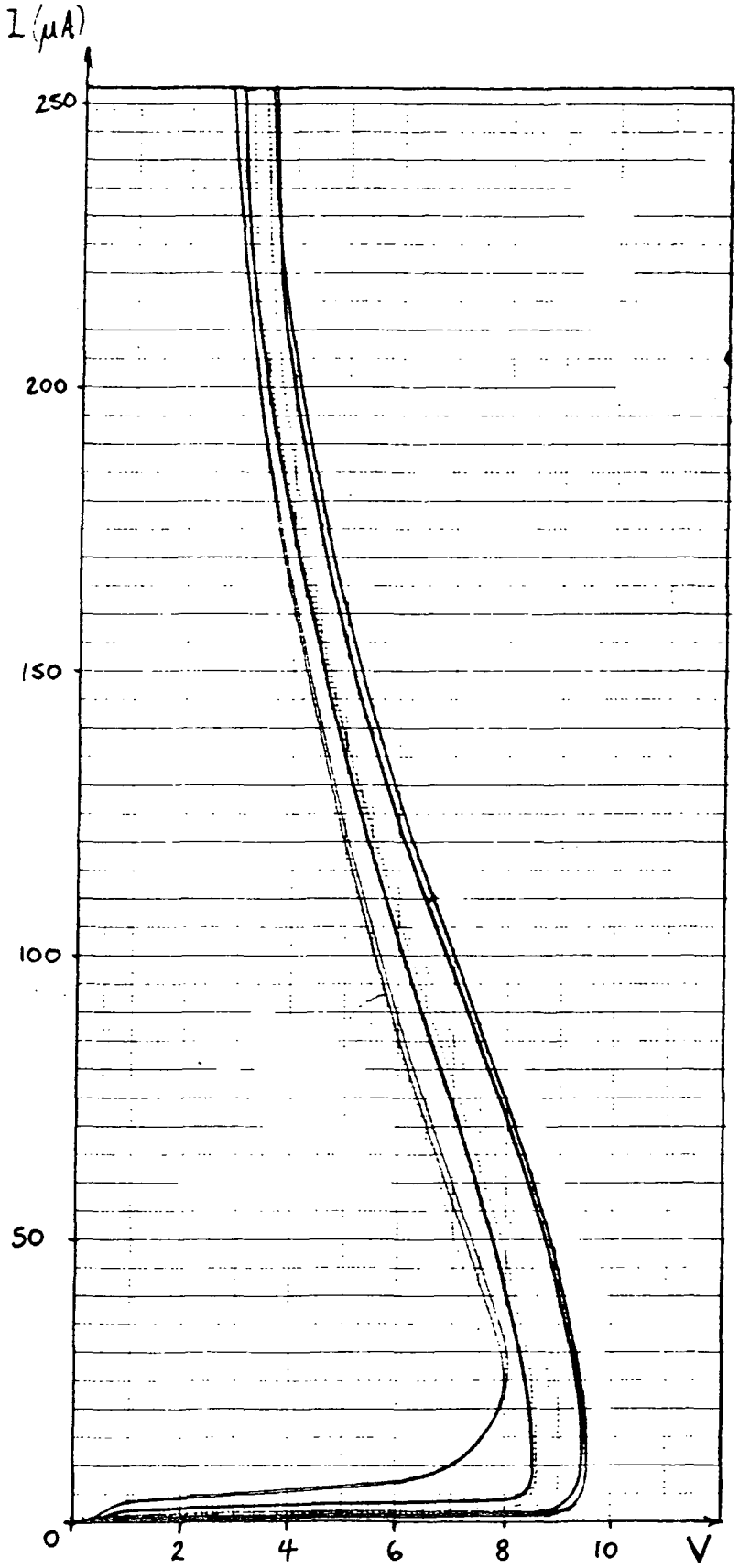


Figure 7.2(a) ; Stable NDR characteristics obtained from a range of large SRO MISS devices (Device areas 0.88 mm^2 -2.36 mm^2).

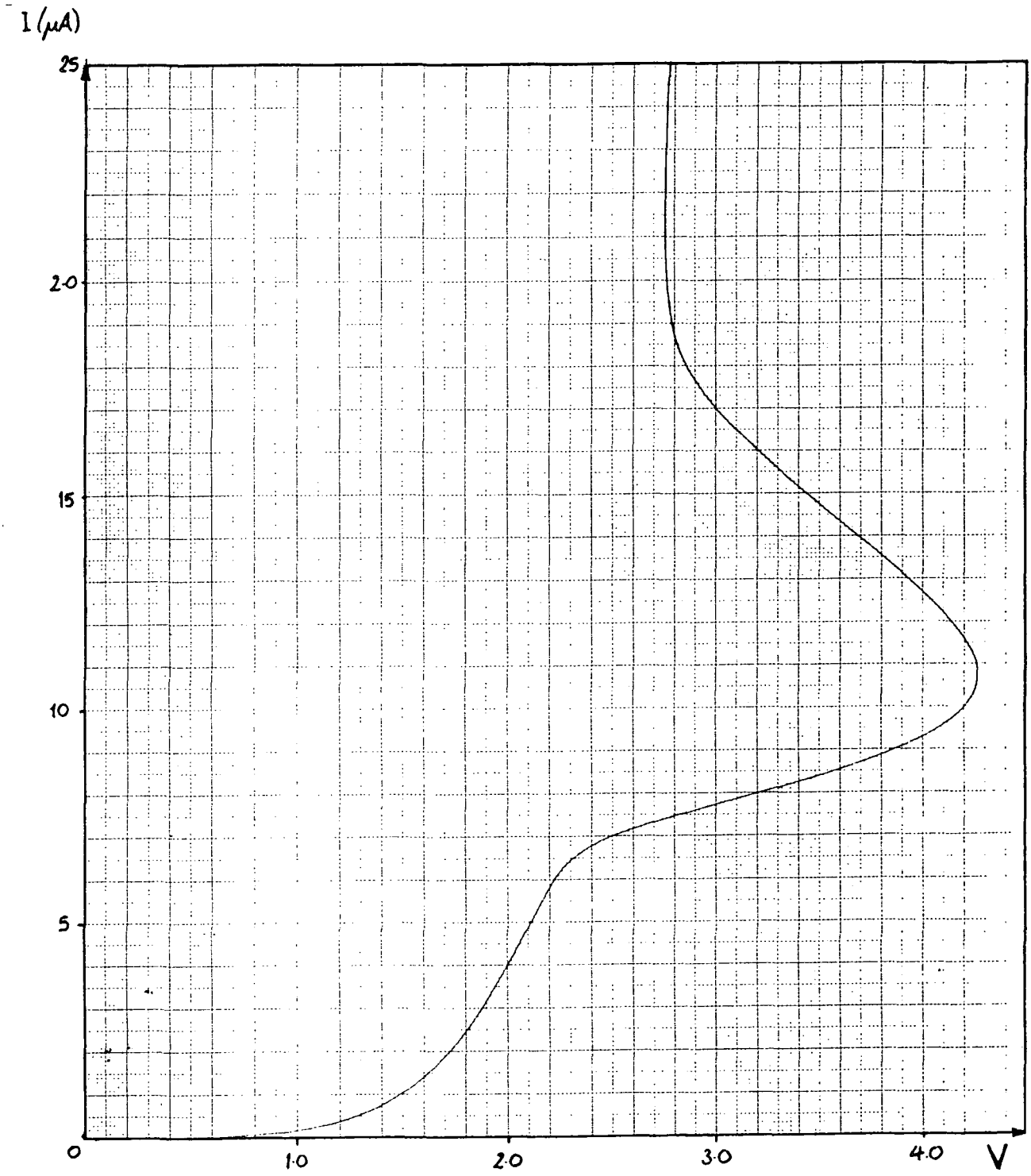


Figure 7.2(b) ; The first stable NDR characteristic obtained from a tunnel oxide MISS device. (Device area 0.17 mm^2, oxide growth time 6 minutes).

(ii)Even if a continuum of equilibrium states can exist in the N.D.R. region, circuit considerations suggest that this stability should not be observable with any real measurement circuit. In principle, any N.D.R. element placed in a circuit containing finite amounts of capacitance and inductance must cause that circuit to become unstable[1] if the condition of equation (7.1)is satisfied.

(iii) At the time of this discovery, the author was unaware of any characteristic demonstrating stable current-controlled N.D.R. being previously obtained from any form of regenerative switching device.†

S-shaped I-V characteristics with smoothly traced N.D.R. regions are normally associated with bulk effects in semiconducting materials. According to a classification of such devices proposed by Fritzsche[3] the MISS has until now been held to be an example of a 'switching device'. The new evidence presented here now requires that the MISS be re-categorised as a negative resistance device.

7.2.2 Stability and Device Dimensions

The stable curves shown in figure 7.2 were obtained from devices with large MIS contact areas. Devices with contacts smaller than a certain minimum area were found, when measured in the same circuit, to have NDR curves which were discontinuous over at least part of their range, figure 7.3. Instability is characterised by the departure of the plotted curve from a smoothly varying S-shaped function and is accompanied, when a mechanical pen plotter is used, by audible rapid vibration of the plotter pen. The extent of the unstable region is greater the smaller the device area and in all cases is located mid way along the NDR curve. The drawn curve is not, in general, smooth in the unstable state although it may appear so if there is sufficient inertia in the plotting device to damp out the oscillations and record time-averaged values of current and voltage. In such a case, the unstable region may take on the appearance of a third stable state intermediate between the 'on' and 'off' conditions, as was pointed out in chapter 5. There it was suggested that reported observations of a third state[4] and

† Since then, the author has found a solitary report of stable NDR in a thyristor. Wei[2] reports in a letter that 'negative resistance in quiescence' could be observed in a thyristor using a load of 500kΩ. Indeed, the I-V characteristics he shows are of a very similar form to those presented here. However, he fails to explain how stability can be reconciled with either of the conditions (i) or (ii).

Surprisingly, Wei's letter seems to have excited little interest and no subsequent comments on this effect appear to have been made.

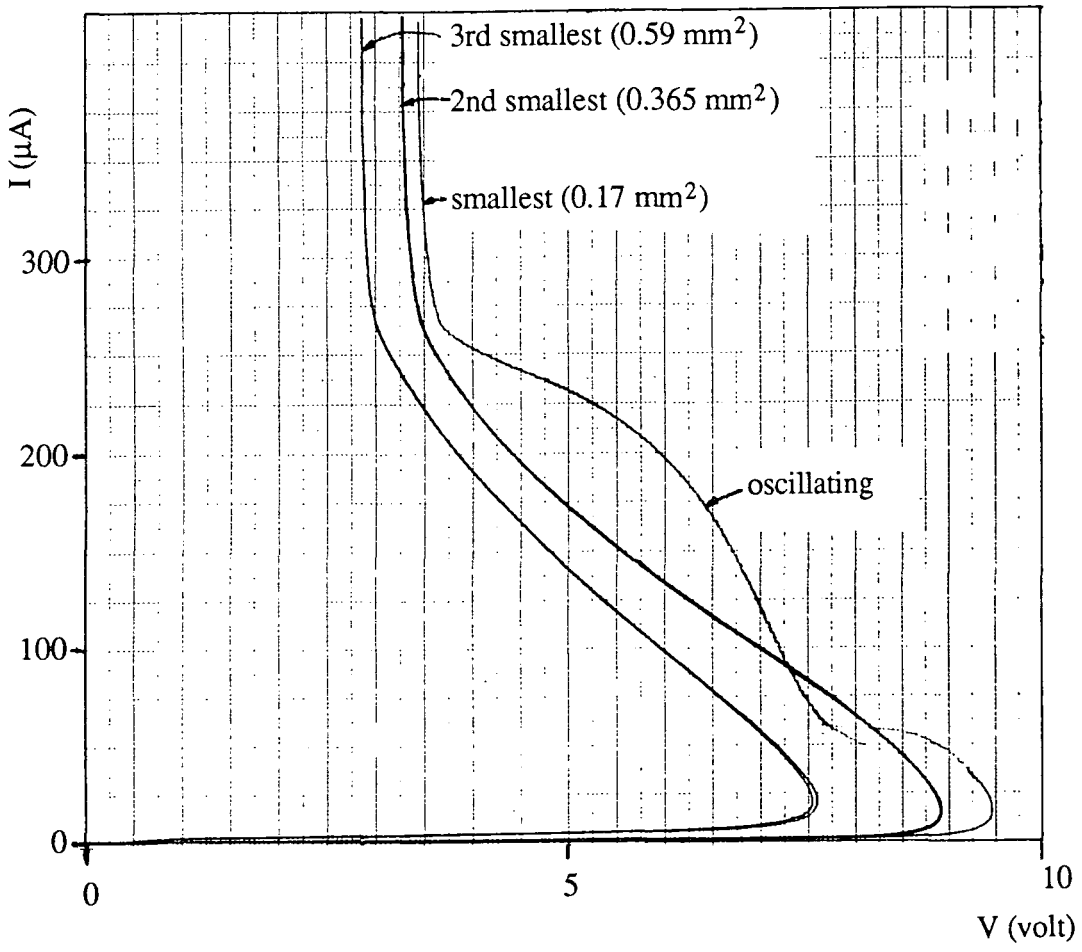


Figure 7.3; The switching characteristics of three SRO MISS devices on the same wafer but with a range of sizes. The smallest device displays an unstable NDR region.

the accompanying attempts to explain the phenomenon theoretically[5] are probably misguided and an unnecessary distraction from a proper understanding of the NDR effect. The question of why the size of a MISS device should influence the stability of its NDR characteristic will be addressed later in this chapter.

7.2.3 Detection of Oscillations using an Oscilloscope.

The conflict between the observable results of figure 7.2 and established theory will be considered presently. However, given the strength of the theoretical arguments against the existence of stable NDR in MISS devices, it was first important to test the stability of the NDR region more rigorously. To this end, an oscilloscope was included in the measurement circuit to monitor current fluctuations in the NDR region. This refinement required an additional current sensing resistor, R_S, in series with R_L to provide a voltage signal, as depicted in figure 7.4. Here the oscilloscope is not connected directly in parallel with the MISS, as would be the obvious arrangement for measuring voltage oscillations, because the parallel configuration was found to increase the possibility of the NDR region becoming unstable. In some cases, introduction of the oscilloscope in parallel could initiate oscillations in a device which had previously given a stable response.

Firstly, measuring a small area device which exhibited a discontinuous I-V characteristic, current oscillations were easily detected in the unstable portion of the curve, figure 7.3. These oscillations were consistently found to be highly asymmetric (skewed), as indicated in the diagram, with frequency and amplitude dependent on the biassing current. The nature of this dependence will be investigated later in this chapter.

Turning to a large area MISS whose current was a smooth and continuous function of voltage, it could in the same manner be shown that no oscillations were detectable, even on the most sensitive scale of the CRO, at any point on the NDR curve.

7.2.4 Use of a Constant Current Source

On first discovering that the NDR region was observable in some devices if a load resistor of sufficient magnitude were used, the assumption was made that a constant current source would provide an ideal biasing arrangement. The near infinite output impedance of such a supply provides a load line which lies essentially parallel to the V-

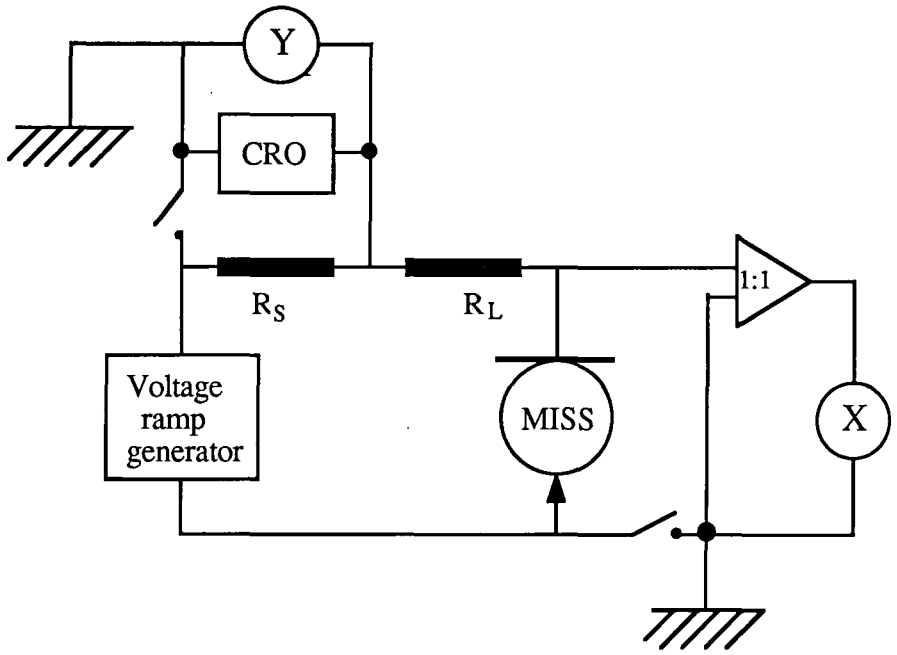


Figure 7.4; Connection of an oscilloscope for the purpose of detecting oscillations in the NDR region.

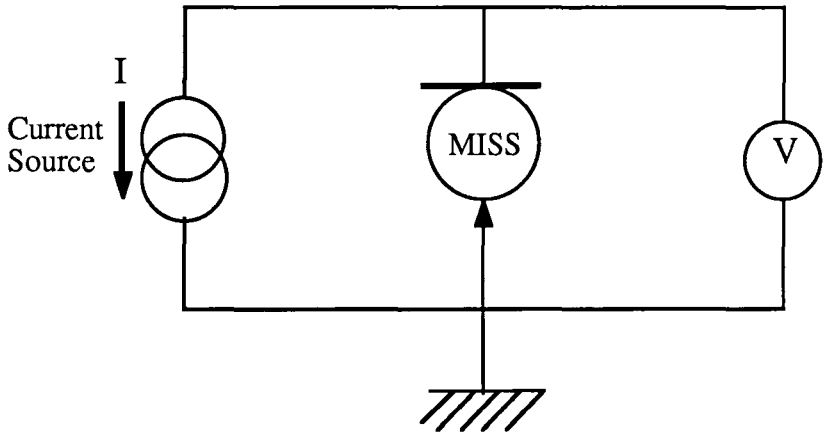


Figure 7.5; Implementation of a programmable current source for the measurement of switching characteristics.

axis of the I-V plot. To this end a Keithley 220 programmable current source with IEEE bus was purchased and implemented in the circuit of figure 7.5. Surprisingly, the use of this circuit only served to decrease the stability of the NDR, just as the oscilloscope had done. Some devices which had proved to have stable characteristics when using the resistive load now exhibited oscillations. As both the constant current source and the oscilloscope had a similar effect, it had to be assumed that an electrical property common to both, namely their internal capacitances, was responsible for forcing the device into an oscillating state.

7.2.5 The Effect of Capacitive Loading on NDR Stability

Having surmised that it was the reactive loading presented by the current source and the oscilloscope that was responsible for the reduced stability of the NDR region, it was easily shown experimentally that a parallel capacitance greater than a certain critical value did indeed initiate oscillations in a normally stable device.

Figure 7.6 shows the I-V characteristics obtained from a device which was nominally stable when measured using the circuit of figure 7.7. In this arrangement, care has been taken to minimise the stray reactance in parallel with the MISS. The x-scale (V-axis) of the pen plotter is driven by the analogue output from a Keithley 617 electrometer,the input capacitance of which is effectively de-coupled from the MISS by the use of the two 1MΩ resistors. Given that the input impedance of the electrometer is of the order of $10^{12}\Omega$, these resistors have negligible effect on the measurement of the potential difference across the device. The load resistance, R_L, is composed of three elements with the intention of de-coupling the capacitances of the voltage source and oscilloscope.

