Failure mechanisms in MOS devices

This item was submitted to Loughborough University’s Institutional Repository by the/an author.

Additional Information:

- A Doctoral Thesis. Submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy of Loughborough University.

Metadata Record: [https://dspace.lboro.ac.uk/2134/11005](https://dspace.lboro.ac.uk/2134/11005)

Publisher: © E.A. Amerasekera

Please cite the published version.
This item was submitted to Loughborough University as a PhD thesis by the author and is made available in the Institutional Repository (https://dspace.lboro.ac.uk/) under the following Creative Commons Licence conditions.

Attribution-NonCommercial-NoDerivs 2.5

You are free:
- to copy, distribute, display, and perform the work

Under the following conditions:

Attribution. You must attribute the work in the manner specified by the author or licensor.

Noncommercial. You may not use this work for commercial purposes.

No Derivative Works. You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.

Your fair use and other rights are in no way affected by the above.

This is a human-readable summary of the Legal Code (the full license).

Disclaimer: 

For the full text of this licence, please go to:
http://creativecommons.org/licenses/by-nc-nd/2.5/
FAILURE MECHANISMS
IN
MOS DEVICES

by

EKANAYAKE AJITH AMERASEKERA BSc.(Hons), DIS.

A Doctoral Thesis

Submitted in partial fulfillment of the requirements for the award of Doctor of Philosophy of the Loughborough University of Technology.

8th August 1986

Research Supervisor: Professor D.S. Campbell

© by E.A. AMERASEKERA 1986
One must have a meditative quality of the mind, not occasionally, but all day long. And this something that is sacred, affects our lives not only during waking hours but during sleep.

J. KRISHNAMURTI

Truth and Actuality
ABSTRACT

Continuous and pulsed voltage stressing of metal oxide semiconductor (MOS) transistors and capacitors has been investigated. The experimental work followed a survey of failure mechanisms in semiconductor devices which identified Electrical Overstress Damage (EOS)/Electrostatic Discharge (ESD) damage as the most frequent cause of failure, accounting for over 50% of all damage observed. The survey itself, covered all aspects of semiconductor reliability including reliability modelling and quality assurance.

A qualitative model of oxide breakdown in MOS structures was developed as a result of the experimental work. Two different mechanisms have been proposed for continuous and pulsed voltage breakdown.

Continuous voltage breakdown simulating EOS conditions, was temperature and voltage dependent. The long time-scales involved, lead to a model whereby breakdown is the result of conduction of charge carriers through the oxide, via electron traps and impurity sites with energies in the forbidden gap. Pulsed voltage breakdown simulating ESD, was voltage dependent but not temperature dependent. The very short time-scales involved indicate that breakdown is the direct result of electron transport in the oxide conduction band. Electrons are injected into the conduction band via quantum-mechanical tunnelling from the cathode.

Both mechanisms were found to be dependent on the surface charge concentration of the silicon and, therefore, polarity dependent. The models explain this effect by analysing the charge injection process under high electric fields.
ACKNOWLEDGEMENTS

During the course of any research studentship there are many people who would have helped in some manner, towards the final success of the thesis. I wish to thank all my friends for their encouragement and support (usually over the glass of an intoxicating liquid) during the numerous dark (and bright) days encountered over the last three years.

I owe my thanks to colleagues in the Electronic Component Technology Group and in the Physics Department at Loughborough University for the many rewarding discussions I have had with them. In particular I would like to express my gratitude to my supervisor Professor David Campbell, for his willingness to discuss this research and provide advice, as well as for his tremendous moral support during the course of my studentship. Dr. Vaughan Williams, who has since departed to conduct research in the warm Caribbean, gave me much assistance and encouragement during the early days of my studentship. I am indebted to Roger Tomlinson for his electronic wizardry in producing some of the more complex circuits needed for the experiments. My thanks to Andy Franklin for his help with the preparation of this manuscript, the coffee and the gossip. To Jeff Jones for providing the computation for working out the numerical values of the integrals used in the course of this work, and to Sana, Ian, Danny, Ratne and Joe for being there.

Plessey Research (Caswell) Ltd. provided the wafers used in this study and my thanks to them, particularly Nick Armstrong, Ray Oakley, Bill Holt and John Cutter for their help. The work has been carried with the support of the Procurement Executive, Ministry of Defence, and I am grateful to them for the opportunity. Abstracts of the three reports associated with this contract are presented in Appendix 3.
I would like to thank Ramani for helping me out with the typing when most needed. Thanks also to Christine for her help in preparing the papers and manuscripts associated with this project, to Yasmin for her typing, and to Eileen for some of the diagrams. I am indebted to Dileepa for his unselfish assistance in putting together the final version of this thesis.

A special 'thank you' to Anoma for her help with this work and for being such a special person.