It can be seen from the results of figure 7.6 that although the NDR remained stable when a parallel capacitance of 94pF was introduced, increasing the capacitance progressively to 220, 470 and 910pF had the effect of initiating an oscillatory response over a progressively larger portion of the NDR curve. Thus it became apparent that a MISS device can exhibit partial stability, where oscillations only occur over a fraction of the NDR region, as was observed for the smaller devices in figure 7.3. It can now be concluded that the extent of the stable zone depends on the magnitude of the capacitive load as well as the size of the device being measured. Also the the NDR is most stable near the turning points of the I-V curve, where the slope $\partial I/\partial V$ is greatest.

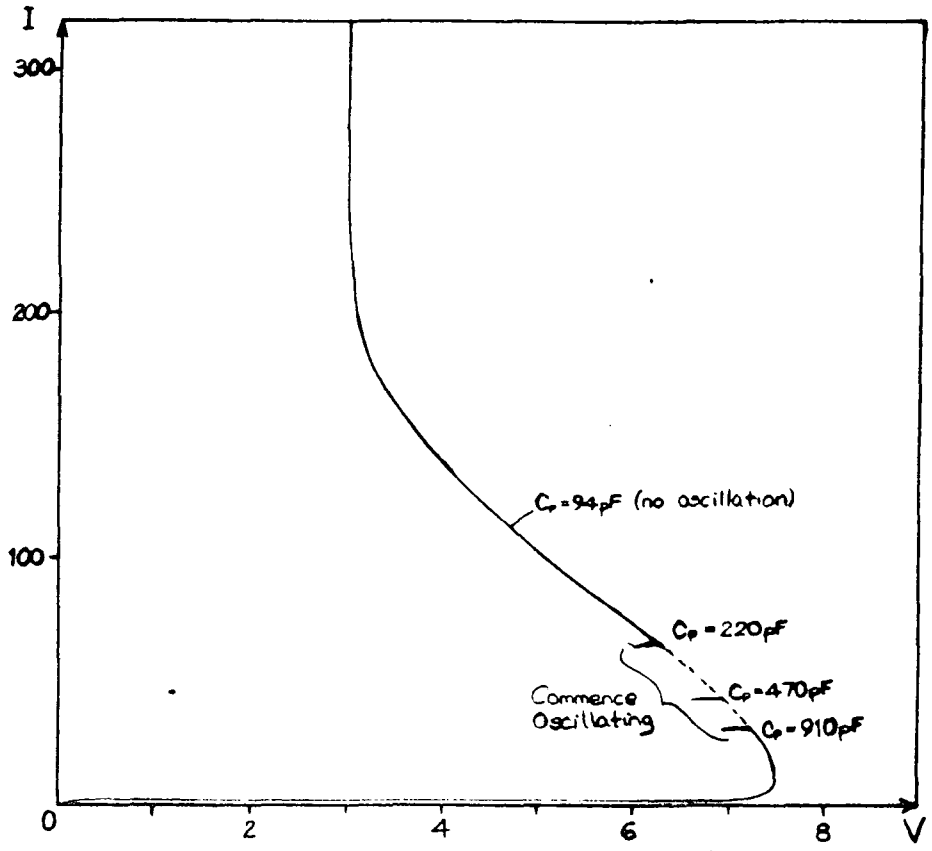


Figure 7.6; The stable NDR characteristic of a MISS device of area 1.24mm^2 indicating the points of onset of instability arising from the addition of a range of values of parallel capacitance, C_P

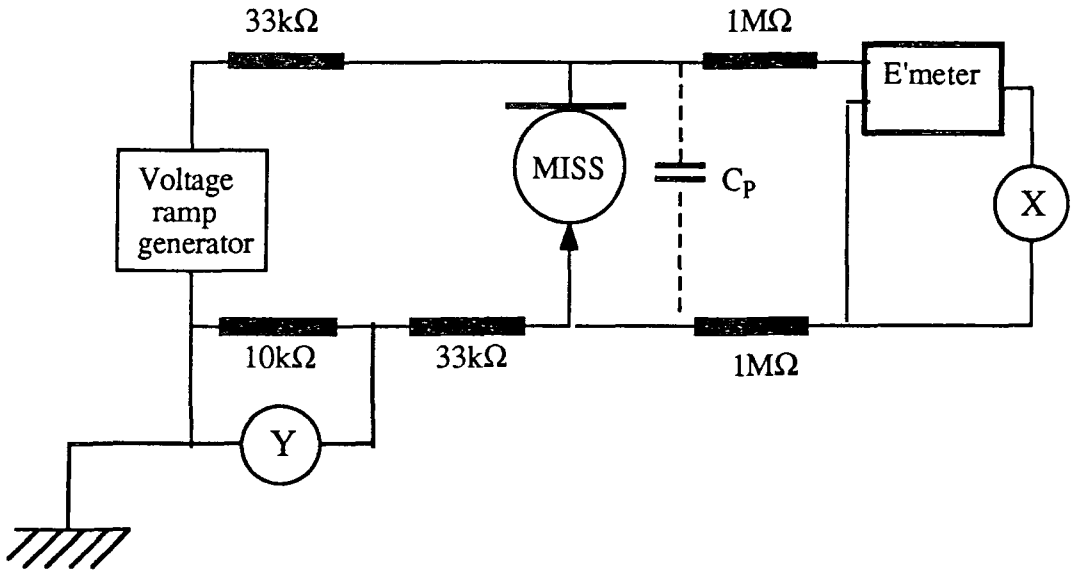


Figure 7.7; The circuit used to minimise the stray reactive load on the MISS device.

7.3 Circuit Stability and the Concept of Negative Differential Capacitance.

One of the principle theoretical questions arising from the observation of stable NDR can now be addressed. In essence, this question can be stated:

How can an element which exhibits current-controlled NDR be placed in a real circuit containing finite amounts of capacitance and inductance without that circuit displaying an unstable response ?

Hines[1] has considered the importance of the load line on the stability of a circuit containing a voltage controlled ('N-type') NDR element. He showed that if such a device is biassed to some point on its NDR curve, then the circuit containing that device will only be stable if the total series load resistance is within a range given by;

$$\frac{L}{R_d.C_d} < R_L < |NDR| \tag{7.2}$$

where |NDR|is defined as;

$$NDR = \left.\frac{\partial V}{\partial I}\right]_{I_S<I<I_H}$$

Thus the NDR region of a device with an 'N-type' I-V characteristic can be traced out if a low value of R_L is used and the same should be true for an 'S-type' curve. However, the condition for the load line stated in equation (7.1) precludes the use of such a low value when tracing an S-type characteristic. The two requirements are mutually exclusive.

Nevertheless, it has been shown here that an S-type characteristic can be traced in direct contravention of Hines' analysis. A highly innovative approach to this dilemma has been proposed by Morant[6] with a solution which invokes the concept of negative differential capacitance (NDC). This concept is not entirely new with regard to the MISS device. Millän et. al.[7] effectively derived this result from an analysis of the switching effect based on a charge control model. They predicted that the differential, or small-signal, capacitance given by $\partial Q/\partial V$ goes instantaneously to infinity and changes sign at the switching and holding points. Between these points, (V_S, I_S) and (V_H, I_H), it

assumes a finite negative value;

$$\left.\frac{\partial Q}{\partial V}\right]_{V=V_S, V=V_H} \rightarrow -\infty$$

$$\left.\frac{\partial Q}{\partial V}\right]_{V_H<V<V_S} < 0 \tag{7.3}$$

These results are easily explained. In evaluating $\partial Q/\partial V$ at any bias point, the total charge Q_S in the semiconductor is mirrored by the charge on the metal cathode, Q_M, which is given by Gauss' law;

$$Q_M = A_{ins}.\epsilon_{ins}.\mathcal{E}_{ins} = A_{ins}.\epsilon_{ins}.\frac{V_{ins}}{d_{ins}} = C_{ins}.V_{ins} \tag{7.4}$$

where the terms are as defined in chapter 2.

Now, taking the case of a tunnel oxide device, it can be shown from the electron and hole tunnelling current equations that the potential dropped across the insulating layer, V_{ox} must increase monotonically with increasing total device current I_{TOT};

$$I_{TOT} = I_{NT} + I_{PT} \tag{7.5}$$

where from (2.23);

$$I_{NT} \propto \exp V_{ox} \tag{7.6}$$

and, from equations (2.29), (B6) and (2.8) respectively;

$$I_{PT} \propto p(0) \quad , \quad \mathcal{E}_S \propto \left[p(0)\right]^{1/2} \quad \text{and} \quad \mathcal{E}_S \propto V_{ox} \quad \text{giving} \quad I_{PT} \propto V_{ox}^2 \tag{7.7}$$

Thus as the current drive is increased, the stored charge on the cathode plate must also increase even in the NDR region where the voltage is decreasing. It follows that at both turning points, ΔQ is positive with ΔV equal to zero and that in the NDR ΔQ remains positive while ΔV is negative, as stated in equation (7.3).

Now it must be shown that a negative-valued differential capacitance can provide a stable response in a circuit composing a NDR element. A full treatment of this problem is given by Majlis[8] in terms of the Laplace transforms of the system and a summary of the analysis is given in Appendix A of this thesis. The response of a generalised circuit with lumped L,C,R elements to a small perturbation, δV_S, of the supply voltage is shown to be;

$$h(t) = \frac{\delta V(t)}{\delta V_S} = \frac{1}{L.C.\sqrt{\beta}}\left[\exp -\left(\alpha + \frac{\sqrt{\beta}}{2}\right).t - \exp -\left(\alpha - \frac{\sqrt{\beta}}{2}.\right).t\right] \tag{7.8}$$

where $V(t)$ is the voltage across the MISS device and;

$$\alpha = \frac{L + R.R_L.C}{2.L.C.R} \quad \text{and} \quad \beta = \left(\frac{L + R.R_L.C}{L.C.R}\right)^2 - \frac{4.(R + R_L)}{L.C.R} \tag{7.9}$$

As shown in Appendix A, in any case where the resistance is negative ($R = -|R|$) and the capacitance is positive, this response will only decay with time (and thus become stable) if α is positive and β is negative. Under these conditions, the impulse response becomes;

$$h(t) = \frac{2}{L.C.\sqrt{\beta}}.\exp(-\alpha t).\sin\left(\frac{\sqrt{\beta}}{2}.t\right) \tag{7.10}$$

and therefore constitutes a decaying sinusoidal oscillation.

The circuit conditions required for positive α and negative β are shown to be;

$$\frac{L}{|R|.C} < R_L < |R| \tag{7.11}$$

which are identical to the stability criteria of Hines, equation (7.2). These requirements are inconsistent with the condition for the load line when measuring current-controlled, S-type NDR (7.1) and therefore do not allow the possibility of observing stable NDR in such devices in the presence of net positive capacitance.

However, in any case where both the differential resistance **and** capacitance are negative ($R = -|R|$ and $C = -|C|$), both of the exponents $(\alpha + \sqrt{\beta}/2)$ and $(\alpha - \sqrt{\beta}/2)$ in (7.8) will be positive for all values of R_L greater than $|R|$, leading to a decaying impulse response. Thus if the net parallel capacitance is negative, the load line condition (7.1) may be met and stability in the NDR region maintained.

7.3.1 Measurement of Negative Differential Capacitance

It has been proposed that the essential condition for circuit stability of S-type NDR is that the **net** capacitance in parallel with the NDR element be negative valued. In such a case, the response of the network to a small perturbation (e.g. a delta function) will be a rapidly decaying signal. The point at which this stability condition breaks down is therefore where the total positive capacitance existing in parallel with the MISS exceeds the value of the NDC. Thus, referring back to figure 7.6, it can be seen that the introduction of increasingly large capacitors in parallel with the MISS, causes the stability condition to be violated over increasingly large sections of the NDR curve. This observation led Morant[6] to propose an elegant method for the measurement of NDC.

In the absence of stray capacitance, the NDC at any stable point on the NDR curve of a MISS characteristic is equal in magnitude to the minimum parallel positive capacitance required to initiate oscillations at that point.

Figure 7.8 represents the first measurement of NDC thus performed on a MISS device with a tunnel-oxide layer. The NDC of this particular device (dot area 0.17mm^2) has quite a large value at all points, which explains its relatively high stability. Immediately it can be seen that the magnitude of the NDC bears some relation to the slope of the I-V curve. Such a dependence has been observed repeatedly by Majlis[8] who explores this measurement technique in some detail in his thesis. This type of relationship between $|NDC|$ and $(\delta I/\delta V)$ can be predicted, at least qualitatively, for a tunnel oxide MISS from the tunnelling equations.

The NDC can be given by the expression;

$$NDC = \frac{\partial Q_M}{\partial V_{TOT}} = C_{ins} \cdot \frac{\partial V_{ins}}{\partial V_{TOT}} \tag{7.12}$$

For a tunnel oxide layer in a case where electron tunnelling is significant, this relationship will in general be non-linear since, substituting for V_{ox} from (7.6);

$$NDC \propto C_{ins} \cdot \frac{V_T}{I_{NT}} \cdot \frac{\partial I_{NT}}{\partial V_{TOT}} \tag{7.13}$$

Thus, it is expected that the small signal negative capacitance should be some function of the slope of the I-V curve. Results in chapters 3 and 5 for the tunnel oxide

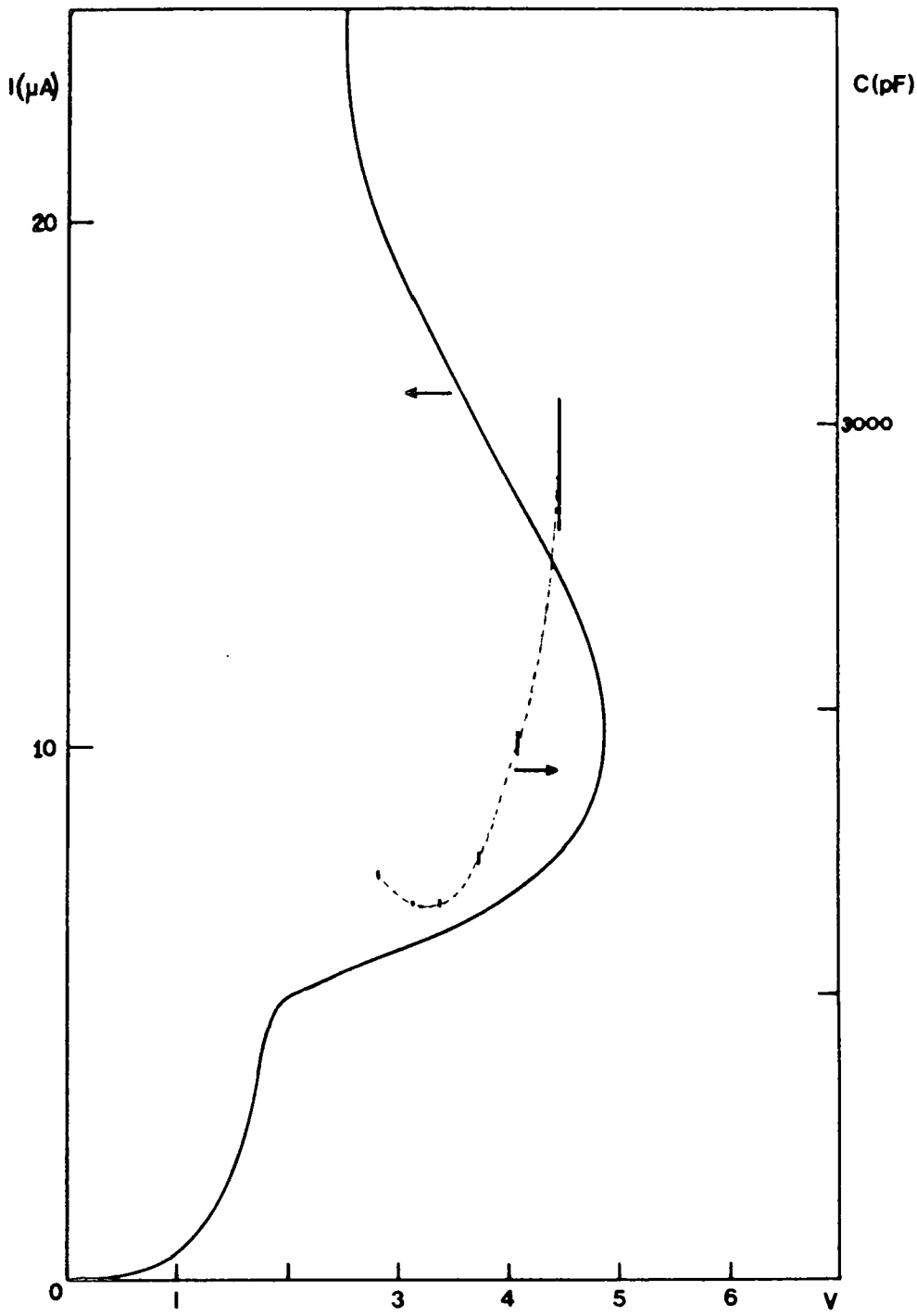


Figure 7.8; The first measurements of Negative Differential Capacitance obtained from a tunnel oxide MISS device using the method described in the text.

MIS and MISS respectively indicated that the electron tunnel current is dominant for the Al-SiO_2 -Si(n) structure. Thus, the NDR.NDC product would not be expected to remain constant in this system, as found by Majlis[8] .

It has thus been shown quite convincingly that the **in-circuit** stability of the NDR of MISS devices can be accounted for on the basis of negative differential capacitance. It remains to be shown, however how this stability can be reconciled with the concept of current-controlled regenerative feedback. Some insight into this problem has been gained through experimental studies of the nature of the unstable response of MISS devices.

7.4 The Unstable State

It has already been shown that instability can be produced in an otherwise stable I-V response if a capacitance greater than the NDC is connected in parallel with the MISS. This provides a means for investigating the unstable state in a controlled manner.

7.4.1 Frequency of Oscillation

Initial results obtained using this approach from a small dot SRO MISS device of area 0.16 mm^2 are shown in figure 7.9. Here the oscillation frequency is plotted as a function of device voltage for three values of parallel capacitance C_P and the steady state I-V characteristic of the MISS is superimposed. Clearly larger drive currents increase the frequency, and larger capacitances cause a reduction in frequency, as is expected when an RC time constant is involved. An additional feature of these curves is that the extent of the unstable region increases with the magnitude of the parallel capacitance in agreement with the NDC stability theory.

Assuming an RC time constant is important, such that $f \propto (R_L.C_P)^{-1}$, then the frequency dependence should be removed by taking the product $f.C_P$. This product is plotted in figure 7.10 as a function of steady state current, which is more representative of the frequency variation. The fact that the three curves fall reasonably well on the same straight line confirms the inverse proportionality between f and C_P which will be discussed in the following section.

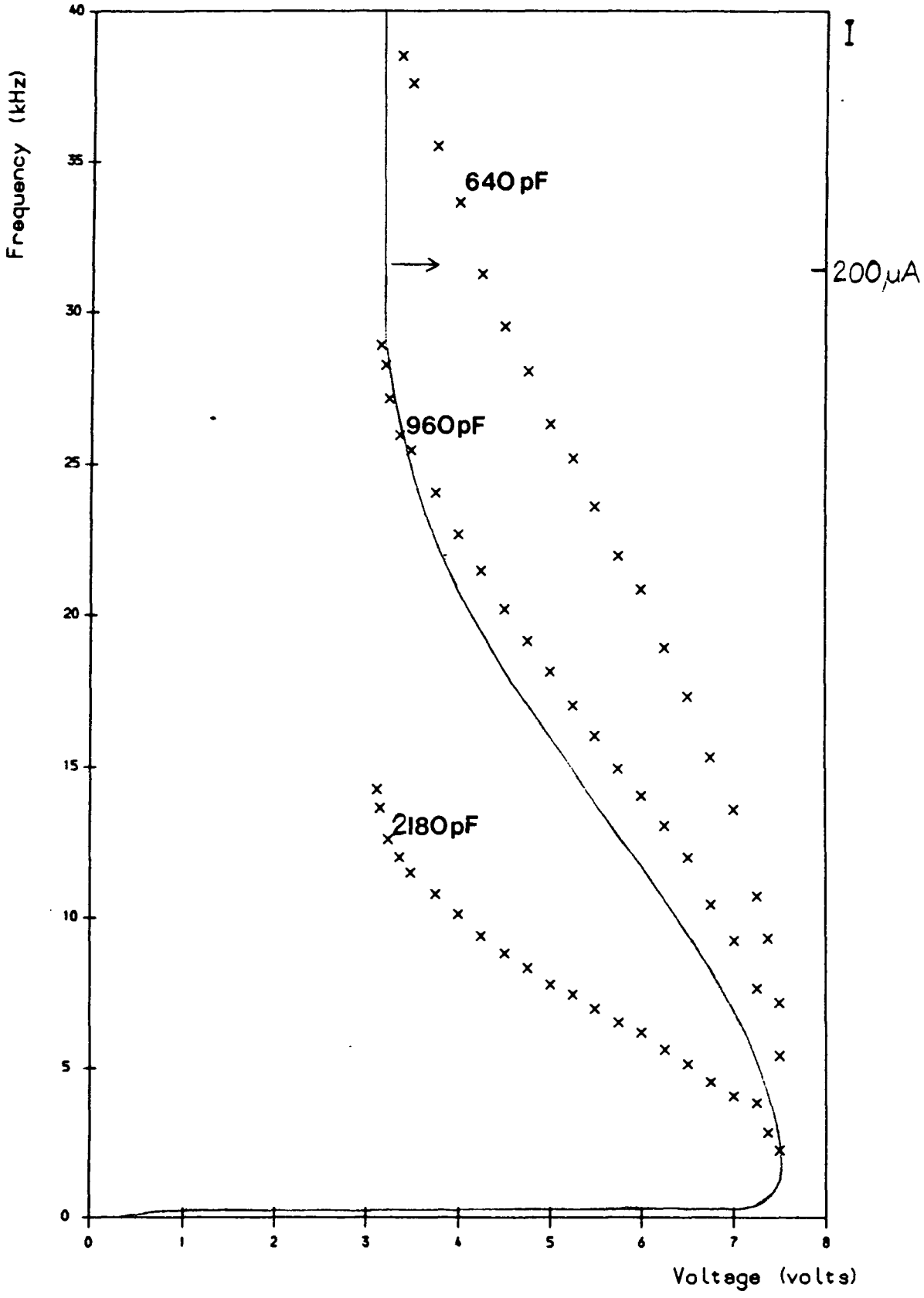


Figure 7.9; Frequency of oscillation of an unstable MISS device as a function of voltage for various values of parallel capacitance

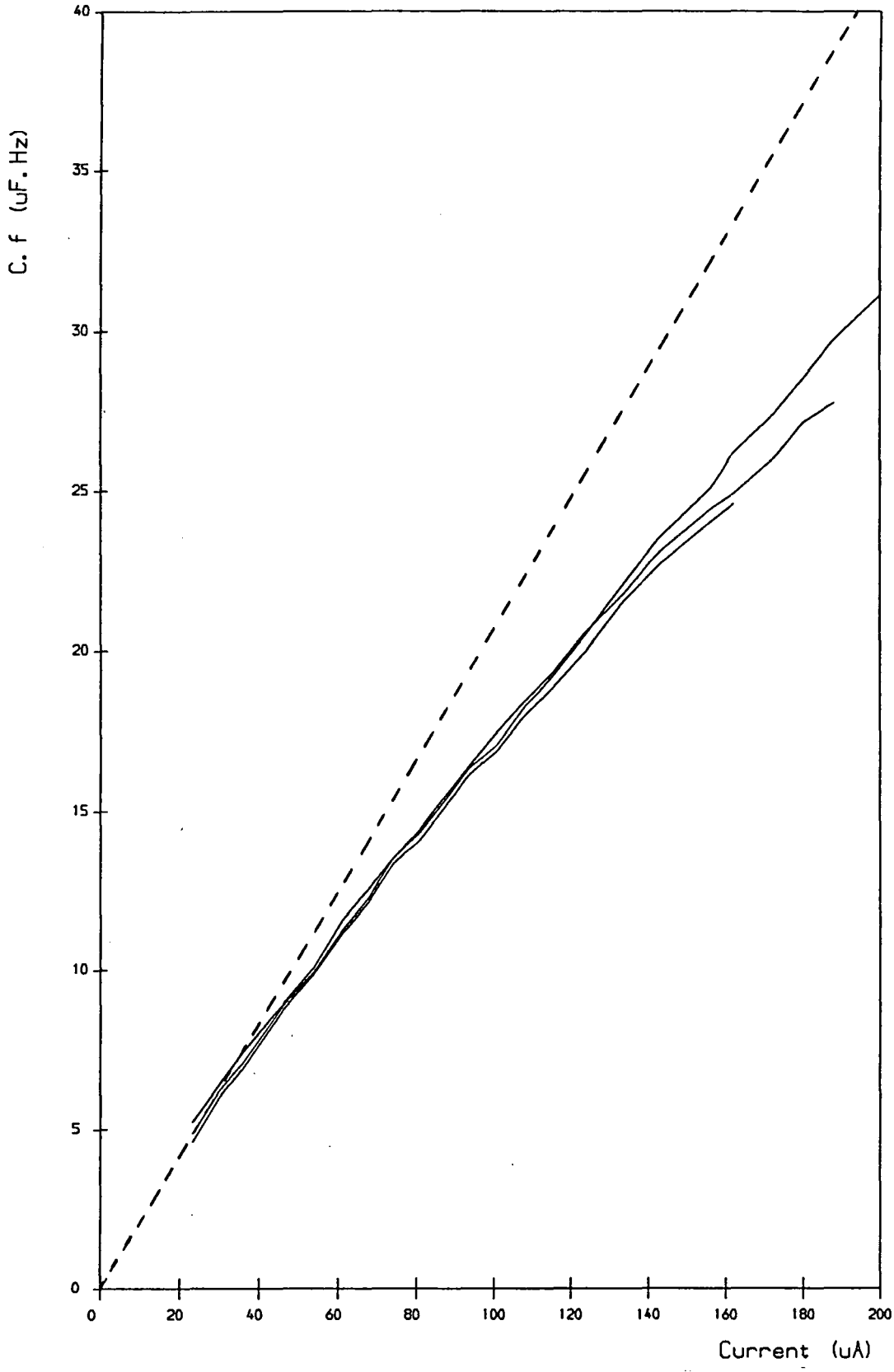


Figure 7.10; The capacitance.frequency product plotted as a function of total current using the same data as figure 7.9.

7.4.2 Oscillations in Small Area Devices

In order to measure the amplitude as well as the frequency of oscillations in the unstable state, the biasing circuit of figure 7.4 was extended to include an oscilloscope and frequency meter, as detailed in figure 7.11. An electrometer was used in place of the buffer amplifier because it has a triaxial input and thus can be operated with both of its input terminals floating with respect to ground. This is important in the present case where the CRO determines the grounded node of the whole circuit and earth loops are to be avoided. As mentioned earlier, connection of the CRO directly in parallel with the MISS is undesirable as it would introduce a sizeable capacitive load. Load resistor R_L is composed of three elements which serve to isolate the MISS from the CRO, frequency meter and voltage sweep generator. The CRO detects current oscillations in the form of a voltage signal produced across the 10kΩ resistor which is also registered in time-averaged form on the 'y-'(current) axis of the plotter.

Subsequent measurements using this circuit have been obtained from MISS devices on a chip which was mounted and ultrasonically gold ball bonded in an 8-pin integrated circuit package. This allowed the measurement circuit to be constructed in a more compact form on a circuit board and thus avoid use of the mechanical probe contact system. By this means it was hoped to reduce stray capacitance which would be more important at higher frequencies. In this case, the devices studied had to be of the oxide-isolated type because these have associated bond pads, as described in chapter 5. It was not thought wise to attempt ultrasonic gold ball bonds directly onto the thin cathode metal of a dot type device.

Figure 7.12shows the I-V characteristic of a MISS of square geometry with sides of 20μm and thus an area 0.0025 times the area of the previously studied dot device. It is evident that this device is unstable even before any parallel capacitance is introduced despite the effort made to minimise parasitic capacitances. Such oscillatory behaviour has been misconstrued by other workers as representing an intrinsic physical instability in the functioning of MISS devices. The results shown here suggest that instability is not an inherent property of the MISS; it is only brought about by the presence of capacitance external to the device. In the case of small area devices, it would appear that very small but non-negligible stray capacitances are sufficient to cause oscillation.

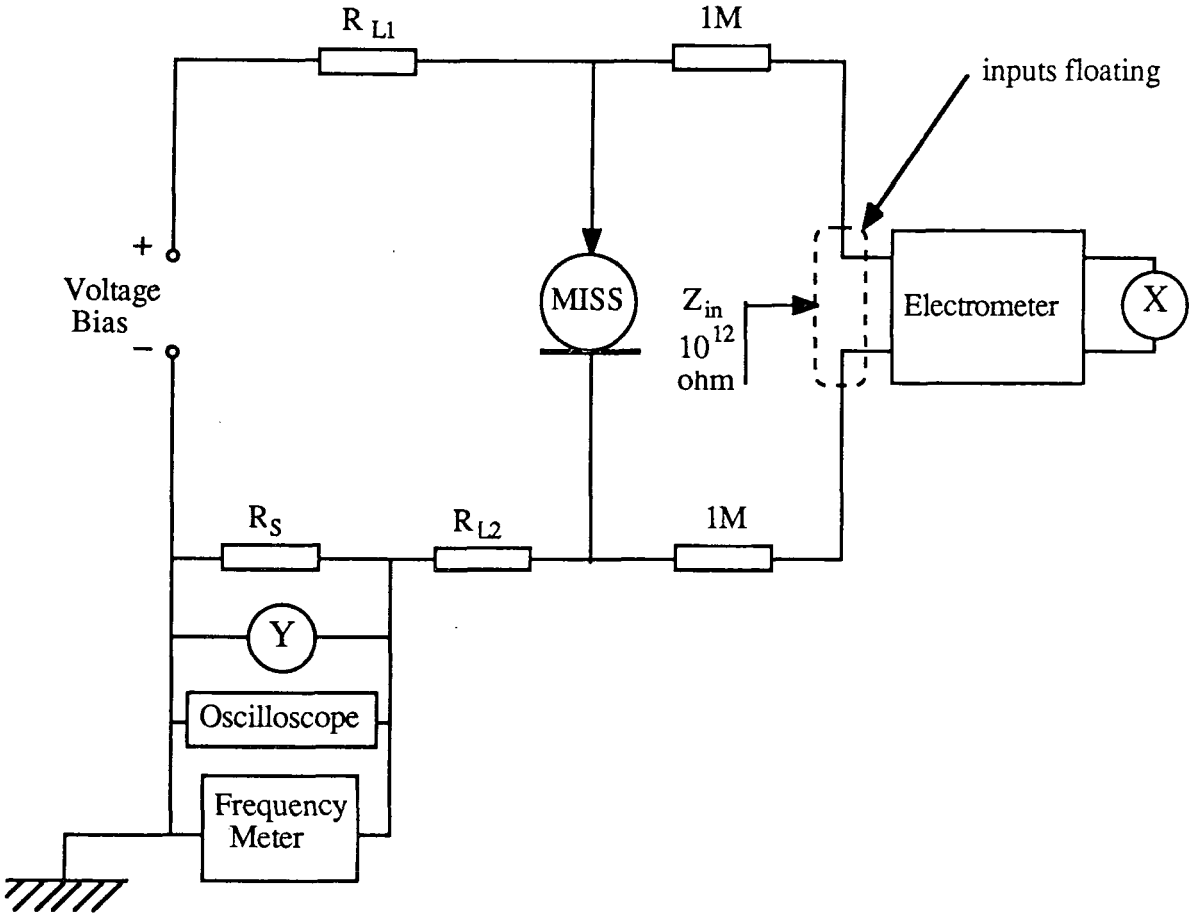


Figure 7.11;. Circuit used to measure the frequency and amplitude of oscillations in the NDR region.

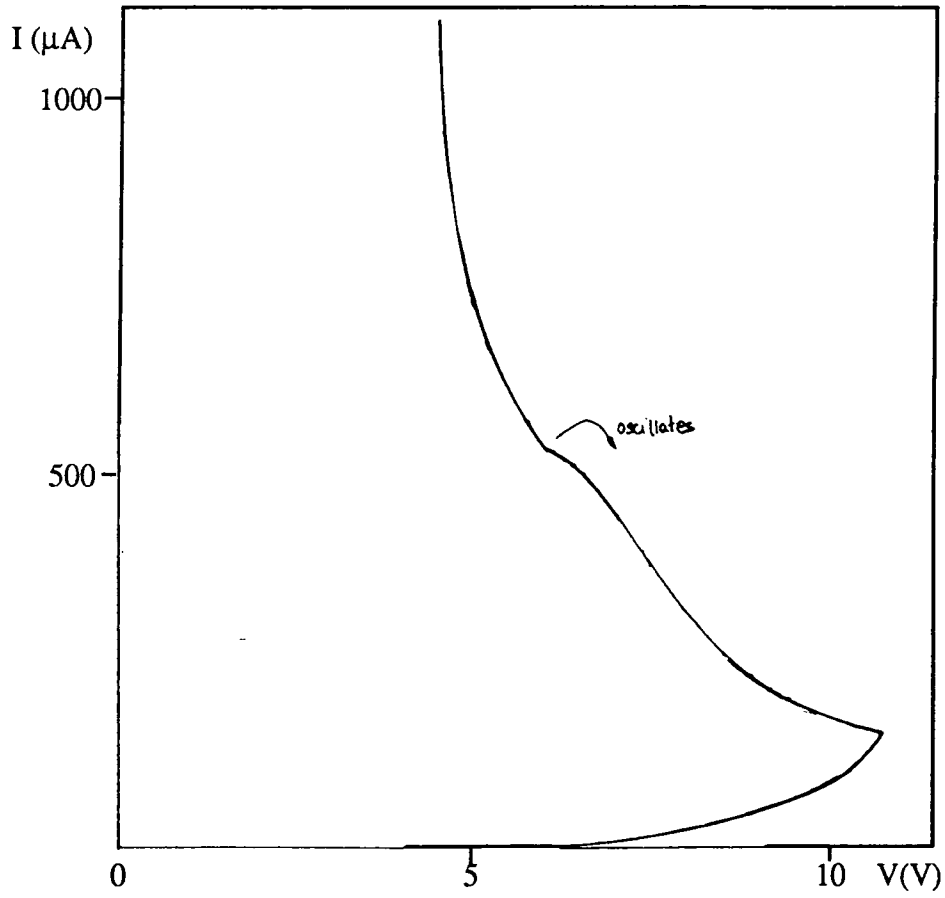
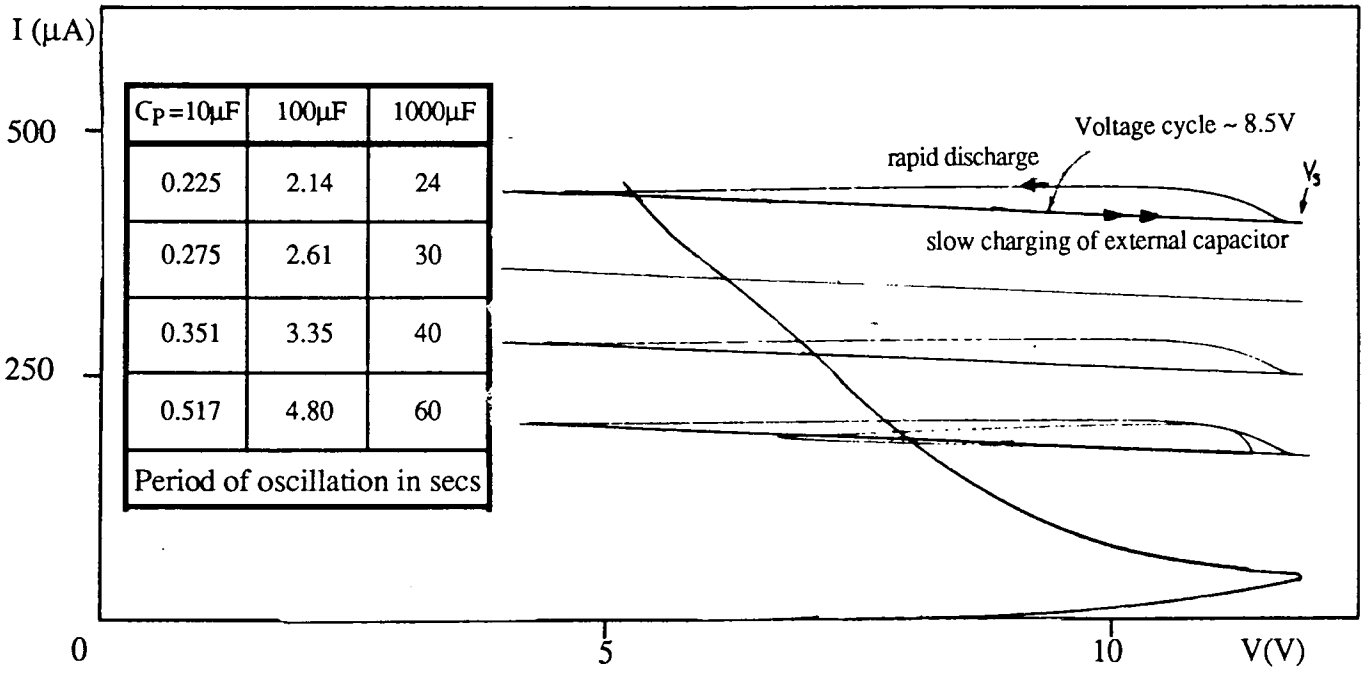


Figure.7.12; The switching characteristic of a [20μm] ^{2}SRO MISS showing a discontinuous (unstable) NDR curve with no added parallel capacitance.



Cp=10μF	100μF	1000μF
0.225	2.14	24
0.275	2.61	30
0.351	3.35	40
0.517	4.80	60
Period of oscillation in secs		

Figure 7.13; The locii of the current / voltage oscillations in the above device caused by adding large values of parallel capacitance as indicated.

7.4.3 Very Low Frequency Oscillations

An interesting result has been obtained from the above mounted device by connecting relatively large value capacitors in parallel and thus increasing the period of oscillation in the unstable NDR region to the range of seconds. This has enabled the pen plotter to follow the time-varying MISS current and voltage and has made it possible to plot their locus throughout a cycle of oscillation. Such locii are shown in figure 7.13, as plotted, for a range of capacitance values and quiescent biassing points.

Immediately, it is clear that the locii lie on the respective load lines with their slopes corresponding to the total series resistance in the current loop, 210kΩ. The straight lines were drawn slowly in the sense of increasing voltage as far as the switching point, whereupon the pen returned rapidly, describing a curve, to a point where the voltage was near but not always equal to the steady state holding voltage. The slow traverse along the load line was understood to correspond to the charging of the external capacitor, a conclusion supported by the proportionality between the cycle periods and the capacitances. Although the locii appear to depart from the load lines in the subsequent discharge phase, this effect is more likely due to the inability of the pen plotter to respond to rapid changes.