These acknowledgements would be incomplete if I did not express my eternal gratitude and love to my mother, Aloma, and my sister, Rosharni, for their continued affection and loving support, across the miles, over the past seven years. I dedicate this work to the two of them.
**NOMENCLATURE AND SYMBOLS**

A - Amperes  
As - Arsenic  
B - Boron  
C - Capacitance  
\(c.m\) - centimeter  
CMOS - Complementary Metal-Oxide-Semiconductor  
\(C_{ox}\) - Capacitance of gate oxide  
\(C_s\) - Total Capacitance across an MOS capacitor  
\(C_{sp}\) - Capacitance of MOS space-charge region  
CT - Curve Tracer  
C-V - Capacitance - Voltage  
\(^\circ\)C - degrees Celsius  

D - Transmission Coefficient  
D.C./d.c. - direct current  
D-mode/type - Depletion - mode/type  
\(D_n\) - Diffusion Constant for Electrons  
\(D_p\) - Diffusion Constant for Holes  
DUT - Device Under Tests  
e - electronic charge \((1.6 \times 10^{-19} \text{ Coulombs})\)  
E - Energy  
\(E_a\) - Activation Energy  
\(E_{Fi}\) - Intrinsic Fermi level  
\(E_{Fn}\) - electron quasi-Fermi level  
\(E_{Fp}\) - hole quasi-Fermi level  
\(E_G\) - bandgap energy  
\(E_i(x)\) - Ionization Energy
EOS - Electrical Overstress
ESD - Electrostatic Discharge
E - mode/type - Enhancement-mode/type
eV - electron volts
FET - Field Effect Transistor
FIT - Failures-in-Time (1 failure in $10^9$ device hours)
G - Generation rate
GaAs - Gallium Arsenide
$g_{ds(sat)}$ - drain-source conductance (saturated)
$g_m$ - mutual transconductance
HTRB - high temperature reverse bias
hrs - hours
Hz - Hertz (cycles per second)
I - current
$\text{I}_{D(sat)}$ - drain current (saturated)
$\text{I}_{DS}$ - drain-source current
$\text{I}_G \text{ leak}$ - gate leakage current
IGFET - Insulated Gate FET
I-V - current - voltage
j - $\sqrt{-1}$
J - current density
$\text{J}_n$ - electron current density
$\text{J}_p$ - hole current density
- Boltzmann's constant \(1.38 \times 10^{-23} \text{ J/K}\)

- Kelvin

- kilo-ohm \((10^3 \ \Omega)\)

- length

- Debye length

- Large Scale Integration

- mass

- effective mass

- electron mass

- hole mass

- metal-semiconductor FET

- metal-oxide-semiconductor

- millimeter

- millisecond

- Mean Time Between Failures

- Megavolt

- Mega Ohm \((10^6 \ \Omega)\)

- electron-doped

- excessively electron-doped

- nanofarad

- acceptor doping concentration

- donor doping concentration

- nanometer

- n-channel MOS

- electron concentration in bulk of p-type material
ns - nanosecond
N(E<sub>x</sub>) - supply function

O - Oxygen

P - Phosphorus
p - hole-doped
p<sup>+</sup> - excessively hole-doped
pF - pico Farad
PMOS - p-channel MOS
polySi - polysilicon

Ppo - hole concentration in bulk of p-type material
ps - picosecond
PX - momentum in x direction

Q - Coulombic charge density per unit area
Q<sub>s</sub> - surface charge density per unit area
Q<sub>ss</sub> - interface states charge density per unit area
Q<sub>eff</sub> - effective charge density per unit area
QA - quality assurance
q - electron/hole charge (1.6 x 10<sup>-19</sup> Coulombs)

R - Recombination rate; Resistance
R<sub>GD</sub> - Gate-drain resistance
R<sub>GS</sub> - Gate-source resistance

s - seconds
S - Siemens
secs. - seconds
Si - silicon
SiO₂ - silicon dioxide

T - temperature
t - time
t-rise - rise time
t-decay - decay time
TTFF - Time To First Failure

V - volts
V_{appl} - applied voltage
V_C - capacitor voltage
V_{DS} - drain source voltage
V_{DS\ eff} - effective drain source voltage
V_G - gate voltage
V_{LSSI} - very large scale integration
V_{SD} - very small dimension
V_{sub} - substrate voltage
V_T - threshold voltage

W - width

Å - Angstrom (10^{-8} cm)
α(F) - Ionization coefficient
β = q/kT = 39 V^{-1}
β_s - Shottky coefficient
β_{PF} - Poole-Frenkel coefficient
$\Delta V$ - change in voltage

$\Delta t$ - change in time

$\Delta I$ - change in current

$\varepsilon_0$ - permittivity of free space ($8.854 \times 10^{-14}$ F cm$^{-1}$)

$\varepsilon_i$ - permittivity of insulator

$\varepsilon_s$ - permittivity of silicon ($1.04 \times 10^{-12}$ F cm$^{-1}$)

$\lambda$ - electron affinity

$\hbar$ - $\hbar/2\pi$, adjusted Plank's constant ($1.055 \times 10^{-34}$ Js)

$\phi_F$ - Fermi potential

$\phi_{ms}$ - metal-semiconductor work function

$\mu$ - mobility

$\mu_n$ - electron mobility

$\mu_h$ - hole mobility

$\mu_s$ - microseconds ($10^{-6}$ s)