This observation of the locus of the unstable state conflicts with the assumption made by several authors that oscillations occur along the steady state NDR curve of the MISS. On reflection, it is clear that such a suggestion contravenes Kirchoffs current law. In order to interpret this result it is first noted that the total current through the loop, I_{TOT} as measured on the y-axis, is composed of two parts; the conductance current through the MISS, I_{MISS} and the displacement current through the external capacitor, I_C (the capacitance of the MISS is considered negligible for present purposes):

$$I_{TOT} = I_{MISS} + I_C \tag{7.14}$$

It can now be shown that nearly all of the current is displacement current. If, given the large value of load resistance, a constant current approximation is made, and this constant current, I_{const} is assumed for the time being to provide only I_C then;

$$I_C = I_{const} = \frac{\partial Q}{\partial t} = C.\frac{\partial V}{\partial t} \tag{7.15}$$

and, if it is assumed that the discharge time is negligible, such that the time required

τ(sec)	Calculated $I_C(\mu A)$	Measured I (μA)
2.14	397	425
2.61	326	345
3.35	253	269
4.80	177	185

Table 7.1

The calculated charging currents and measured total currents through the experimental circuit (figure 7.11) for the four periods of oscillation with $C_P = 100\mu F$.

to raise the potential difference across the capacitor from V_H to V_S constitutes the total oscillation period, τ, then;

$$I_C . \int_0^{\tau} dt = C . \int_{V_H}^{V_S} dV \tag{7.16}$$

giving

$$I_C = C . \frac{(V_S - V_H)}{\tau} \tag{7.17}$$

In order to test this equation, the charging current, I_C has been calculated using the set of measured values of τ obtained for the 100μF capacitor in figure 7.13. It is found that the charging currents calculated using this equation are only slightly lower than the measured total currents, as shown in Table 7.1

Thus, in the unstable mode, almost all the measured current is accounted for by the displacement current through the external capacitor during the charging cycle. It must then be concluded that the MISS is in its low current 'off' state during the charging period since there is insufficient remaining current for it to be sustained in its NDR region. As such, the electrical behaviour of the MISS would appear to vary during each cycle. Although it might be expected to follow its NDR curve during the discharge phase, during which it also has the property of NDC, the device response departs from its stable NDR characteristic and again assumes positive values of resistance and capacitance during the charging period.

On the basis of this newly established result, it is possible to interpret the results of figure 7.10. From (7.17), and again neglecting the current through the MISS such that

$I_{TOT} = I_{const} = I_C$;

$$f.C = \frac{I_{const}}{(V_S - V_H)} \tag{7.18}$$

This expression is also plotted in figure 7.10 where $(V_S - V_H)$ is given the steady state value of 4.8 volts. It can be seen that (7.18) does in fact correspond quite closely with the measured results, particularly towards the low frequency limit. The small departure of the measured curves from this simple relation, which is greater the higher the frequency, is due to the various approximations made;

(i) The current taken by the MISS during the charging cycle, although small, will generally serve to increase the charging delay and thus reduce the slope of the curves.

(ii) The finite times taken for the MISS to switch and for the capacitor to discharge must increase the total oscillation period and will become more significant at higher frequencies.

(iii) As the MISS is in its 'off' state during the charging cycle, it contributes an additional capacitance, further increasing the period.

(iv) As the biasing circuit does not provide a true constant current source, the charging cycle will be exponential rather than linear (*i.e.* for $R_L \neq \infty$, $V(t) = V_{max}.[1 - \exp(-t/R_L.C_P)]$).

(v) The measured quiescent current (taken from the steady state MISS characteristic) does not, in general, correspond to the average current during the charging cycle.

An important consequence of this proposed behaviour in the oscillating state is that the property of NDC is lost during the greater part of each cycle. It is therefore not correct to analyse the frequency response of the MISS when it is oscillating by treating it as an NDC element in the generalised circuit of figure A1. Indeed, the Laplace analysis of Appendix A would not seem appropriate for modelling the oscillatory behaviour since it assumes that both the dynamic resistance and capacitance of the device remain constant.

In addition, it is now expected that when using the NDC measurement procedure outlined in section 7.3.1, the minimum parallel capacitance required to initiate an unstable response in a MISS biased to a stable point on its NDR curve will differ from the capacitance necessary to re-establish the stable state, as was found experimentally by Majlis[8] .

7.4.4 The Amplitude of Oscillation

Now that greater insight has been gained into the behaviour of MISS devices in the unstable state, further useful results may be achieved by simultaneously measuring the amplitude of current oscillations detected by the oscilloscope.

Results have been obtained from a second oxide-isolated device on the same mounted and bonded chip which has dimensions $30\mu m \times 180\mu m$. The steady-state characteristic of this device, measured using the same circuit, figure 7.11, but with a total load resistance of 76.33kΩ, is shown in figure 7.14. Again, the discontinuity in the I-V plot clearly indicates that the behaviour is unstable over a large part of the NDR curve, as indicated in the figure.

Now, for the purposes of measuring the peak-to-peak voltage swing in the oscillating region, it is noted that at any point in the NDR region the oscilloscope detects a voltage, $V(t)$ which is composed of a time-varying component $v(t)$ and a constant V. These are respectively given by $R_S.i(t)$, where $i(t)$ is the time-varying current through the circuit and $R_S.I$, where I is the quiescent current.

Given that oscillations occur along the load line, the full voltage swing across the MISS, V_0 is then;

$$V_0 = R_L.i_o = \frac{R_L}{R_S}.v_o \tag{7.19}$$

where i_o is the amplitude of the total current oscillation and v_o the amplitude of the oscillating voltage across R_S which is detected on the oscilloscope.

The unstable behaviour of this device has been examined under two sets of conditions. In the first, oscillations have been measured as a function of the quiescent current through the circuit with no positive capacitance connected in parallel with the MISS, and the results are shown in figure 7.15. These oscillations, of high frequency, are assumed to result from the presence of small stray capacitances in the measurement circuit. The frequency is seen to increase with current up to 5MHz in a highly non-linear fashion. This rise is accompanied by a concomitant fall in the amplitude of oscillation, which is at all points considerably lower than $|V_S - V_H|$. The frequency at which the amplitude falls to zero can be defined as the 'cut-off' frequency, f_C above which a device ceases to oscillate, or at least above which oscillations become undetectably small.

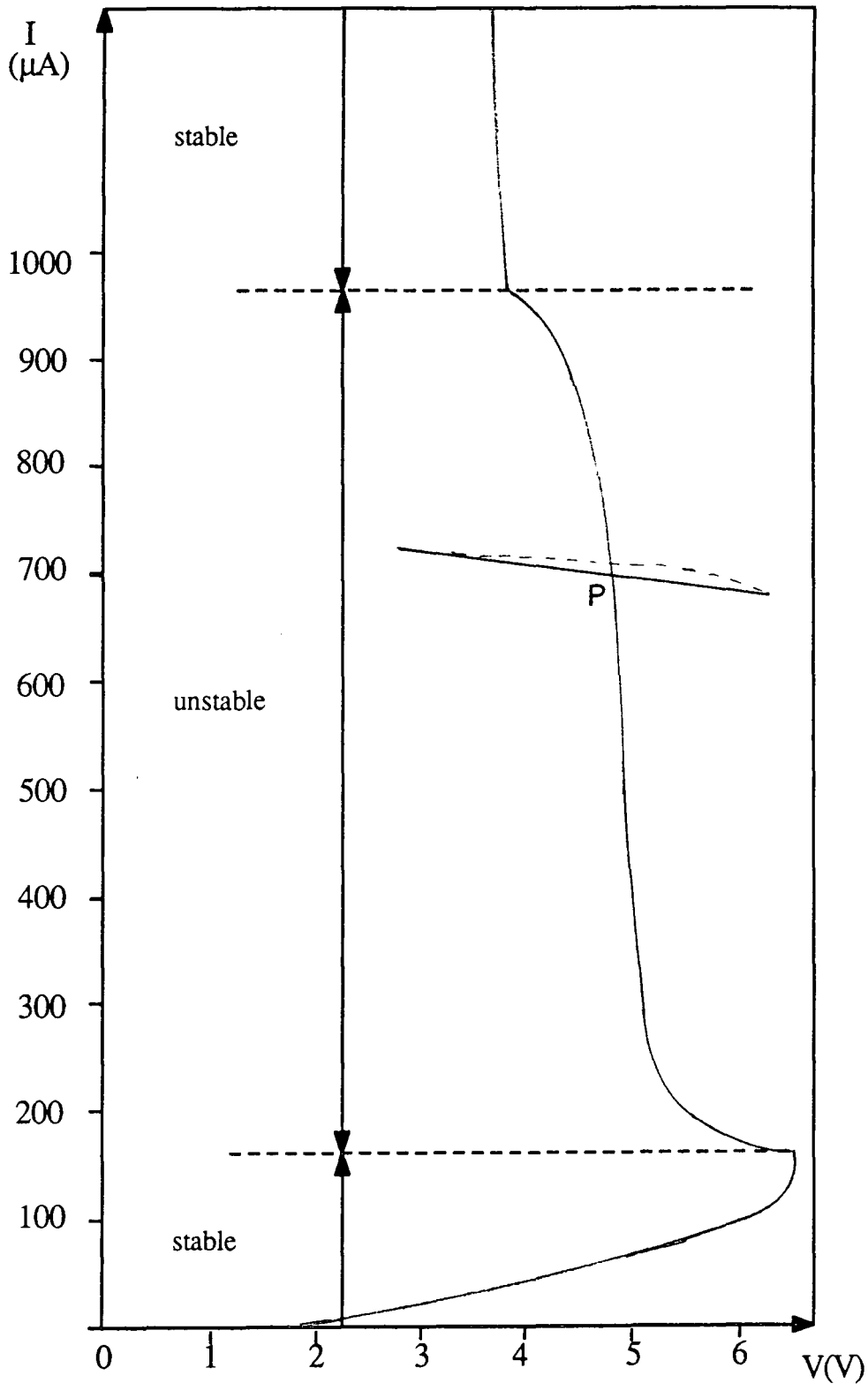


Figure 7.14; I-V characteristic of an SRO MISS device of dimensions 30μm x 80μm showing the locus of the oscillatory response at the biasing point, P.

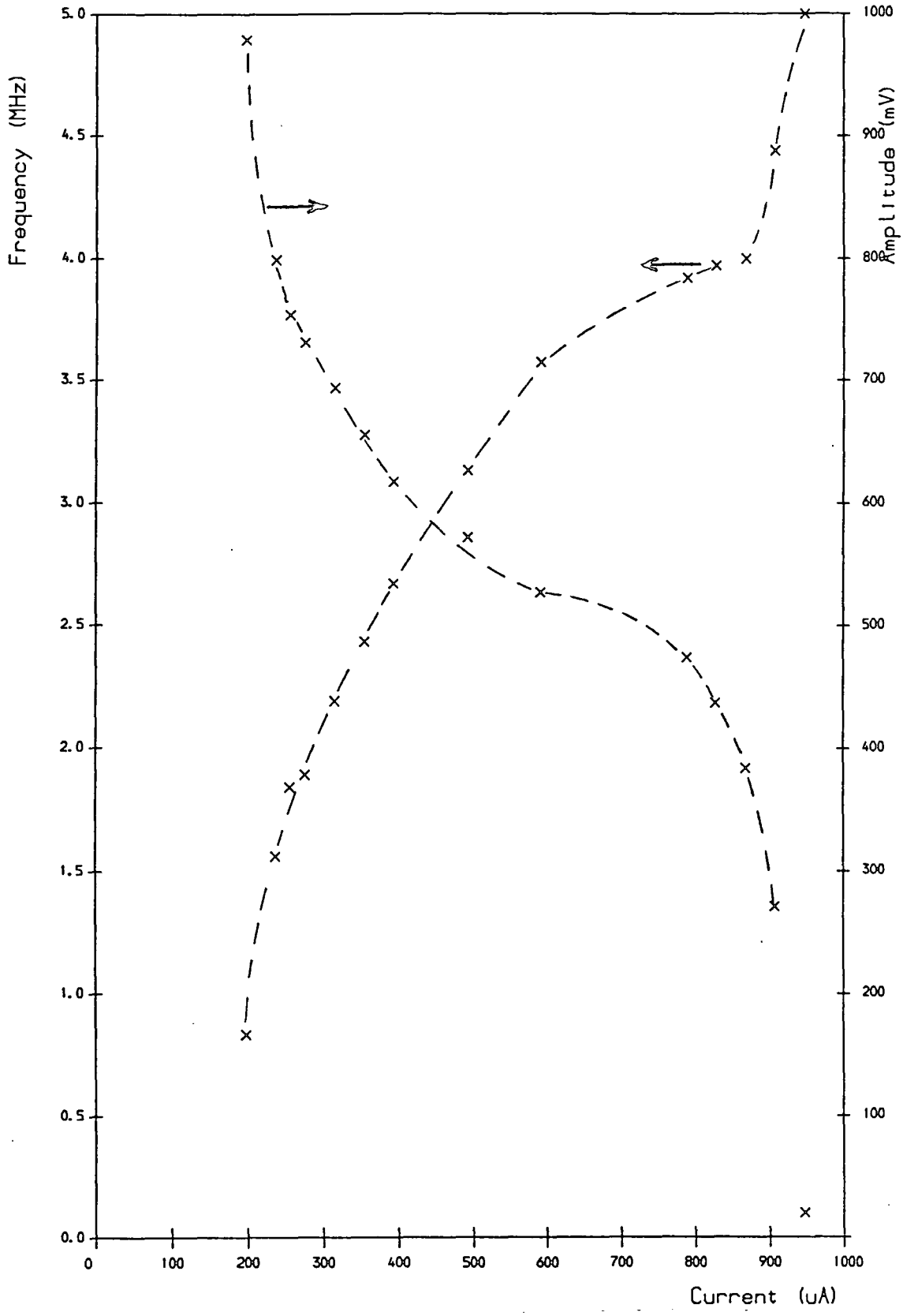


Figure 7.15; Frequency and amplitude of oscillation as functions of quiescent current.

In the second case, figure 7.16, oscillations have been measured as a function of additional capacitance introduced in parallel with the MISS for a constant, arbitrary value of quiescent current. These results effectively extend the range of measurements to lower frequencies by increasing the time constant with capacitive loading. Proportionality between the period and the external capacitance, as observed for the large dot devices, is confirmed over at least six decades of change. The corresponding variation in amplitude from a maximum near $|V_S - V_H|$ to zero is nonlinear and appears to be described by two characteristic decay factors.

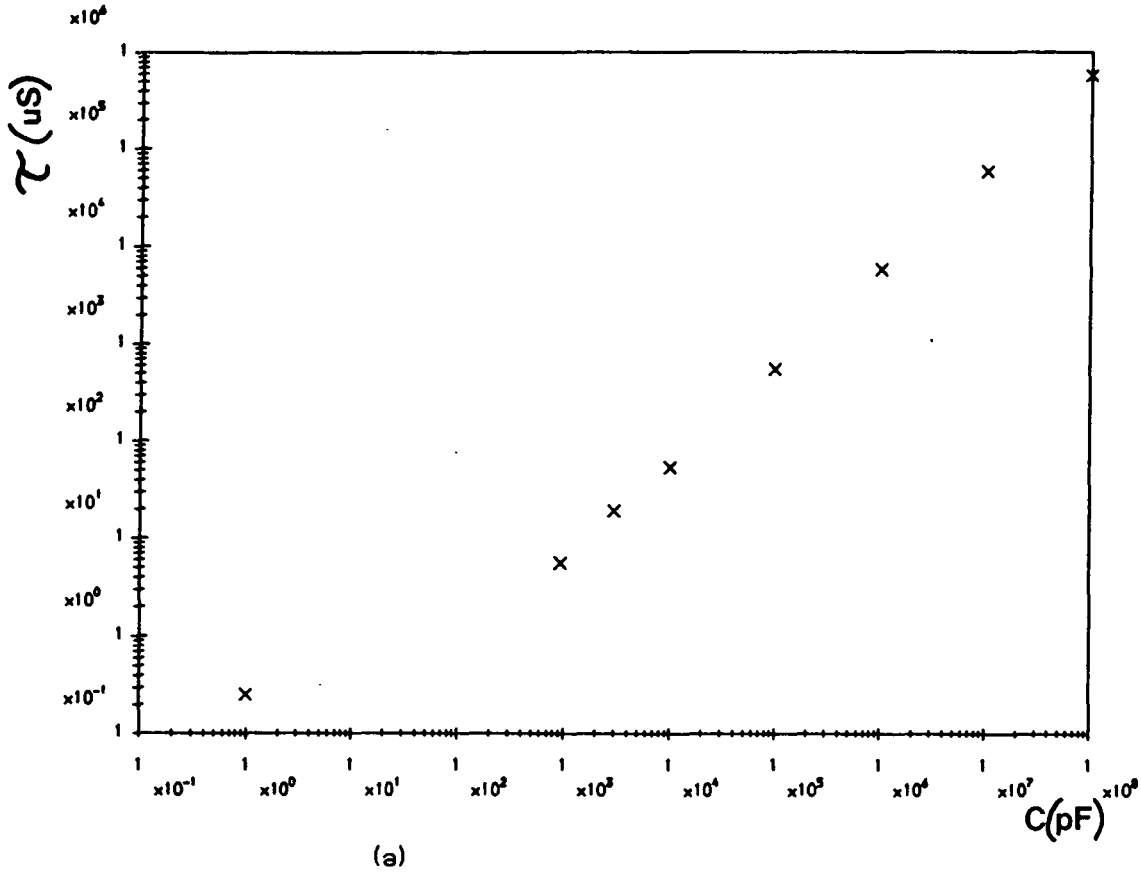
The results of figures 7.15 and 7.16 are combined in figure 7.17 where amplitude is plotted as a function of frequency and the dependences on capacitance and current are thus removed. This plot highlights the important point that the size of the voltage transition is primarily dependent on the delay between each transition.

A reduction in switching voltage for increasing frequency of operation of MISS devices has been reported by several authors[9,10] and investigated at some length. It has been shown[9] that switching under dynamic conditions, even at relatively low frequencies, occurs at a voltage lower than the steady-state value, $V_S(ss)$, due to the presence of residual minority charge at the semiconductor-insulator interface. This is known to allow the switching condition to be satisfied with a lower total voltage drop across the MISS. In effect, the steady-state 'off' condition, corresponding to a low density of minority charge at the I-S interface, is not fully established during the turn-off (discharge) phase before the turn-on (charge) cycle recommences. For the present oscillating case, it may be said that the charge state at the I-S interface is unable to respond fully at the frequency determined by the external circuit.

Adan and Zolomy[9] considered the stored charge to be composed of two parts; the inversion layer and diffusion charge and neglected minority carrier injection when deriving an approximate formula for the case of a punch-through mode device;

$$V_S(dynamic) = \frac{1}{2}.\frac{q.N_D}{\epsilon_s}\frac{\epsilon_{ins}^2.E_{th}^2}{\left\{q.N_D + q.K.exp(\frac{-t}{t_o})\right\}^2} \tag{7.20}$$

Here, t_o is the lifetime, E_{th} represents the field strength in the insulator at the switching point, K is the starting value for the minority carrier charge density which is assumed to decay exponentially and the other symbols have their usual meanings.