$b$ - base failure rate

$p$ - Failure rate

$\pi$ - 3.1416

$\tau_{\text{min}}$ - minority carrier response time

$\tau_{\text{maj}}$ - majority carrier response time

$\rho(x)$ - charge density

$\Phi_s$ - Band-bending at surface.
CONTENTS

ABSTRACT

ACKNOWLEDGEMENTS

NOMENCLATURE AND SYMBOLS

1. INTRODUCTION
   1.1. Historical background
   1.2. The present and the future
   1.3. Failure mechanisms in semiconductor devices
   1.4. Chapter synopses
   1.5. References

2. SURVEY ON FAILURE MECHANISMS IN SEMICONDUCTOR DEVICES
   2.1. Introduction
   2.2. Event-related failures
   2.3. Intrinsic failure mechanisms
   2.4. Extrinsic failure mechanisms
   2.5. Improving reliability
   2.6. The physics of degradation mechanisms
   2.7. Review of published work on MOS breakdown
   2.8. Summary
   2.9. References

3. EXPERIMENTS
   3.1. Introduction
   3.2. The MOS structures
   3.3. Apparatus
   3.4. Measurements
   3.5. Experimental procedure
   3.6. Summary
   3.7. References

4. RESULTS
   4.1. Introduction
   4.2. Pulsed voltage stress experiments
   4.3. Sequential pulsing
   4.4. Continuous voltage stress experiments
   4.5. Summary

5. DISCUSSION
   5.1. Introduction
8. A comparison between GaAs MESFET and Si NMOS ESD behaviour.

9. ESD pulse and continuous voltage breakdown in MOS capacitor structures.
CHAPTER 1

INTRODUCTION

1.1. HISTORICAL BACKGROUND

The origins of research into semiconducting materials can be traced back as far as 1833 [1]. Michael Faraday probably made the first experimental observation in the semiconductor field around this time, when he found that silver sulfide had a negative temperature coefficient of resistance [2]. Other conductors known at the time, all had resistances which increased with temperature. The next significant discovery was in 1874, when Braun found that the resistance of contacts between certain metals did not obey Ohm's law, and depended on the magnitude and polarity of the applied voltage [3]. Schuster (1874), made the same observation for contacts between tarnished and untarnished copper wires [4].

It was not, however, until the 1930's and the advent of quantum mechanics, that a real understanding of semiconductor properties began to evolve. The development of solid state physics based on the concepts of energy levels with one electron per level, electron spin, the Pauli exclusion principle, and Fermi-Dirac statistics, enabled great strides to be made in interpreting the behaviour of electrons atoms and molecules. Sommerfeld [5] presented his free-electron model of metallic conduction based on quantum mechanics in 1928, while perhaps the most important contribution to the theory of semiconductor physics was made by Wilson in 1936 [6]. He presented a quantum mechanical model of a solid semiconductor, in which he showed that the movement of electrons as waves throughout the solid set up interference patterns. Hence, certain energy levels were excluded (forbidden), leading to the concept of energy bands in solids.
In 1925, the first known transistor was designed by Lillienfeld who filed patents for it in Canada and the U.S.A. [7][8]. Current flowed between two gold electrodes, through a copper sulfide channel and was controlled by a potential on a third electrode (aluminium). The principle was similar to that of a MESFET. He later filed a patent in 1928 for an improved FET biased on the insulated gate electrode, as in the MOSFET. Aluminium oxide was used as the insulator with copper sulfide again being used as the semiconducting material. In 1928, Lillienfeld also patented a metal-base bipolar transistor which consisted of several layers of metals and semiconductors with rectifying properties. While it seems certain that Lillienfeld was aware of the empirical theory of the operation of transistors it is not known whether any of these devices was actually constructed. However, it is believed that if manufactured using fabrication techniques available today, those devices would have worked.

Shockley, Bardeen and Brattain, in 1947, constructed the first operational transistor. It was a bipolar device based on germanium and was known as the point contact transistor. They also made a vast contribution to the development of the present theoretical understanding of semiconductors [9]-[11][19]. For this work the three of them received the Nobel Prize for physics in 1956.

Theoretically, the FET was still the most logical structure to form an amplifying device. Experimentally it was impossible to obtain efficient devices, the amplifying effect observed being much less than predicted. It was not until the importance of surface energy states was realised, that an FET was finally constructed and described by Shockley in 1952 [12]. Majority carriers formed the current in these devices making them unipolar transistors.
Insulated gate transistors (IGFETs) had to wait until the advances in semiconductor fabrication techniques had been made. In particular, the development of planar diffusion and the growing of passivation (SiO₂) layers, in 1959. The first metal oxide semiconductor FET (MOSFET) was proposed by Kahng and Atalla in 1960 [7]. Hofstein and Heimann, in the early 1960's, further developed this device to include enhancement and depletion mode transistors, capable of operating with different polarity gate voltages [13]. MOSFETs in the 1980's, form the backbone of the semiconductor industry, comprising a major portion of every piece of electronic equipment manufactured.

Looking back over the years, it is perhaps possible to identify three important aspects of the developments in semiconductor technology which enabled the rapid advancement towards today's devices.

1) The potential difference between p-doped and n-doped semiconductors, which is 0.3 V at a typical silicon p-n junction. The formation of a single-phase interface between p and n-type regions of the same material is important to the semiconducting properties of the devices.

2) The development of single crystal growth techniques which can provide long (>15 cms.), large diameter (>7.5 cms.) crystals, with impurities of $=1 \cdot 10^9$ has been vital to the advancement of semiconductor device technology.

3) The ability to grow a natural amorphous oxide on silicon, with good dielectric and passivating properties, has made it possible to produce high quality, high performance devices.
1.2. THE PRESENT AND THE FUTURE

The shrinking of the semiconductor device has been one of the great phenomena of recent times. Very Large Scale Integration (VLSI) has made it possible to pack large numbers of active devices into very small areas of silicon. Today, commercially available VLSI circuits can have >800 devices per mm². Such circuits are capable of performing a host of complex functions, ranging from high speed computation to automated control systems.

Research in semiconductors is still concentrating on making the transistor even smaller [14][16]. The reasons behind this are basically economic. Since the cost of the silicon itself makes up the major portion of a manufacturer’s financial burden, it stands to reason, that, by increasing the number of transistors per unit area of silicon, the cost of each transistor is effectively reduced. Large packing densities as found in VLSI circuits, have the added advantage of enabling complex circuitry to be designed with a minimum of external connections. Internal interconnections are more reliable and hence VLSI design can increase the reliability of complex circuits. The number of pins as a function of the number of gates for silicon microprocessors and random logic i.c.’s, is shown in Figure 1.1 [36]. It can be seen that 50 gates would have 10 pins in an i.c., while, if the number of gates is increased to ≈10⁶, the number of pins increases by only a factor of 10.

Silicon is currently the premier semiconductor material and is believed that it will remain so until well into the 21st century. Recently, a great deal of attention has been focussed on gallium arsenide (GaAs), a compound semiconductor with certain properties which can prove to be an improvement on those of silicon [16]. The higher mobility of carriers in GaAs and their lower effective mass means that GaAs devices would be more suitable in very fast switching systems. Improvements in modern fabrication techniques have also
FIGURE 1.1. The relationship between the number of pins and the number of gates found for silicon microprocessors and random logic LSI and VLSI chips (after ref. 36)
made it possible to consider new device structures which can take advantage of the properties of GaAs to produce VLSI circuits capable of operating at very high speeds.

Advances in semiconductor fabrication techniques, means that it is now possible to manufacture very small dimension (VSD) silicon devices with operational sizes in the submicron region [16]-[18]. MOSFETs incorporating submicron channel dimensions would introduce several new physical effects which must be considered in any analysis.

1.3. FAILURE MECHANISMS IN SEMICONDUCTOR DEVICES

The survey described in this thesis explored the available literature on failure mechanisms and combined it with several discussions with quality and reliability personnel in the semiconductor and equipment manufacturing industry. A shortlist of the most common mechanisms was, therefore, obtained. Analysis of the quality assurance side of the manufacturing industry also showed that the understanding of failure mechanisms was important in achieving higher quality devices.

1.4 CHAPTER SYNOPSIS

The survey on failure mechanisms is described in Chapter 2. The chapter concludes with a review of published work on MOS oxide breakdown.

Chapter 3 describes the experimental work on MOSFETs and MOS capacitors. Results of pulsed and continuous voltage stressing are in Chapter 4. These are discussed in Chapter 5.
Models of the breakdown processes are qualitatively developed in Chapter 6. A description of the theories used is given, although only the important mathematical equations are presented in the text.

Chapter 7 discusses the major points made in each of the preceding chapters and presents an overall view of the whole project.

1.5 REFERENCES


CHAPTER 2

SURVEY ON FAILURE MECHANISMS IN SEMICONDUCTOR DEVICES

2.1 INTRODUCTION

This chapter begins with a study of the literature on failure mechanisms in semiconductor devices. The work has been coupled to discussions with reliability and quality engineers in the semiconductor and equipment manufacturing industries. Hence, it has been possible to evaluate the dominant failure mechanisms and also to evaluate the methods used to improve device reliability.

The next part of the chapter reviews the published work on oxide breakdown in MOS devices. Investigations into the breakdown of oxides is the focus of the work in this thesis.

2.2. EVENT-RELATED FAILURES

Event-related failures are due to influences external to the actual device. They can be caused by improper application of equipment, careless handling of components or operations outside the component specifications.