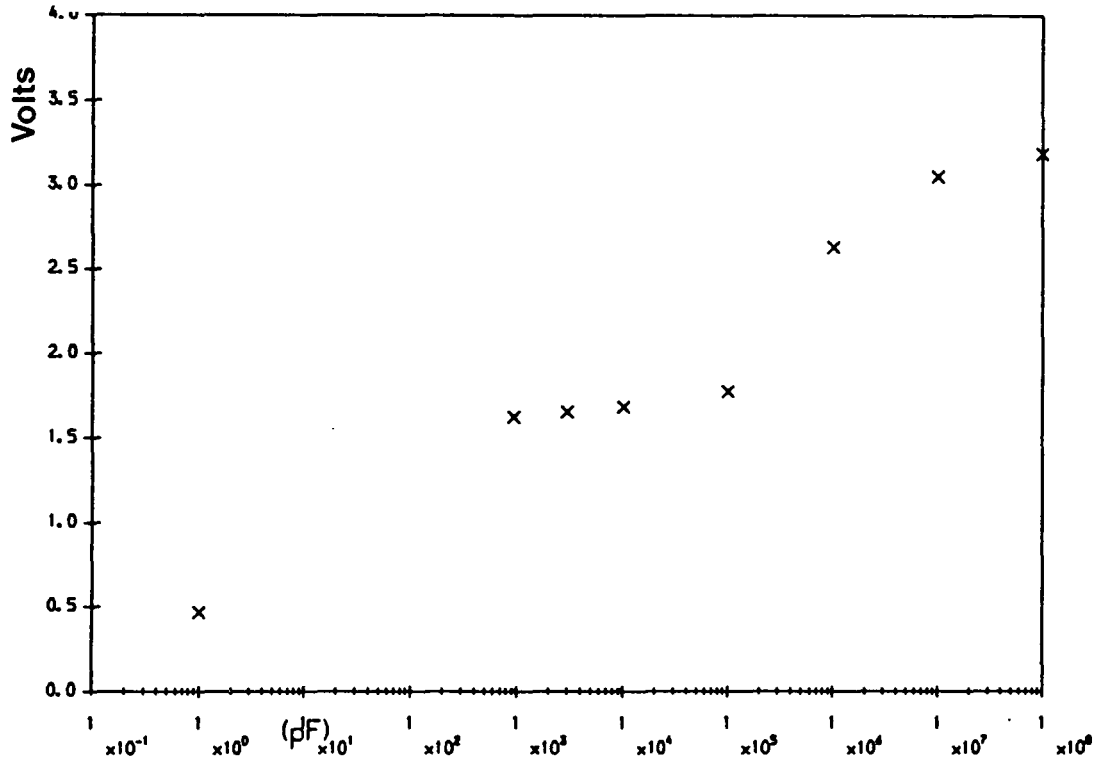


Figure 7.16; (a) Period and (b) amplitude of oscillation as functions of the capacitance connected in parallel with the MISS.

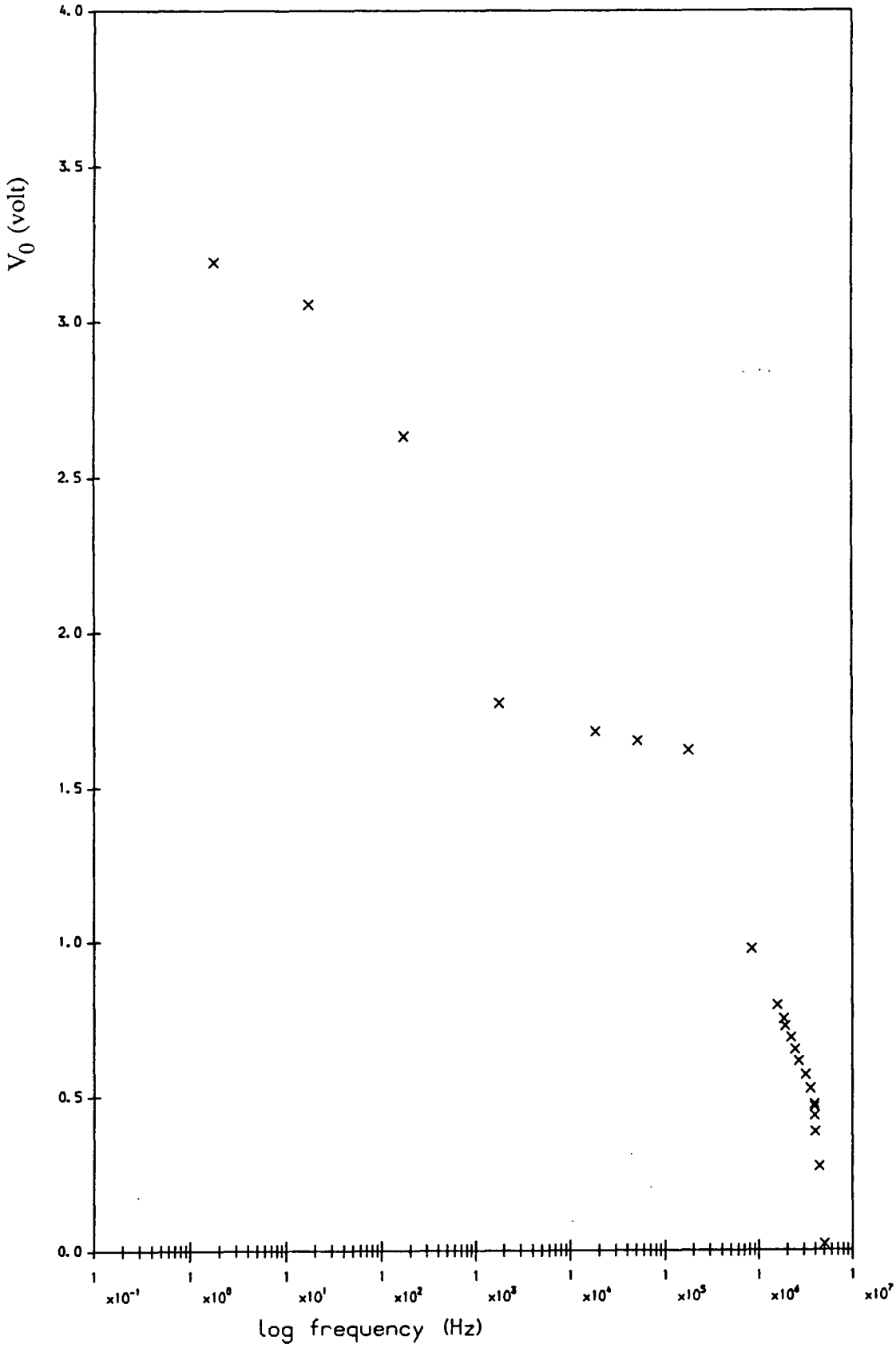


Figure 7.17; Amplitude of oscillation plotted against frequency, using the data of figures 7.15 and 7.16.

Phan et al.[10] developed this model to include the injected minority carrier current and derived an expression;

$$V_S(dynamic) = \frac{V_S(ss)}{\left\{1 + I_{inj}.\alpha.exp(-\frac{t}{t_o})\right\}^2} \tag{7.21}$$

Clearly for such an effect to account for a fall in amplitude at frequencies as low as 1Hz, some of the interface charge must have a characteristic lifetime of the order of one second. It must be assumed that such charge resides in deep surface state traps at or very close to the I-S interface. It has been shown in the characterisation of tunnel oxide and SRO MIS diodes in chapters 3 and 6 respectively that a high interface state density does indeed exist in both the MIS systems under study here. Free inversion charge contained in the semiconductor bulk will have a considerably shorter lifetime, of the order of 1μS and cannot therefore be expected to influence the fall in amplitude seen at low frequency.

The small plateau in the frequency range 2kHz to 200kHz, figure 7.17, may possibly arise because the surface state charge fails to respond in this range and ceases to be effective in further reducing V_S. If this is the case, a quasi-equilibrium density of interface state charge has been established.

The second fall in amplitude above 200kHz must therefore be due to another time-dependent mechanism in the MISS system with a characteristic lifetime of the order of $1\mu S$. As already mentioned, such a time is typical of semiconductor inversion layer decay in MIS systems and corresponds to the values of t_o determined experimentally by Zolomy and Phan et al (0.5 μS and 0.45 μS respectively). The depletion regions of both the reverse biased MIS and forward biased pn junctions can be expected to react on a much faster time scale than this as a change in potential across a space-charge region requires only the movement of majority carriers in bulk semiconductor.

Having established that the effective switching voltage is reduced when the device is operating in a dynamic mode, it is now proposed that for the present case where the voltage is oscillating under the influence of an external capacitance, it is not only V_S that is affected. There is reason to expect that the lower limit to the voltage excursions will in fact exceed the steady-state holding voltage, $V_H(ss)$. Considering that in the

unstable mode the quiescent current is insufficient to maintain the device in its 'on'-state, the only drive current available to send the operating point towards the holding point is that provided by the discharging capacitor. Referring to figure 7.13 then, the fall in voltage during the discharge phase is limited by the amount of charge stored on the parallel capacitor. That charge is determined by the magnitude both of the capacitance and of the voltage existing across its terminals at the switching point where the latter corresponds to $V_S(dynamic)$ at any given frequency of oscillation. Thus it may be postulated that as the frequency increases, the voltage swing, V_0, about the steady state quiescent point will decrease due to

(i) the falling magnitude of V_S(dynamic) and

(ii) the consequentially increasing magnitude of V_H(dynamic).

It is therefore proposed that the second and final fall in amplitude of oscillation results from the inability of minority charge in the inversion layer to respond to the rapidly changing potential across the MISS. In the limit, at $f = f_C$, no change in the inversion charge density, $p(0)$ is possible with time and a 'quasi-equilibrium' value is established. No variation in $p(0)$ must imply, following the analysis of chapter 4, no change in either the partition of potential across the device structure or in the component currents. As such, a quasi-equilibrium state is established in the device corresponding to the true steady state described by the current continuity equations and the charge neutrality equation.

7.5 Stable NDR and the Regenerative Feedback Mechanism

The theory of current-controlled regenerative feedback (CCRF) provides a powerful and persuasive means for analysing the switching behaviour of thyristors (SCRs), triangular barrier switches and MISS devices. The basic features of CCRF,as applied to the MISS device have been discussed in chapter 4 where it was shown that the response of the total current, I_{TOT} to an incremental change in generation current ΔI_G in the depletion layer is given by;

$$\Delta I_{TOT} = \frac{(1 + G'_{IBT}).(1 + G'_{BJT})}{1 - G'_{IBT}.G'_{BJT}}.\Delta I_G \tag{7.22}$$

where G'_{IBT}, the small-signal common-emitter current gain of the MIS, which was considered as an Inversion Base Transistor (IBT) and G'_{BJT}, the equivalent gain of the

epitaxial pnp bipolar junction transistor were respectively given as;

$$G'_{IBT} = \frac{\partial I_{NT}}{\partial I_{PT}} \qquad \text{and} \qquad G'_{BJT} = \frac{\partial I_{PJS}}{\partial I_{NJ}} \tag{7.23}$$

However, it is now suggested that this analysis is flawed in one important respect. Although this approach perports to be dynamic in that it considers small incremental changes in the component currents, it in fact neglects all real time dependences. In essence, such a simple description assumes that all changes in charge distribution occur much more rapidly than the rate at which the change in bias is being imposed.

Results reported in this chapter clearly show that time dependence cannot be neglected for the MISS system. In particular, the response of these devices at frequencies greater than about 10^5 Hz is probably limited by the inversion layer relaxation time. Thus it is concluded that although CCRF does provide a useful description of the switching process, it is restricted in its range of application. When the frequency of operation of a MISS exceeds f_C, change is suppressed by the inertia of the system and the loop gain is reduced to a value less than unity.

7.6 The Condition for Stable NDR

The existence of a cut-off frequency for current oscillations constitutes a stability condition which is based on rate-limiting physical processes internal to the MISS device. This condition is additional to that established for in-circuit stability which was based on the concept of NDC. It is interesting to consider which of these two quite independent criteria supercedes the other.

Considering a MISS device in an oscillating state, the NDC condition, $C = -|C|$, probably does not apply during the greater part of each cycle (since the device is not biased on its NDR curve while the parallel capacitance is charging) and the MISS itself will present a small additional positive capacitance. However, the NDC effect will still arise during each cycle for the short time taken for the device to switch on. Now, as the amount of parallel positive capacitance is reduced;

(a) the frequency of oscillation will increase;

$$f \propto \frac{1}{(C_P + C_{MISS})} \longrightarrow \infty \tag{7.24}$$

and (b) the total capacitance during the switching phase will tend towards the 'net negative capacitance' condition;

$$C_P - |NDC| \longrightarrow 0 \tag{7.25}$$

As such, the cut-off frequency may be exceeded before the net capacitance becomes negative, a situation which is represented in figure 7.18 where the two zones of stability are indicated. An unstable device may thus attain internal stability and cease oscillating without necessarily satisfying the 'in-circuit' stability condition arising from negative capacitance.

What is more, the observation of stability in a thyristor reported by Wei[2] may also now be explained. The same frequency cut-off effect that occurs in the MISS would be expected to apply to the thyristor and indeed to all other forms of regenerative switching device, such as the bulk barrier switch.[11] Thus it is proposed that all regenerative feedback devices should exhibit stable NDR under the correct circuit conditions without the requirement of negative differential capacitance.

On reflection, that class of devices which exhibit NDR as a result of a physical change in a material property, such as $\alpha-$Si[12] and chalcogenide memory devices[13,14], may also be expected to satisfy the internal stability condition and therefore exhibit a stable NDR curve without the requirement of negative differential capacitance.

7.7 Conclusions

The first observation of stable negative differential resistance in a MISS device has been reported. Although this experimental result appears to conflict with both circuit theory and the theory of regenerative switching, a successful resolution of both points has been achieved.

On the basis of simple Laplace analysis, it has been shown that if a large value of load resistance is used, then in-circuit stability of an NDR element becomes possible if that element has the property of negative differential capacitance. Furthermore, it has been shown that MISS devices would theoretically be expected to exhibit negative capacitance in the NDR region of their switching characteristic. However, dynamic I-V measurements appear to indicate that the property of NDC is lost when the device is

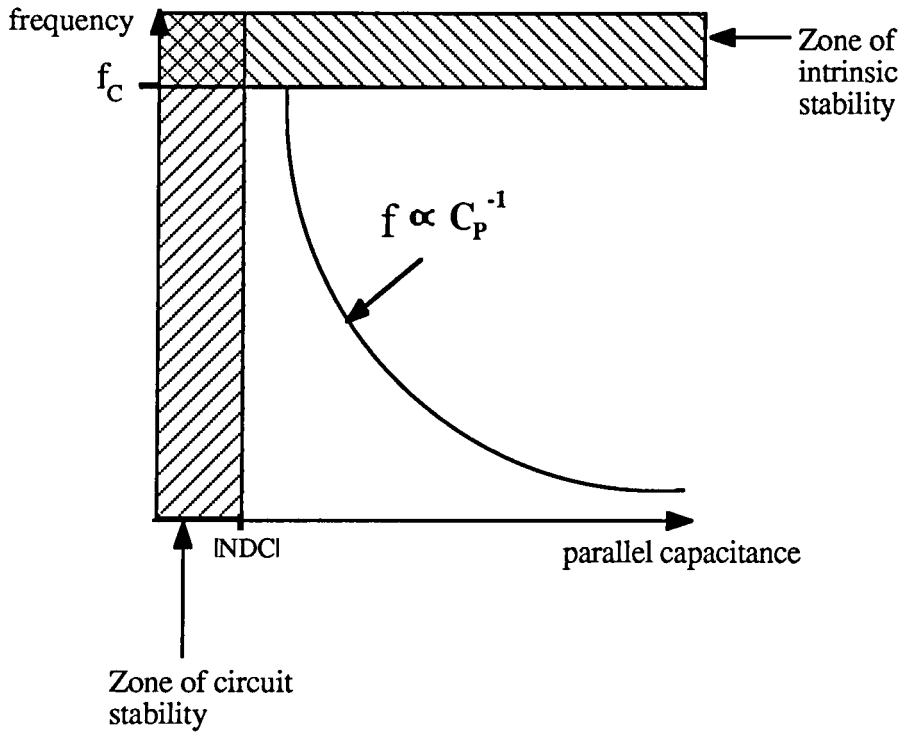


Figure 7.18 ; The zones of stability of a negative differential resistance device with negative differential capacitance.

oscillating since its quiescent point departs from the NDR curve for the greater part of each cycle.

The question of how a regenerative switching device can be held in quiescence between its stable 'on' and 'off' states has been answered on the basis of the frequency limitations inherent in the device. It has been shown that the frequency at which a MISS device oscillates in its NDR region is determined principally by the $R.C$ time constant of the circuitry around it. Any electronic device will take a minimum length of time to respond to a change in bias. In the case of a MISS, the minimum time is probably limited by the inversion layer discharging delay. Therefore, if the parasitic capacitance in parallel with the MISS is reduced sufficiently, the oscillation frequency will exceed the maximum (cut-off) frequency of the device.

In terms of the regenerative feedback theory proposed in chapter 4, this stability condition may be considered to be equivalent to the small-signal gain of the current feedback loop falling below unity;

$$G'_{IBT}.G'_{BJT}\rceil_{f=f_c} < 1 \tag{7.26}$$

It is concluded finally that this condition will override that pertaining to circuit stability and furthermore that the notion of a cut-off frequency probably accounts for the stability observed in other types of current-controlled negative resistance device.

REFERENCES

1. M.E.Hines, Bell Sys. Tech. J. 477 (1960)
2. D.T.Y.Wei, Solid State Electronics **23** 509 (1980)
3. H.Fritzsche, IBM J. Res. Develop. 515 (1969)
4. H.Kroger and H.A.Wegener, Solid State Electronics **21** 643 (1978)
5. S.E-D.Habib and J.G.Simmons, Solid State Electronics **23** (1980)
6. M.J.Morant, private communication
7. J.Millän, F.Serra-Mestres and J.Buxö, Rev. Phys. App. **14** 921 (1979)
8. B.Y.Majlis, PhD thesis, University of Durham (1988)
9. A.Adan and I.Zolomy, phys. stat. sol. (a) **57** 113 (1980)
10. H.K.Phan, P.H.Binh and L.H.Phu, phys. stat. sol. (a) **81** K1 (1984)
11. K.Board, K.Singer and R.Malik, Electronics Letters **18** 676 (1982)
12. A.E.Owen, P.G.LeComber, G.Sarrabayrouse and W.E.Spear, IEE Proc. I; Sol. State Elec. Dev. **129** 51 (1982)
13. S.R.Ovshinsky, Phys. Rev. Letters **21** 1450 (1968)
14. A.D.Pearson, IBM J. Res. Develop. 510 (1969)

CHAPTER EIGHT

Conclusions

8.1 Negative Differential Capacitance

The concept of NDC is obviously novel and it is a matter for conjecture what practical applications it might be put to. In an analogue circuit, the 90° phase shift introduced by an element with NDC will be in the same sense as that due to an inductor. However, a negative capacitance is quite distinct from an inductance in that its impedance bears an inverse rather than a direct proportionality to the signal frequency. An NDC element, if implementable, would therefore appear to present a whole new set of possibilities in analogue circuit design.

However, several points have arisen during the course of this work which suggest that utilisation of the NDC is perhaps a forlorn hope. A point of major concern is that in the unstable, oscillating mode, the operating point of the MISS departs from its NDR curve. As such, the property of NDC is lost when the device is not operated in the steady state. To hold the device at a quiescent point on its NDR curve, a shallow load line is necessary, as discussed in chapter 7. Now, if a signal were to be applied to the MISS (through a high-pass capacitive filter for instance), the I-V locus of that signal would be along the load line, not along the NDR curve. Thus, an applied signal would also be expected to force the operating point of the MISS off its NDR curve with consequential loss of the NDC. In practical terms, the property of NDC, although conceptually well-founded and theoretically useful for explaining the stability of the NDR, may prove to be rather elusive. The application of a test signal is expected to result in the destruction of the desired effect. As a consequence, it would seem likely that the only means of measuring the NDC is by the indirect method described in section 7.3.1. Certainly, all attempts by the author to measure NDC using capacitance meters have been unsuccessful.

A more general factor to consider, whatever the intended application of the MISS is the limited frequency range over which the device is able to respond. The highest oscillation frequency observed during the course of this work was only 5MHz. As such, even if the NDC effect was usable, the range of applications would be very restricted.

8.2 Applications of the MISS in Microelectronics

In the introduction to this thesis, it was pointed out that most of the suggested applications of the MISS in integrated circuits are now redundant. As a memory element, or as part of an oscillator circuit, the MISS presents no advantages over mature and well-established MOS technologies. Following the discussion of chapter 5, a major problem would appear to be reduction of the holding current which does not scale with the device dimensions. It is difficult then, given the recent proliferation of MOS technology and continued advances in design rule reduction to below a micron, to see many remaining applications for the MISS.