Manufacturers of semiconductor devices, as well as equipment manufacturers and users claim that electrical stress damage accounts for over 50% of their in-circuit failures. This makes these failure mechanisms the largest contributors to reliability degradation in the electronics industry [1] -[5].
FIGURE 2.1. Photomicrograph of a GaAs MESFET, showing the effect of EOS. Burnout has occurred in the channel region between the gate and the source. (12500×)

FIGURE 2.2. Schematic of an input protection circuit.
Electrical stress can be defined in two categories, namely, Electrical Overstress (EOS) and Electrostatic Discharge (ESD). Figure 2.1. shows a photomicrograph of a typical EOS/ESD damaged area on a GaAs MESFET. The hole in the semiconductor material is due to the localised heating caused by the high electrical stress between the gate contact and the source/drain contact of the transistor.

EOS failures are associated with hot-spot development in bipolar devices [6]. In MOS devices the stress causes the oxide (field or gate) to breakdown [7].

These failures can be prevented by incorporating protection circuitry capable of dissipating the excess electrical stress. A schematic of an input protection circuit is shown in Figure 2.2. The diode is in reverse bias and is designed to turn on at a voltage below the breakdown threshold of the transistor. More complex structures are used to obtain the high levels of protection required under ESD conditions (up to 4kV) [8][9].

ESD is the result of the discharge of static through a semiconductor device. Voltages may range from 100V to 20kV depending on the environment. The discharge could take place by touching the device after accumulating static on oneself. It is also possible to transfer static charge onto a device. In this case discharge takes place when the device is placed in contact with a ground plane [3][4][10][11][18].
2.3. INTRINSIC FAILURE MECHANISMS

Failures which can be traced to the fabrication stages of device manufacture are termed intrinsic. It is here that inherent flaws such as crystal defects and low quality thermal oxide growth can lead to failure during the working life. Contamination can also be introduced at this stage which will eventually limit the device lifetime as discussed by Lange [12] and Schmidt [13].

Edwards [14] in a review paper on MOS failure mechanisms identifies defective gate oxides, the trapping of charge in the oxide at impurity sites, and the presence of contaminating ions as being the major problems. Partridge [15] in a review paper on bipolar failure mechanisms shows crystal defects to be important. Diffusion of dopant ions along these crystal defects may result in parametric shifts of the electrical characteristics. Low frequency noise has been traced to the presence of crystal defects [17]. In extreme cases devices may short-circuit across a junction resulting in catastrophic failure [16]. Stojadinovic [64] looks at the principal failure mechanisms in bipolar and MOS devices in a review paper published in 1983.

As technologies mature and fabrication processes improve, intrinsic failures are reduced. Instead, device reliability is more influenced by extrinsic failure mechanisms.

2.4. EXTRINSIC FAILURE MECHANISMS

Extrinsic failure mechanisms are introduced by the packaging and the interconnections. The term packaging describes the die attach, the lead frame and the
ceramic or plastic encapsulation of the device. The metallisation deposited on the chip, the bonds and the attached leads are considered to be the interconnections.

2.4.1. The packaging

Semiconductor devices can be packaged either hermetically in ceramic or completely in a solid epoxy [20]. The former are known as Ceramic Dual-in-Line Packages (CERDIPs), while the latter are Plastic Encapsulated Devices (PEDs).

The critical element in the CERDIP structure is the leaded "sealing" glass which must provide a reliable hermetic seal against corrosive elements. In addition, moisture ingress along the lead frame causes corrosion of the metallisation on the die with the occurrence of eventual failure [20].

PEDs are prone to moisture ingress through the porous plastic which means low reliability can be expected in high humidity environments. Also, the epoxy introduces contaminants which, because of the direct contact between the plastic and the silicon, will be a reliability hazard [23][24]. However, the introduction of inhibitors against moisture and contamination has improved the reliability of PEDs to the extent that they are found to be comparable to that of the CERDIPs in benign environments [25][26]. This is supported by data obtained from both laboratory lifetests [20][21], and from field failures [22][25].

2.4.2. The die attach

Two main failure mechanisms are associated with the die attach. The first is the integrity of the contact made between the die and the lead frame. Voids in the die attach lead to eventual detachment or thermal/electrical failure due to bad contact
[27][28]. Secondly, there is the problem of contaminants and moisture introduced by the epoxy die attach system or the lead frame. These are discussed by Manchester [29] and Powell [30]. PEDs are found to be inferior to CERDIPs in this respect because of the problems associated with the eutectics. The use of side-brazed hermetic packaging as opposed to the dual-in-line lead frame is found to reduce moisture ingress [20]. This is because, unlike the CERDIPs, the lead frame in side-brazed packages is not in direct contact with the interconnects. These devices, therefore, tend to be less prone to contamination and more highly resistant to moisture.

2.4.3. The metallisation

Mechanisms affecting the metallisation are a) corrosion, b) electromigration, c) contact migration, d) stress relief migration. In addition, problems related to oxide steps on the die surface can lead to cracks forming in the metal film at those points. This is illustrated in Figure 2.3. along with the fabrication technique used as a solution.

Aluminium is the most extensively used element for metallisation and interconnections in i.c's.

**Corrosion:** A survey of corrosion failure mechanisms in microelectronic devices has been conducted by Schnable et al [31]. Corrosion is an electrochemical mechanism which occurs in the presence of moisture and a d.c. operating potential. Chlorine or sodium ions act as catalysts to the process [32].

**Electromigration:** When the current density in the metal is greater than $10^6 \text{A cm}^{-2}$, electromigration takes place. The continuous impact of electrons causes the aluminium grains to move in the direction of the electron flow. A void is created at one end of the track while the metal is piled up at the other end [33]. The Al grain size
Figure 2.3. (a) Metal deposited over a steep step.

Figure 2.3. (b) Formulation of a microcrack.

Figure 2.3. (c) Tapering of steps to eliminate microcracks

Figure 2.3. Schematic illustration of microcracks in Aluminium metallisation.
Wire (Au)

Loop - if too tight, tension in wire is high and tends to fracture; if too loose then the wire is free to move and may short circuit with adjacent wires.

Lag - tension is important (as described for the loop).

Heel - weak point if wire in high tension

Wire attached to lead frame by wedge bond

Figure 2.4. Diagram of a bond wire showing potential weak points.
is important in this process; smaller grain sizes lead to an increase in mass transport. As a result of work by researchers such as Black [33][34] and D’Heurle [35] into the physics of electromigration methods have been developed to minimise the problem. These methods centre on the use of Al alloys incorporating small percentages of Cu or Si [77].

Contact migration: Migration can also take place at the Al or Si interface of the metal contacts [36]. The migrating species could be either Al in Si or Si in Al. Failure may manifest itself as either a short-circuited junction or an open contact [36][37]. Inhibiting the diffusion of Al and Si is accomplished using Al alloyed with Si or Cu. Complex alloys such as PtSi-Ti/W-Al is found to greatly enhance contact integrity [2].

Stress relief migration: Deformation of the metallisation occurs as a result of the movement of atoms from areas of high stress. Failures occur due to whisker growth [38][39] under compressive stress, or by the formation of hillocks and voids [15][40]. Again, the use of alloys can prevent this mechanism.

2.4.4. Bonding

The bond can be considered to be one of the weakest areas of the package and interconnects, although improvements in the manufacturing processes and tighter quality controls are reducing the number of inferior products reaching the consumer [41][45]. Figure 2.4. illustrates the areas in a wire bond where failures are commonly observed.

At the interface between the gold bond and the Al metallisation, the formation of Au₂Al in excess is a serious problem. The quality of the contact at the base of the bond is undermined by the intermetallic and results in failure [42]. Adjacent bond
wires may short-circuit if the loop between the die and the lead frame has too much sag. If the loop is too tight, the tension created at the heel of the bond and in the wire itself can cause fracturing of the bond metal [43].

Moisture aided migration at the base of the bond and the presence of contaminants in the metallisation can expedite voiding and bond detachment [44][45]. Whisker growth as a result of excessive bonding pressure has already been discussed. If the bond pressure is too low, fractures can result at the bonding interface.

In the moulding process used for PEDs, the backwash of the plastic compounds could force the bond wires to short-circuit. The moulding process can also cause high stress in the bond wires [46][47].

Process improvements and design optimisation have reduced failures due to intermetallic formation. Careful monitoring of the bonding pressure and the loop formation has improved the quality of the bonds.

2.4.5. Alpha particle radiation

Trace impurities of radioactive elements (e.g. Thorium, Uranium) in the packaging material can emit alpha particles with energies up to 8 MeV [48] - [51]. Interaction of these particles with ions in the bulk material results in the generation of hole-electron pairs. The charges move by diffusion through the bulk of the device and significantly affect the operation of a device such as a dynamic RAM. The alteration of the stored charge in a memory cell due to the radiation causes the stored data to be changed.
Prevention techniques are focused on limiting the collection efficiency of electrons at the storage nodes [48]. A trade-off between device parametric degradation and soft error rates is recommended [52].

Other forms of radiation, e.g. gamma-rays and X-rays can also result in similar degradation [53] - [56]. One of the major problems encountered with external radiation is that of CMOS latch-up [55][57]. The radiation causes a parasitic bipolar transistor in the CMOS structure to turn on, and the output of the device latches to one of the supply rails, rendering the device useless.

2.5. IMPROVING RELIABILITY

2.5.1. Screening

Screening is the process by which defective devices are identified and removed from a production patch. A study of the device physics and its failure mechanisms would enable a good screening procedure to be developed based on the activation of the relevant defect mechanisms. For example, Crook [7] shows how a high voltage prestress across an oxide dielectric can be used to cause premature failure of weak oxides. Therefore, the reliability of the screened batch of devices is improved. In Figure 2.5., the failure rate of a batch is plotted as a function of time. An electric field of 2.5 MV/cm is applied to the device for one second. The dotted line shows the increased failure rate due to the stress. Infant mortalities due to weak oxides have, therefore, been eliminated.

The standard tests for microelectronic devices are given in the MIL-STD-883C [58], BS 9300 [59] and BS 9400 [60]. An example of the screening processes used by a semiconductor manufacturer is given in INTELs reliability report on the 2164 A D-
FIGURE 2.5. An example of screening. A high voltage prestress across an oxide dielectric is shown to eliminate defective oxides.
RAM [63]. Ryerson [61] presents a comprehensive review paper on reliability testing and screening methods up to 1978.