However, returning to the concept of the MISS as a switching device rather than as a logic element, one area where the device might still have a role is in 'smart power' applications. MOS circuitry is fast and power efficient when not under any load. However, at the output of an I.C, buffer circuitry is required to drive even the capacitive load presented by a bond pad if intolerable loss of speed is to be avoided. Here, the ability of the MISS to discharge or charge a reactive load rapidly may be useful. Indeed, an inverting buffer circuit containing two MISS devices, as shown in figure 8.1, was suggested and patented by Sarrabayrouse et. al.[1] in 1982. This configuration makes use of the capability of the MISS to pass a large current over a short period and subsequently remain in a high impedance state between switching transients. As such, the large change in output impedance of the MISS as it switches between its on and off states is put to use rather than its bistability. The authors reported that a hybrid, non-optimised inverter was capable of discharging a 100pF load in 50nS with an on:off state current dissipation ratio of $\sim 10^4$.

In an application such as this, the MISS is only used in the output stages and therefore in a low-density role to which it is probably better suited.

8.3 Limitations of the Epitaxial MISS

Whatever the possible applications of the MISS in I.C.s, there remain two other practical considerations which prejudice its chances of being adopted;

(a) The epitaxial form of the MISS is not amenable to integration, nor is it compatible with conventional MOS technology. All devices on the same substrate share a common

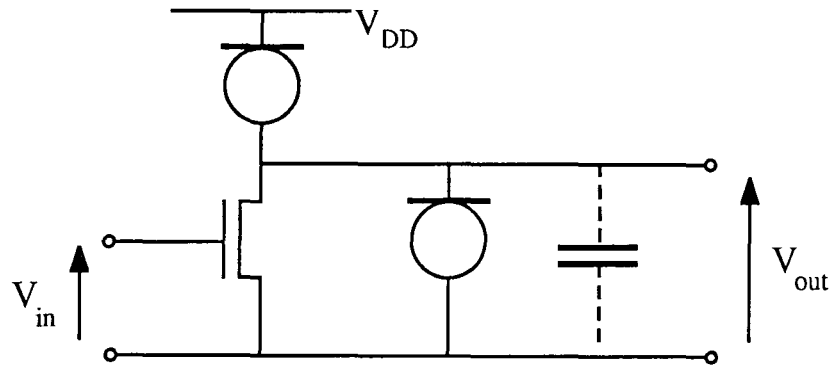


Figure 8.1; An inverter containing two MISS devices after Sarrabayrouse et. al. [1]

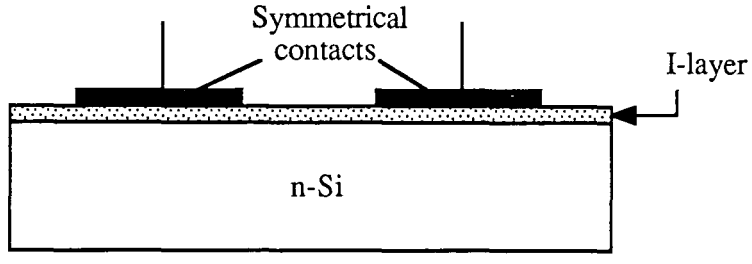


Figure 8.2; Schematic diagram of a MISIM lateral switching device after Darwish and Board [2]

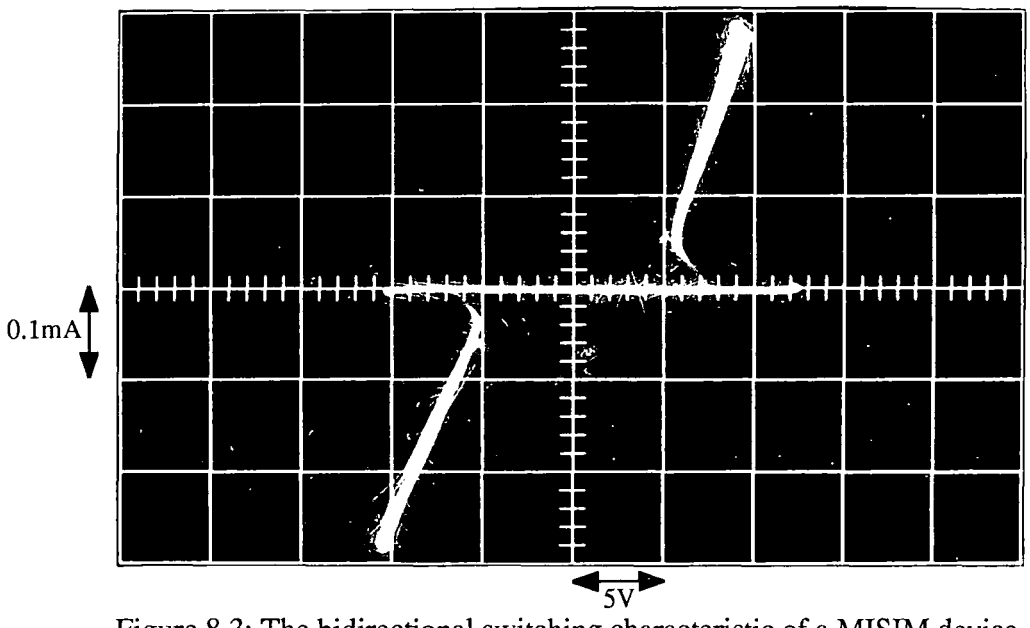


Figure 8.3; The bidirectional switching characteristic of a MISIM device fabricated with a tunnel oxide I-layer. Electrodes 30x80μm, separation 20μm.

terminal which clearly imposes a severe restriction on the circuit design and prohibits most applications such as the buffer of figure 8.1.

(b) The reproducibility of the intrinsic switching voltage of a two terminal MISS is likely to be poor.

The first problem may be overcome by the adoption of a lateral configuration for the MISS, using a diffused or implanted, rather than an epitaxial pn junction. This offers several benefits;

(i) High injection efficiency of minority carriers (holes) at the n-p^+ junction without the need for isolation. The p^+ implanted region has its dimensions defined photolithographically and may thus be small. In addition, the abruptness and quality of the junction should be markedly improved.

(ii) The processing required is more compatible with standard MOS technology; photolith / etching / implantation etc.

(iii) All contacts to the MISS are on the top surface and independent of each other.

The second problem of control over the switching voltage may be overcome by use of a third, or even a fourth control (or gate) terminal. Then, as in most thyristor applications, the tolerance on the intrinsic switching voltage, V_{S0}, is relaxed. It is only necessary that V_{S0} be greater than the maximum voltage the device is required to block. Switching is initiated, not by exceeding the V_{S0} but by applying a base current to either the IBT or BJT in the MISS and thus reducing the switching voltage.

8.4 Lateral Forms of the MISS Device

8.4.1 MISIM

A very simple variant of the MISS which has a lateral configuration is the metal-insulator-semiconductor- metal structure, proposed and demonstrated by Board et. al.[2,3] In this device, the forward biased pn junction is replaced by a forward biased MIS junction, as indicated in figure 8.2. As such, the structure is symmetrical and switching is possible whatever the polarity of the applied potential, a property which might be useful in some applications.

A number of such devices have been fabricated in the course of the present study using tunnel oxide for the I-layers. However, they were characterised by a low yield, high

holding voltage and poor reproducibility. Indeed, it has not been possible to obtain a truly symmetrical switching characteristic from any of the hundred or so devices tested. The characteristic shown in figure 8.3, which was obtained from a device with $80\mu m \times 30\mu m$ MIS contacts separated by $20\mu m$, is one of the better examples. These problems may be attributed to the poor reproducibility of the electrical properties of tunnel oxide layers, as discussed in chapter 3 and to the poor injection efficiency of a MIS junction as compared to a p^+-n junction. Thus, despite the obvious attraction of its inherent simplicity, the MISIM is not expected to be a viable component for microelectronic applications.

8.4.2 The Three-Terminal Lateral M-I-n-p^+ Switch

Based on the arguments of the previous sections, it may be concluded that the form of MISS best suited to I.C. applications should be the lateral M-I-n-p^+ structure with contacts either to the base of the IBT, figure 8.4(a) or to the base of the BJT, figure 8.4(b).

Despite some effort, it has not proved possible to fabricate the first of these structures, mainly due to the limitations of the facilities available at Durham. However, the second device has been successfully produced at the University of Southampton microelectronics centre as part of an SERC collaborative project. An extensive mask set was created by the author on the SERC 'PR1ME' I.C. design system for the multiple purposes of defining isolated epitaxial and lateral MISS device structures. Two batches of wafers were produced using fabrication processes devised by the author which were as close to a standard CMOS process as possible. However, difficulties arose in the course of both fabrication runs due mainly to the minor modifications required to the normal process sequence. Despite these problems, a number of devices from the first batch were found to function albeit with very poor reproducibility and yield. Figure 8.5(a) shows the switching characteristics of a three-terminal lateral MISS with the geometry indicated in figure 8.5(b). In this particular device, p^+ and n^+ implanted stripes were aligned on either side of the $100\mu m$ square MIS contact with a separation of $10\mu m$. The low-doped epitaxial layer was grown on an n^+ substrate in the hope of achieving a degree of confinement of the minority holes in the upper layer. It can be seen from the switching characteristics that the effect of a positive base current (negative bias on the n^+ contact) is to considerably reduce the switching voltage below its intrinsic value ($I_b = 0$), as expected. Indeed, a current of only $250\mu A$ is required to reduce the switch-

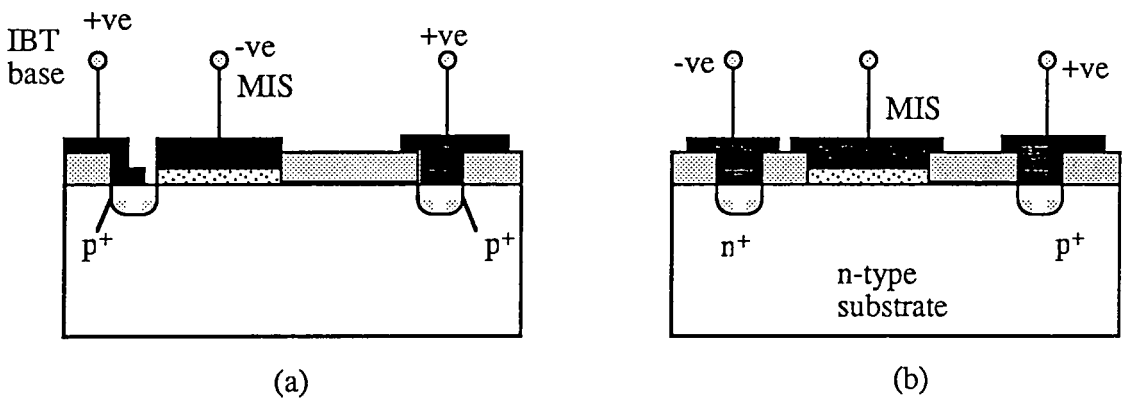


Figure 8.4; Proposed forms of three-terminal lateral MInp (or MIpn) switching devices;
(a) third (self-aligned) contact to the base of the IBT
(b) third contact to the base of the BJT.

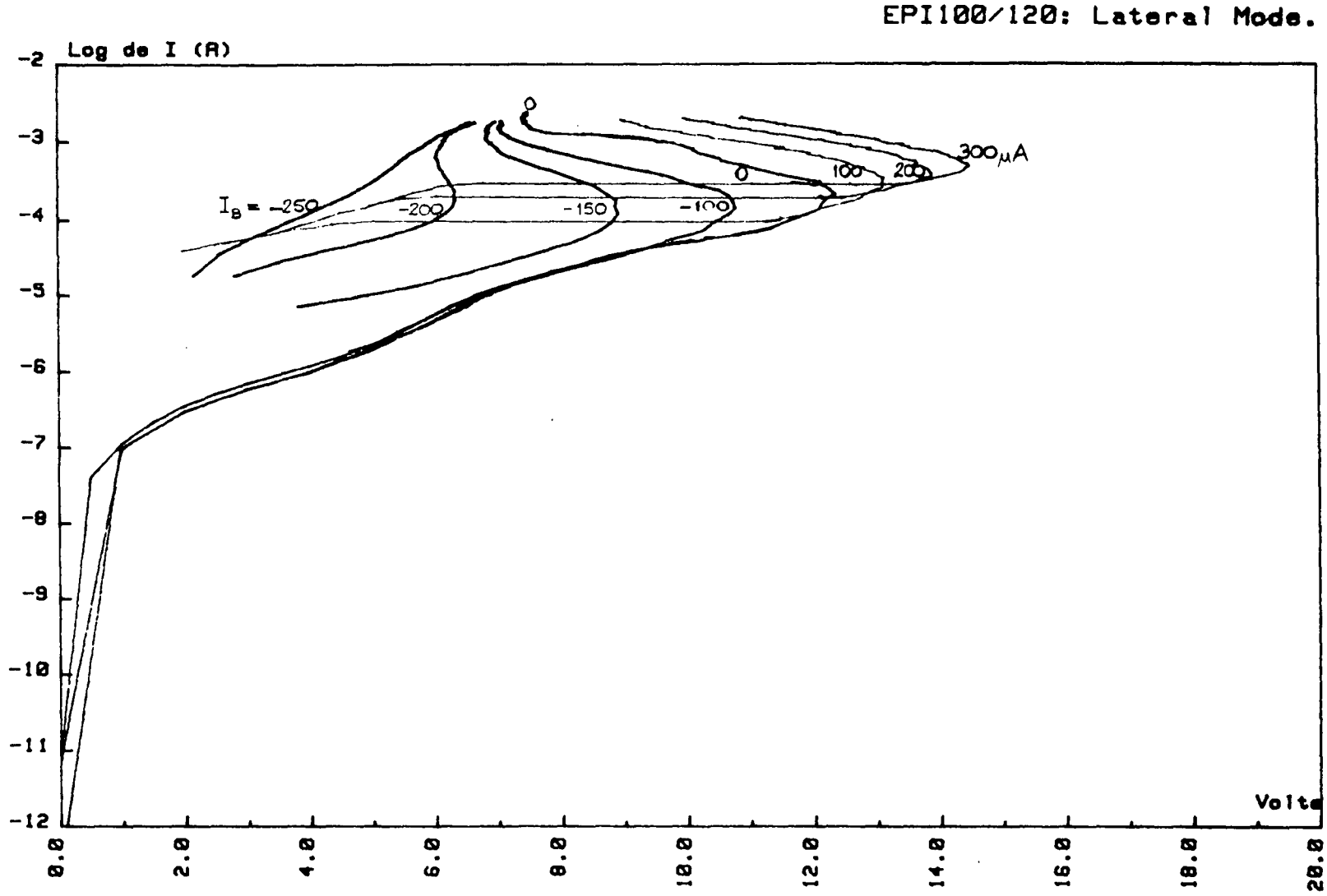
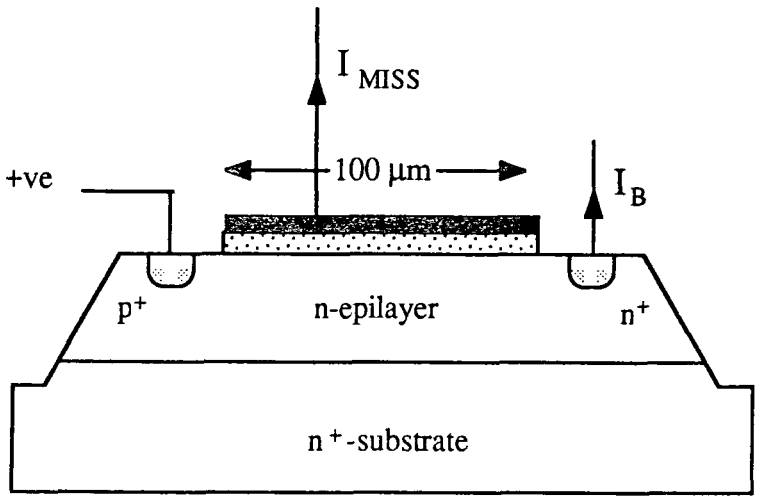
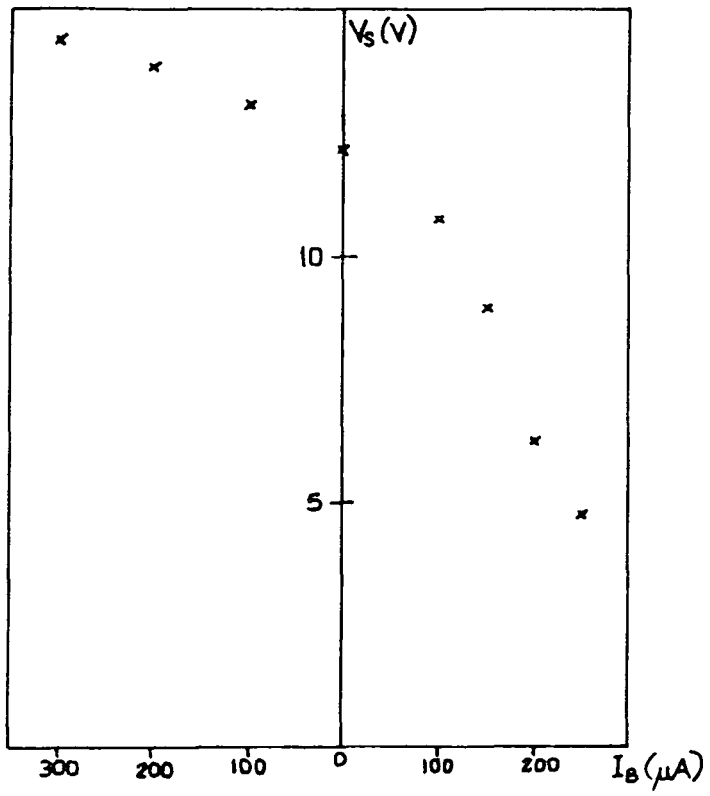


Figure 8.5 (a); Switching characteristics of a lateral 3-terminal MISS device obtained for several values of base current, I_B in the range -250μA to 300μA, as indicated.



(b)



(c)

Figure 8.5; (b) Schematic cross-section diagram of the switching device.
(c) Plot of the control curve, V_S against I_B obtained from (a).

ing voltage below the holding voltage. Moreover, the opposite effect is achieved if the base current is reversed but with a rather lower efficiency of control. The effect of the base current is summarised in figure 8.5(c), from which a maximum control efficiency, $\partial V_S/\partial I_B$ of 56.2 kΩ may be extracted.

8.5 Future Trends in MISS Switches

In the final analysis, it has to be concluded that the MISS is not a very likely candidate for future VLSI circuits. Any realistic application will be where there is a need for a robust and easily manufactured switching device, working in the range of about 5 to 60 Volts. The MISS might be either a discrete component or form part of the output stage of an I.C. The semi-insulator used would probably be SRO which is easily and cheaply deposited and does not have to withstand such high fields as tunnel oxides.

Two particular fields in which these criteria are met are automotive electronics (12V standard line voltage) and telephony (50V standard in Britain). In the former, much greater use of electrical instrumentation is foreseen in the near future and suitable devices will be required to form the interface between the low-power MOS control electronics and the high power devices (lamps, motors displays etc.) being controlled.

However, some improvement would have to be be achieved in the holding voltage of the MISS (whereas in a switching application, the holding current is of less importance). In view of the current densities that the MISS would have to pass in such applications, it is clearly intolerable to have more than about one volt dropped across the device in its on-state. Otherwise, the power dissipated internally would be excessive and probably lead to failure.