Principal tests used in screens are:

1. High temperature burn-in; devices are typically subjected to 125 °C for 48 hours. This is considered to be more a process monitor than a screen since it is intended to remove devices which fail as a result of manufacturing defects, rather than to address a specific failure mechanism [62].

2. High temperature storage test; devices are baked at around 250 °C for hermetic devices and 150 °C for PEDs. The intention being to detect any instability mechanisms in the devices [63].

3. High temperature reverse bias test; overvoltages (50% above the maximum specified voltage) are combined with high temperatures (150 °C) to make an effective screen against mobile ion contamination, particularly in MOS devices [64].

4. Thermal shock; a technique to evaluate the integrity of the package. The device is alternately subjected to temperatures between -65 °C and +125 °C for about 10 seconds at each level. Bad thermal matching, for example between the passivation layer and the die, would be detected by this test. Crystal defects are also sensitive to this screen.

5. Temperature cycling; structural defects are aggravated by temperature cycling. Devices are placed in hot/cold air-to-air cycling chambers for 10 cycles of -65 °C to +150 °C and held for approximately 10 minutes at each extreme.

6. Humidity tests; usually at 85 °C/85% Relative Humidity (RH) or 120 °C/2 atmospheres at 100% RH. These tests indicate the resistance of the packaged device to failure mechanisms associated with moisture ingress, e.g. corrosion.
The above list is not exhaustive and other tests are also carried out depending on the level of screening specified. These would test for hermeticity, loose debris in hermetic packages, and the quality of the bonds. All screens must be followed by electrical functional tests [65].

2.5.2. QUALITY ASSURANCE

Generating quality specifications, control of quality conformance with specifications through incoming inspection, and quality inspection of the finished products are all parts of the QA system [66] - [68].

Wernik [67] and Hutchins [68] recommended the policy of "getting-it-right-first-time", as opposed to testing-in quality. In order to achieve this goal, the physics of the device and the mechanism of degradation must be understood. Hutchins shows that the failure rate of microprocessors was halved as a result of making improvements to fabrication techniques and converting to improved material. This is an example of progress along the learning curve. Reduction of flaws generated by poor quality workmanship was observed after the workforce was made aware of required product quality levels.

Quality is measured in terms of acceptable quality levels (AQLs) where sample numbers of a batch are tested to a given specification. The batch is passed depending on the percentage number of defects present [66]. High risk samples, where the sample size is too small for the volume of production, results in low confidence levels in the reliability estimates. Procurement specifications dictate the levels of sampling and screening required for a defined application.

2.5.3. Reliability Estimations
Reliability estimates provide a yardstick by which the quality of a component can be evaluated. The principal reliability model which comes closest to universal usage is that presented in the U.S. Department of Defense publication MIL-HDBK-217D [69]. The model is based on data covering 10^{10} device hours collected from accelerated lifetests, screening, burn-in and field experience [70].

Objections to the MIL-HDBK model are based on the variance between the calculated and observed failure rates. Blanks [71] and O'Connor [72] point out that the high values given to the quality factors are unjustified. These quality factors are assigned values depending on the level of screening used and are, "...intended to represent the risk associated with inadequate screening and the reliability enhancement which can be realized through an effective screening program." [70]. Hence, although a disparity is observed in failure rates, the estimates can be considered to be safe.

Reliability estimates are also made by accelerated testing of devices; i.e. prematurely inducing failures by exposing the devices to elevated temperature ambients. The failure rates are then extrapolated to standard operating conditions using the Arrhenius Equation [16][73]. One important requirement before accelerated testing, is to ensure that expected failure mechanisms are governed by this equation. Typical accelerated tests use temperatures between 75 °C and 225 °C, although lower or higher temperatures are used for specific mechanisms. The device may be operated during the test. Additionally, an ambient of humidity or pressure can be included [74]. Sample sizes vary according to expected failure rates. Low failure rates require large sample sizes to provide a high degree of confidence in the results.

Objections to the validity of accelerated test results are based on the point that the die surface may not be at the ambient temperature and humidity [75][76]. Stanley [75]
has also found that temperature excursions about the mean die temperature ranged from -16 °C to +17 °C. Herr et al [76] have shown that a device exposed for over a week to an ambient of 85% RH, shows only 12% RH at the die. However, if careful control is maintained over experimental conditions, valid estimates of device failure rates can be made from accelerated lifetests.

2.6. THE PHYSICS OF DEGRADATION MECHANISMS

A study of the physics of degradation processes and failure mechanisms enables a better understanding to be gained. From the work described in this chapter, four areas of research, significant to the next generation of semiconductor devices, can be identified. The results of the research, when applied to the development of such devices, would then mean that the commercial manufacture of these devices begins further along the learning curve discussed by Hutchins [68] and Brambilla [78].