REFERENCES

1. G.Sarrabayrouse, A.Essaid and J.Buxo, Revue de Physique Appl. **17** 681 (1982)
2. M.Darwish and K.Board, IEE Proc. I, *Solid State and Electron Devices* **128** 161 (1981)
3. M.Darwish and K.Board, IEE Proc. I, *Solid State and Electron Devices* **128** 165 (1981)

APPENDIX A

Laplace Analysis of NDR in a Circuit

In order to analyse the response of a circuit containing a NDR element, it is most practical to treat a notional circuit given in figure A.1 which contains a minimum of lumped elements. Here, L is representative of the inductance of the connecting leads and R_L is the total series load in the current loop. C consists of the parallel combination of the effective capacitance of the MISS (which may be negative), the lumped effect of all the distributed capacitance in the circuit wiring and other circuit components plus any extra capacitance that may be connected in parallel with the MISS. R is the negative differential resistance component of the MISS impedance, $\delta V/\delta I$.

From this circuit, the small change in the device potential, $\Delta V(t)$ that will arise from a small change in the supply voltage, $\Delta V_S(t)$ is given by;

$$\frac{\Delta V(t)}{\Delta V_S(t)} = \frac{\left\{\frac{1}{R} + \frac{1}{1/j\omega C}\right\}^{-1}}{R_L + j\omega L + \left\{\frac{1}{R} + j\omega C\right\}^{-1}} = \frac{R}{R + (1 + j\omega CR).(R_L + j\omega L)} \tag{A1}$$

Now, if the perturbation of the supply potential is assumed to take the form $\Delta V_S.\delta(t)$, where $\delta(t)$ is a delta function, the Laplace transfer of this function becomes;

$$\Delta V_S(s) = \mathcal{L}[\Delta V_S.\delta(t)] = \Delta V_S.\mathcal{L}[\delta(t)] = \Delta V_S.1 \tag{A2}$$

The response of the system to such a perturbation is therefore given by (A1) and can be put into the complex frequency domain by the transfer function;

$$H(s) = \frac{\Delta V(s)}{\Delta V_S(s)} = \mathcal{L}\left(\frac{\Delta V(t)}{\Delta V_S(t)}\right) = \frac{R}{R + (1 + sCR).(R_L + sL)} \tag{A3}$$

which can be simplified to;

$$H(s) = \frac{1}{LC}.\frac{1}{(s + \alpha + \sqrt{\beta}/2).(s + \alpha - \sqrt{\beta}/2)} \tag{A4}$$

where α and β are defined as;

$$\alpha = \frac{L + RR_LC}{2LCR} \qquad \text{and} \qquad \beta = \left(\frac{L + RR_LC}{LCR}\right)^2 - \frac{4(R + R_L)}{LCR} \tag{A5}$$

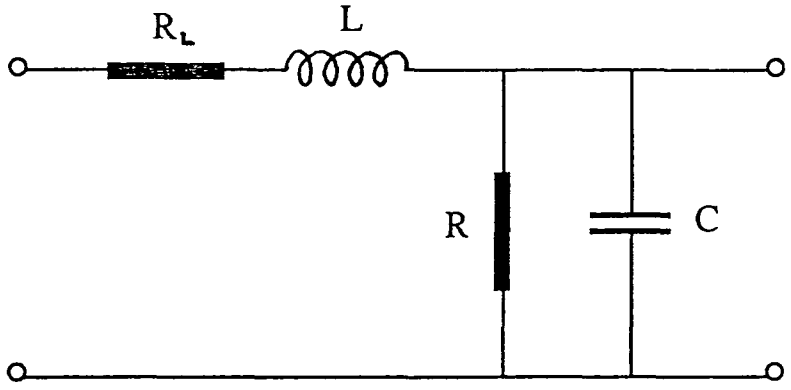


Figure A1; The 'lumped' element equivalent circuit considered in the Laplace analysis.

The inverse transform of this expression then gives the impulse response of the system in the time domain;

$$h(t) = \mathcal{L}^{-1}[H(s)] = \frac{1}{LC\sqrt{\beta}}.\left\{\exp-\left(\alpha+\frac{\sqrt{\beta}}{2}\right).t - \exp-\left(\alpha-\frac{\sqrt{\beta}}{2}\right).t\right\} \tag{A6}$$

or alternatively;

$$h(t) = \frac{-2}{LC\sqrt{\beta}}.\exp(-\alpha t).\sinh\left(\frac{\sqrt{\beta}}{2}.t\right) \tag{A7}$$

There are two possible ranges of solution for this response function, depending on whether β takes (I) a negative or (II) a positive value.

Case (I); β negative.

For $\beta < 0$, $\sqrt{\beta}$ may be considered as $j.\sqrt{\beta}$ where β now represents the magnitude of this parameter. Using the trigonometric relation $\sinh j.(A.x) = j.\sin(A.x)$, equation (A7) may be rewritten as;

$$h(t) = \frac{2}{LC\sqrt{\beta}}.\exp(-\alpha t).\sin\left(\frac{\sqrt{\beta}}{2}.t\right) \tag{A8}$$

Clearly the response in this case is a sinusoidal oscillation whose amplitude either decays (for $\alpha > 0$) or grows (for $\alpha < 0$) exponentially with time, or remains constant in the specific case $\alpha = 0$. Thus, only the condition $\alpha > 0$ provides a stable response of the circuit if $\beta < 0$.

Examining the parameter α from (A5) where $R = -|R|$ represents an NDR element;

$$\alpha = -\frac{1}{2C|R|} + \frac{R_L}{2L} \tag{A9}$$

the requirement for a stable response is simply;

$$R_L > \frac{L}{|R|.C} \tag{A10}$$

If, however, the capacitance, C also takes a negative value, $-|C|$, this condition is satisfied independently of the size of the load;

$$R < 0\ ,\ C < 0 \Rightarrow \alpha > 0 \qquad \forall R_L \tag{A11}$$

Clearly then, for the case $\beta < 0$, the response of a circuit containing a negative resistance element will be stable if the net capacitance in parallel with that element is also negative. It is now necessary to consider the circuit parameters required to ensure $\beta < 0$. The expression for β in (A5) may be rearranged, after substitution of $R = -|R|$, as;

$$\begin{aligned}\beta &= \frac{L^2 + (|R|.R_L.C)^2 + 2.|R|.R_L.C.L - 4.|R|^2.C.L}{(L.C.|R|)^2} \\ &= \frac{(L + |R|R_L C)^2 - 4|R|^2 LC}{(LC|R|)^2}\end{aligned} \tag{A12}$$

Now, assuming the capacitance, C is positive, the condition for $\beta < 0$ may be stated;

$$R_L < 2.\sqrt{\frac{L}{C}} - \frac{L}{|R|.C} \tag{A13}$$

Taking typical values for the parameters L, C and $|R|$ of $10\mu H$, $100pF$ and $10k\Omega$ respectively, the value of load resistance required to maintain stability is $R_L < 600\Omega$. Such a value clearly contravenes the condition for a unique quiescent point in the NDR region of an 'S-type' characteristic, $R_L > |R|$. Thus, $\beta < 0$ is not satisfied for realistic values of the circuit elements if the parallel capacitance is positive.

However, for negative values of small signal capacitance, $C = -|C|$, (A12) becomes;

$$\beta = \frac{L^2 + (|R|.R_L.|C|)^2 - 2.|R|.R_L.|C|.L + 4.|R|^2.|C|.L}{(L.|C|.|R|)^2} \tag{A14}$$

As such, $\beta < 0$ requires;

$$2.|R|.R_L.|C|.L > L^2 + (|R|.R_L.|C|)^2 + 4.|R|^2.|C|.L \tag{A15}$$

Using the previous values for L, C and R and evaluating the terms in (A15), it can be shown that the left-hand side is always very much less than the right-hand side whatever the magnitude of the load.

In conclusion then, Case (I) can not arise if the capacitance is negative, nor is it likely to arise if the capacitance is positive.

Case (II); β positive.

Returning to the impulse response equation (A6), it can be seen that the fluctuation in the device voltage $\delta V(t)$ due to a perturbation $\delta V_S(t)$ will only decay if both the exponents $(\alpha + \sqrt{\beta}/2)$ and $(\alpha - \sqrt{\beta}/2)$ are positive.

(a) Assuming C is positive.

Since in Case (II) β is by definition always positive, the first term $(\alpha + \sqrt{\beta}/2)$ will be positive if $\alpha > 0$. From (A9), this condition corresponds to;

$$R_L > \frac{L}{|R|.C} \tag{A16}$$

In a practical circuit, this inequality holds unless the inductance is very large. For instance, taking a load of $10k\Omega$ and $|R| = 10k\Omega$, $|C| = 100$ pF, (A16) is only untrue for $L > 0.01$H.

From (A9), the second term, $(\alpha - \sqrt{\beta}/2)$ will be positive if;

$$\alpha = -\frac{1}{2C.|R|} + \frac{R_L}{2L} > \frac{\sqrt{\beta}}{2} \tag{A17}$$

or, restated;

$$\beta < \left(\frac{|R|.R_L.C - L}{LC|R|}\right)^2 \tag{A18}$$

Substituting for β from (A12), this inequality may be reduced to;

$$R_L < |R| \tag{A19}$$

which forbids a unique operating point on the NDR curve. As a consequence, it is not possible to observe stable NDR in a current controlled, 'S-type' I-V characteristic if the parallel capacitance in the measuring circuit is positive.

(b) Assuming C is negative.

Substituting $C = -|C|$ in expressions (A9) and (A12), α and β become;

$$\alpha = \frac{1}{2|C|.|R|} + \frac{R_L}{2L} \quad \text{and} \quad \beta = \frac{(L - |R|.R_L.|C|)^2 + 4|R|^2.|C|.L}{(L.|C|.|R|)^2} \tag{A20}$$

In this case, both α and β and hence the term $(\alpha + \sqrt{\beta}/2)$ in (A6) are always positive. Moreover, the requirement for the term $(\alpha - \sqrt{\beta}/2)$ to be positive is $\beta > 4.\alpha^2$ or, substituting for α and β;

$$\frac{(L - |R|.R_L.|C|)^2 + 4|R|^2.|C|.L}{(L.|C|.|R|)^2} > 4.\left(\frac{1}{2|C|.|R|} + \frac{R_L}{2L}\right)^2 \tag{A21}$$

which simplifies to;

$$R_L > |R| \tag{A22}$$

Thus, it may be concluded that stable NDR may be measured with $R_L > |R|$ if the differential capacitance in the measuring circuit is negative.

To summarise, as shown in the table below;

(i) Where the parallel capacitance is positive, one stability condition is $R_L > L/|R|.C$ and this may be satisfied. However, a further condition is that $R_L < |R|$ which does not provide for a unique operating point on the NDR curve.

(ii) Where the parallel capacitance is negative, the circuit is stable for $R_L > |R|$.

	$C > 0$	$C < 0$
$\beta < 0$	$R_L > \frac{L}{\lvert R\rvert C}$ but for $\beta < 0, C < 0$ require very low R_L	$\forall R_L$ but $\beta < 0, C < 0$ not possible
$\beta > 0$	$R_L > \frac{L}{\lvert R\rvert C}$ and $R_L < \lvert R\rvert$	$R_L > \lvert R\rvert$
	No stable operating point	Stable condition

APPENDIX B

The Silicon Surface Field in Depletion and Inversion.

A rigorous derivation of the electric field strength in the surface depletion layer of a reverse biased MIS diode must start from Poissons equation. This may be stated in general as;

$$\frac{\partial^2\psi}{\partial x^2} = -\frac{\rho}{\epsilon_s.\epsilon_o} = -\frac{1}{\epsilon_s.\epsilon_o}.[N_D^+ - N_A^- + p(x) - n(x)] \tag{B1}$$

where N_A^- and N_D^+ are the ionised acceptor and donor densities and $p(x)$ and $n(x)$ the free hole and electron concentrations within the depleted layer. For n-type silicon, N_A may be neglected and the latter are given by;

$$p(x) = p_{n0}.\exp\left(\frac{\psi - \xi}{V_T}\right) \qquad \text{and} \qquad n(x) = n_{n0}.\exp\left(-\frac{\psi}{V_T}\right) \tag{B2}$$

Here, $p_{n0} = n_i^2/N_D$ and $n_{n0} = N_D$ are the equilibrium concentrations of holes and electrons in the neutral n-type silicon at room temperature and $V_T \equiv kT/q$ as usual.

The electric field, $\mathcal{E}$ at any point, x, in the depletion layer may be obtained from (B1) by noting;

$$\mathcal{E}(x) = \frac{-\partial\psi(x)}{\partial x} \quad \Rightarrow \quad \mathcal{E}.\frac{\partial\mathcal{E}}{\partial\psi} = \frac{\partial^2\psi}{\partial x^2} \tag{B3}$$

Then $\mathcal{E}_S$ may be obtained from (B1) by integrating between the boundary conditions at the silicon surface (where $\mathcal{E}(0) = \mathcal{E}_S$ and $\psi(0) = \psi_S$) and the edge of the depletion region (where $\mathcal{E}(W_S) = 0$ and $\psi(W_S) = 0$);

$$\int_0^{\mathcal{E}_S} \mathcal{E}.d\mathcal{E} = -\frac{q}{\epsilon_s.\epsilon_o}.\int_0^{\psi_S} N_D + \frac{n_i^2}{N_D}.\exp\left(\frac{\psi - \xi}{V_T}\right) - N_D.\exp\left(-\frac{\psi}{V_T}\right).d\psi \tag{B4}$$

with the solution;

$$\frac{\mathcal{E}_S^2}{2} = -\frac{q}{\epsilon_s.\epsilon_o}.\left[N_D.\psi_S + V_T.\frac{n_i^2}{N_D}.\exp\left(\frac{\psi_S - \xi}{V_T}\right) - V_T.N_D.\exp\left(-\frac{\psi_S}{V_T}\right)\right] \tag{B5}$$

For a modest reverse bias such that $\psi_S > 3V_T$ and for low current levels, the space charge of free electrons in the depletion layer may be neglected with little error.

Then, noting the surface hole concentration $p(0) = n_i^2/N_D.\exp(\psi_S - \xi/V_T)$, a simplified expression for $\mathcal{E}_S$ is obtained;

$$\mathcal{E}_S = -\left(\frac{2.kT.N_D}{\epsilon_s.\epsilon_o}\right)^{1/2}\left[\frac{\psi_S}{V_T} + \frac{p(0)}{N_D}\right]^{1/2} \tag{B6}$$

Here, the two terms in square brackets represent the contributions to the total field from

(i) the depletion charge due to fixed uncovered donor ions and

(ii) inversion layer charge due to free holes at the surface.

For a MIS diode in depletion, before the onset of weak inversion, the term involving p(0) may also be neglected and (B6) reduces to;

$$\mathcal{E}_S(\text{dep}) = -\left(\frac{2.kT.N_D}{\epsilon_s.\epsilon_o}\right)^{1/2}\left[\frac{\psi_S}{V_T}\right]^{1/2} \tag{B7}$$

It is sometimes useful, however to consider these two volumes of charge as sheets of charge with negligible thickness at the I-S interface. For instance, the charge density on the metal, Q_M (C/cm^2) may be given in terms of the total notional sheet charge in the semiconductor, Q_S by the so called 'charge neutrality' equation;

$$Q_M = \epsilon_{ins}.\epsilon_o.\mathcal{E}_{ins} = \epsilon_s.\epsilon_o.\mathcal{E}_S - Q_{ss} = -Q_S - Q_{ss} \tag{B8}$$

Here, Q_{ss} represents the contribution from surface states which are also treated as a lamella of charge located at the interface. Q_S (C/cm^2) may be considered to consist of the sum of depletion and inversion charge reduced to two dimensions;

$$Q_S = Q_{dep} + Q_{inv} \tag{B9}$$

where Q_{dep} and Q_{inv} are given by the integrals;

$$Q_{dep} = q.\int_0^{W_S} N_D.dx = q.N_D.W_S \qquad \text{and} \qquad Q_{inv} = q.\int_0^{x_i} p(x).dx \tag{B10}$$

Although the depletion approximation facilitates a simple solution for Q_{dep}, evaluation of Q_{inv} is more difficult. The hole population near the interface must be integrated between the surface, $x = 0$ and the plane $x = x_i$ which is defined where the weak inversion condition $E_{Fp} = E_i$ is satisfied.

Under similar circumstances, Brews[1] in a 'charge sheet model' of the channel of a MOS transistor avoided evaluating Q_{inv} by this means. Instead, he used the Gauss law relations, $Q_S = \epsilon_s.\epsilon_o.\mathcal{E}_S$ and and $Q_{dep} = \epsilon_s.\epsilon_o.\mathcal{E}_S(\text{dep})$ which, substituted into (B9) yield;

$$Q_{inv} = \epsilon_s.\epsilon_o.\mathcal{E}_S - \epsilon_s.\epsilon_o.\mathcal{E}_S(\text{dep}) \tag{B11}$$

Then, substituting from equations (B6) and (B7);

$$Q_{inv} = -\left(2.kT.\epsilon_s.\epsilon_o.N_D\right)^{1/2}\left\{\left(\frac{\psi_S}{V_T} + \frac{p(0)}{N_D}\right)^{1/2} - \left(\frac{\psi_S}{V_T}\right)^{1/2}\right\} \tag{B12}$$

REFERENCE

J.Brews, Solid State Electronics **21** 345 (1978)

APPENDIX C

The Mask Set for Fabrication of Experimental Epitaxial and Lateral MISS Devices at Durham.

A set of masks has been designed and fabricated by the author for the purpose of defining both lateral and epitaxial MISS devices with the option of V-groove isolation. The layouts of eight masks were created on the SERC 'PR1ME' system for I.C. design. Hard copies, obtained at a size corresponding to 134× the desired dimensions of the masks, were then used as templates for transfer of the patterns to 'Stabilene' (otherwise known as 'Rubylith') film using the 'cut-and-strip' technique. The first reduction was achieved by mounting the finished Stabilene sheets on a light box and photographing them using a Sinar camera at f22 onto a Kodak 'High Resolution Plate' (HRP). This provided the negatives for the second reduction stage which involved creating 4 × 8 arrays of each image on HRP's using a step-and-repeat camera.

The mask identities and details of their intended applications are given in table B.1.

Mask #	Purpose
1	V-groove isolation.
2	Gate oxide windows.
3	Diffusion windows.
4	MIS windows.
5	Substrate contact windows.
6	Metallisation mask.
7	Supplementary metal mask.
8	Diffusion windows. (complement of #3)

Table B.1

Details of the eight masks created in-house for the photolithographic processing of experimental MISS devices

The eight masks are shown superimposed in figure C.1 and the bond pad of each discrete device is labelled with its identifying number.

Devices in the series 1 to 8 have stripe-geometries (defined by mask # 4) and were intended to take the forms of alternate p^+ doped emitters and MIS collectors, thus forming a sequence of parallel lateral M-I-n-p^+ switching devices. Masks #3 and #8 serve the purpose of selecting which of the parallel stripes etched into the field oxide are doped and which host the MIS contact. Unfortunately, lateral devices of this type that have been fabricated by the author have failed to work, due, it is thought to imperfect masking of the MIS windows from the boron doping. Similarly, these masks have been used to fabricate working lateral MISIM switches and could be used to form MISM devices. Mask #6 serves its usual purpose of defining the top contact metallisation and bond pads. Mask #2 was included to provide the option of a gate oxide between the parallel stripes. The intention was to etch a large window in the field oxide and grow a gate oxide prior to definition of the stripe windows. It was then hoped to form metal gate contacts using a combination of mask #6 as a 'lift-off' mask and mask #7 as an etch mask.

The remaining devices, in the series 9 to 17, all have square geometries. Numbers 9 to 13 have the same nominal MIS contact area of $[20\ \mu m]^2$ but have epitaxial pn junction areas, defined by V-groove isolation (mask #1), varying in the sequence $[100\ \mu m]^2$, $[120\ \mu m]^2$, $[160\ \mu m]^2$ and $[240\ \mu m]^2$. Device 13 is not isolated.

Numbers 14 to 17 have the same mesa dimensions, $[120\ \mu m]^2$, but have MIS contacts increasing in size in the sequence $[10\ \mu m]^2$, $[20\ \mu m]^2$, $[30\ \mu m]^2$ and $[40\ \mu m]^2$.

Finally, it should be mentioned that mask #5 was included to allow the definition of contact windows to the p^+ substrate revealed at the bottom of the isolation moats. However, contact to the back face of the p^+ substrate was found to be quite adequate for the purpose of these investigations and mask #5 has not been required.

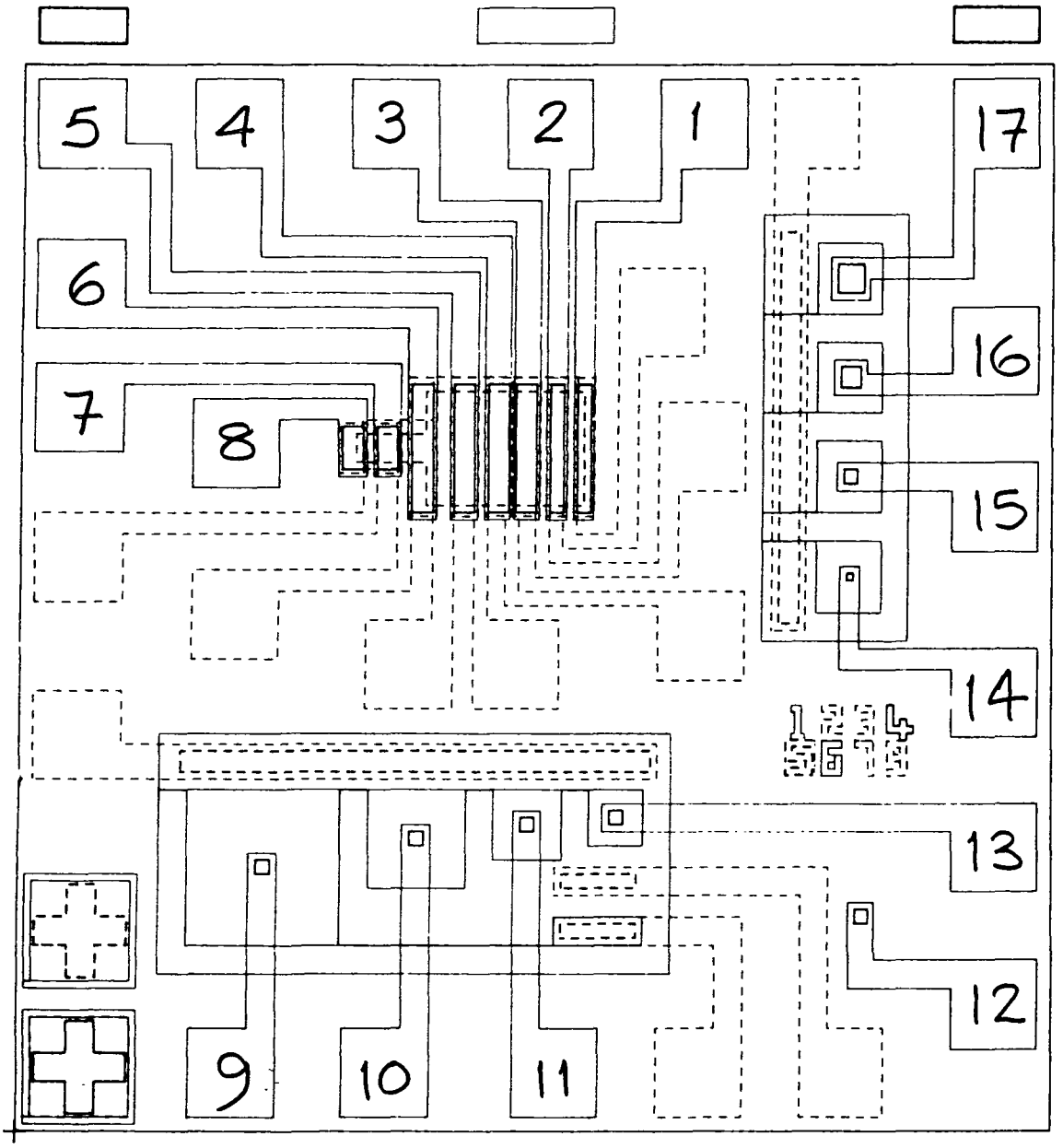


Figure C.1; The mask set created by the author for the fabrication of lateral and isolated epitaxial MISS devices. The 8 masks are shown superimposed.

APPENDIX D

Ellipsometry

Ellipsometry is a non-destructive optical technique which is used for the measurement of the thickness and refractive index of thin transparent films and is easily applied to the case of oxides on silicon[1] . The technique involves the determination of the effect of a sample on the state of polarisation of a monochromatic light beam.

The ellipsometer used for the present study was built in-house at Durham and is configured in a 'PCSA' (polariser - compensator - sample - analyser) sequence, as shown in figure D1[2] . The purpose of the polariser (a rotatable polarising plate) and compensator (a quarter wave plate) is to provide an elliptically polarised beam of light. The sample is inclined at 20° to the beam axis and the analyser plate is located in the path of the reflected beam. The incident elliptically polarised light undergoes multiple reflections at the film-air and film-substrate interfaces and emerges with its state of polarity altered. For particular states (or 'zones') of elliptical polarisation of the incident beam, a linearly polarised reflected beam will result. The angle of polarisation of the reflected light is then measured by adjustment of the analyser plate to obtain extinction (cross-polarisation).

In order to obtain cross polarization, it is necessary to alternately adjust the polariser and analyser plates, an operation which is performed manually on the system used. The extinction condition (null) is found at first by eye and then more accurately using a photomultiplier tube and meter. At the null, the angles of the two plates provide the two measured parameters, Ψ and Δ necessary for the determination of the two unknowns, n and d_{ox}, through the simple relations;

$$\Psi = A \text{ (the analyzer angle)} \quad \text{and} \quad \Delta = 2P + 90^\circ \text{ (P, the polarizer angle)}$$

The sharpness of the null depends on the lateral uniformity of the transparent film, both in terms of its thickness and composition, across the sampled area (several square millimetres in the present case). Broad nulls are generally detected for tunneling thickness oxides, indicating some non-uniformity on this scale. However, in this thesis, as in published work, no attempt has been made to quantify the effect.

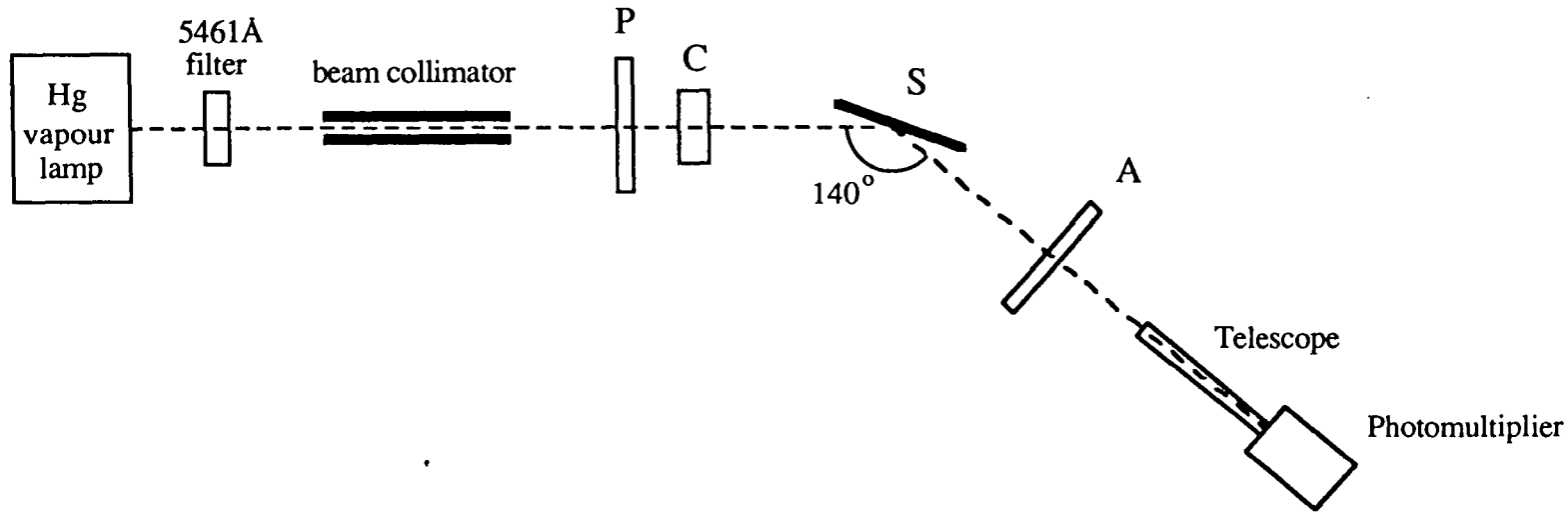


Figure D1; Schematic diagram of the ellipsometer used for the measurement of thin films on silicon.

The basic equation of ellipsometry is;

$$\rho = \tan \Psi . e^{i\Delta}$$

where ρ is the ratio of the parallel and perpendicular complex reflection coefficients, $\rho = R_p/R_s$. The calculation of n and d_{ox} from Ψ and Δ is very complicated[3] . As such, it is normal practice in ellipsometry to produce tables or graphs of Ψ and Δ (calculated for a given substrate, light wavelength and angle of incidence) as functions of d_{ox} for a range of n values. A typical set of curves obtained by this means for the case of a silicon substrate is shown in figure D2(a). Here, the range of n values is close to that of bulk silica (1.462) and is thus relevant to the study of tunnel oxides. Important features to note in these plots are that the curves converge for $d_{ox} < 100$Åand that the film thickness is a strong function of the Ψ parameter, making a precise determination more difficult in this range. Figure D2(b) shows the curve for $n = 1.46$ expanded over $d_{ox} = 0$ to 100Å.

Similar curves have been produced for the range of refractive indices relevant to silicon rich oxide films and are shown in figure D3. However, in practice these graphs were not used since more accurate data may be extracted by computation. A program written by Dr D.Buchanan (Dept. of Applied Physics and Electronics, University of Durham) to calculate Ψ and Δ from n and d was adapted by the author to automatically find a best fit to the measured values by incrementing n in steps of 0.005 and, within those steps, incrementing d by 0.5Å. By this means, n and d may be extracted to a much greater degree of accuracy than is inherent in the measurement technique itself. Experimental errors arising in null detection and in the initial calbration of the apparatus are expected to amount to about 10Å[2] , far exceeding the increments used in the computation.

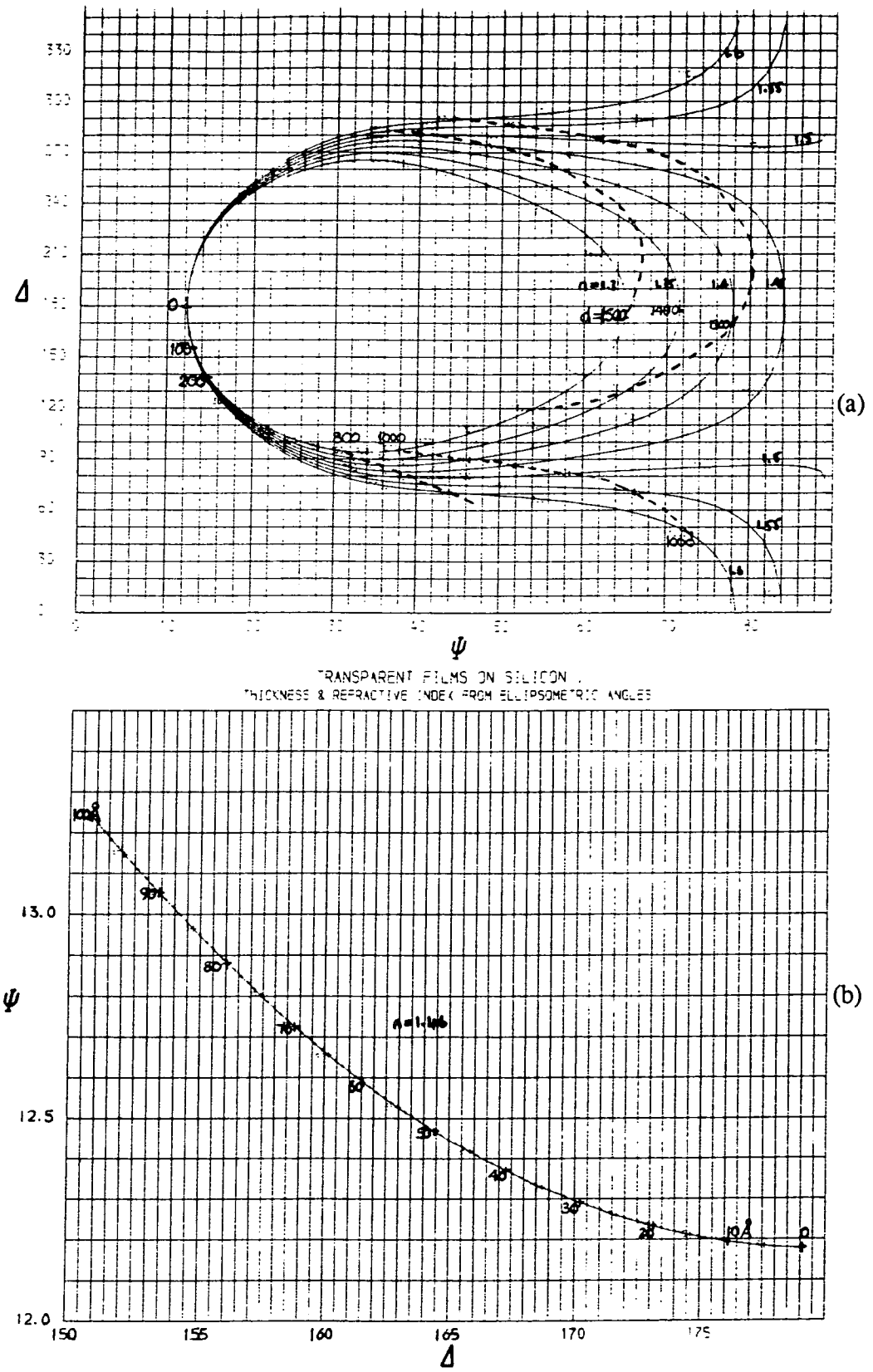
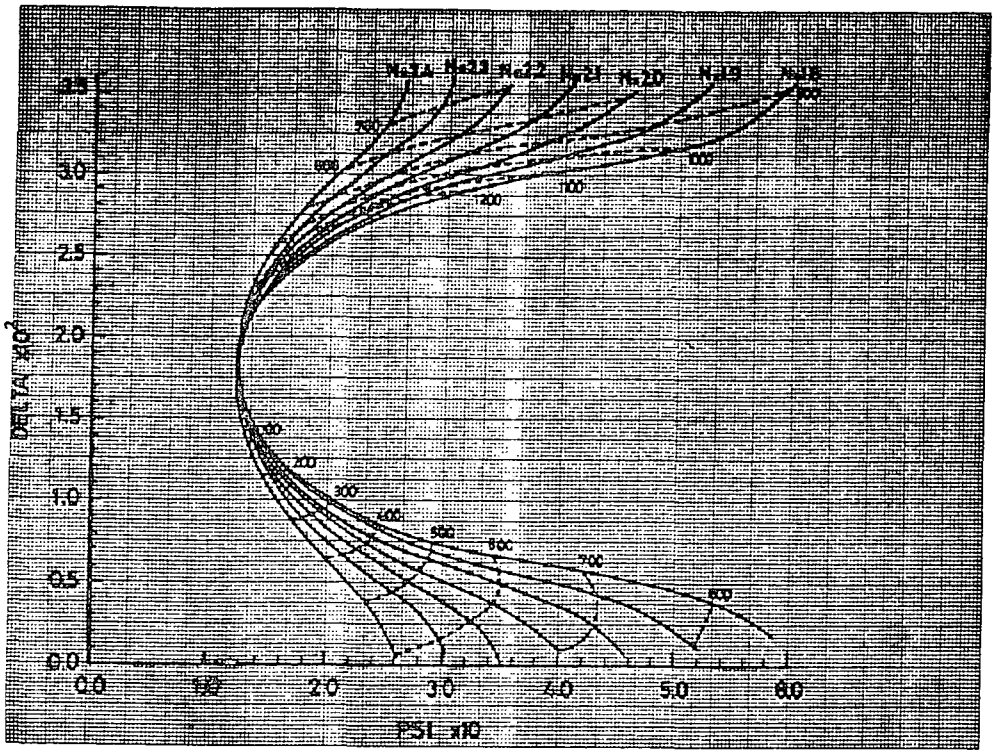
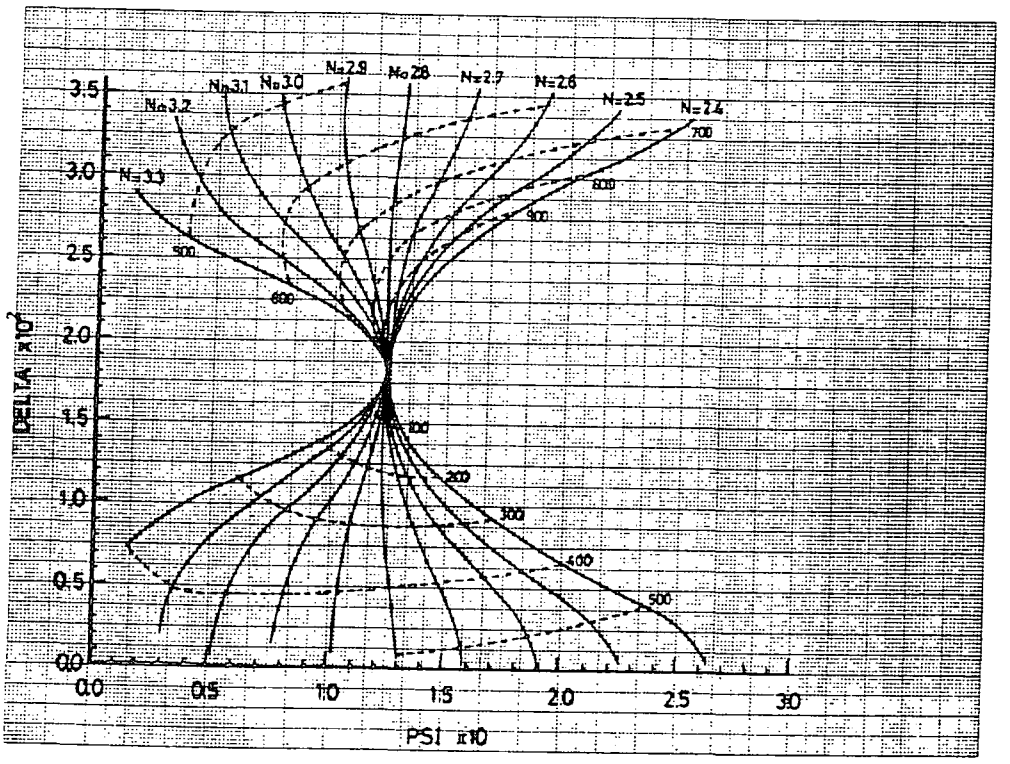


Figure D2; (Ψ,Δ) plots computed as continuous functions of thickness for various values of refractive index;
(a) n=1.3 to n=1.6
(b) n=1.46, d=0 to 100Å



(a)



(b)

Figure D3; (Ψ,Δ) plots computed as continuous functions of thickness for a range of values of refractive index typical of SRO films;
(a) n=1.8 to n=2.4
(b) n=2.5 to n=3.3

REFERENCES

1. R.J.Archer, J. Opt. Soc. America **52** 970 (1962)
2. E.G.Lloyd, Internal Memo, Department of Applied Physics and Electronics, University of Durham (1982)
3. K.A.Azzam and N.M.Bashara, *Ellipsometry and Polarized Light* (North-Holland, 1977)