The four areas are:

1) Electrical Overstress (EOS)/Electrostatic Discharge (ESD) damage. A high proportion of failures are attributed to this mechanism. Research would foster a better understanding of the effects of the fabrication techniques and device structures on the sensitivity of components to overstress effects.

2) Submicron Technology. Devices based on this technology are expected to form the backbone of the next generation of semiconductor devices. Studies into possible degradation mechanisms or the extent to which existing mechanisms would effect such devices will enable potential problem areas in the development to be anticipated.
3) GaAs. Integrated circuits using GaAs are still very much at the initial stages of development. As with submicron technology, studies on GaAs would aid this development.

4) Radiation. The effects of radiation become more important as device dimensions get smaller. Research in this area will help to improve the quality of radiation-hardened devices.

Of the above four areas, the effects of electrical overstress voltages, both continuous (simulating EOS) and pulsed (simulating ESD) has been investigated by the author. It has also been attempted to use these results to predict the effects of electrical overstress in submicron silicon MOSFETs.

2.7. REVIEW OF PUBLISHED WORK ON MOS BREAKDOWN

Electrical conduction and breakdown in solid dielectrics has been a subject of research since the mid 1920’s. Wagner [100] investigated dielectric breakdown related to weak spots in the insulator. Fowler and Nordheim in 1928 [79], proposed their theory of quantum tunnelling through a potential barrier at a metal-vacuum interface. This was adapted for dielectrics by Fröhlich in 1937 [80]. Frenkel [81] looked at the effect of electric fields on the conduction band in a dielectric or a semiconductor in the presence of impurity sites. Fröhlich later expanded on his work in the light of discoveries by other researchers, to develop an empirical model for the breakdown of ionic crystals based on avalanche multiplication by collision ionization [82][83]. Setz [84] correlated the breakdown field strength to the number of collisions required for breakdown. Whitehead [85] presents an extensive survey on existing work upto 1951 as well as expanding upon dielectric breakdown theories. O’Dwyer presented a modification of the simple avalanche model, to take into account the field distortion as a result of the space charge build up due to the avalanche [101]. His book [86] on dielectric
breakdown is a comprehensive work on the subject (upto 1972) covering experimental data and existing theories on breakdown.

Most of the early work on dielectric breakdown concentrated on ionic crystals. Klein [87] investigated the breakdown properties of $10^3\text{Å}$ to $10^4\text{Å}$ silicon oxide films on glass substrates using Al electrodes. The samples were subjected to continuous voltage stress and the analysis of breakdown fields was made on the principle of electronic avalanche in the dielectric. Further work by Klein [88], and work by Osburn and Ormond [89], and Osbourn and Weitzmann [91] have looked at dielectric breakdown, particularly, in amorphous silicon dioxide, in terms of the electron avalanche effect. Osburn et al observed different breakdown thresholds with different Si dopant concentrations, but do not qualitatively analyse this effect.

Silicon dioxide is a wide bandgap insulator. DiStefano and Shatzkes [92] present a theoretical breakdown model based on impact ionization. Their model considers the behaviour of the electrons only after they have entered the silicon dioxide. The influence of the electrodes on the injected charge has not been studied by them. However, the role of the silicon surface in the injection process cannot be neglected. Weinberg [93][94] shows that under high electric fields, the band-bending at the silicon surface leads to quantisation of charge in the Si conduction band. This, in turn, can effect the tunnelling current. Studies on surface quantisation have been conducted by Stern et al [95][96] and Nakamura et al [97], who have evaluated the separation between the quantised energy levels at high fields.

Oxide breakdown as a result of trap generation in silicon dioxide was first proposed by Harari [98] to explain the non-saturating behaviour of the applied voltage in constant current experiments. Others have observed similar behaviour [99] in the form of shifts in capacitance-voltage curves as a function of the injected charge. They have also
regarded the non-saturating behaviour as an indication of the generation of traps in the oxide layer. The generation of new electron traps under high-field stress has been inferred by Badihi et al [102] and Heyns et al [103]. Heyns observes a large number of electron traps at the p-Si-SiO₂ interface when the p-Si surface is in strong accumulation, while Badihi concludes that the injection of charge into the oxide at high fields is the cause of damage. Wolters et al [104] - [106] present a model for dielectric breakdown in MOS devices based on charge injection. They conclude that with continuous current injection (d.c. voltage stresses) the energy gained by the injected charge is not sufficient to support impact ionization. On the other hand, Thies et al [107] show that with the same constant electric fields (5MV/cm to 12 MV/cm) electrons have sufficient energy to tunnel into the SiO₂ conduction band. They calculate that this energy (3 to 4 eV above the Si conduction band) is sufficient to sustain impact ionization. Wolter's has based his model of the discharge pattern of the oxide breakdown on Budenstein's tree model of breakdown in insulators [108]. Budenstein's work, based on KBr single crystal dielectrics, Al₂O₃ and SiO₂, showed that the discharge patterns in the dielectrics were tree shaped. The trunk of the tree is at the anode, and the branches are formed by the movement of electrons, from traps distributed throughout the sample, towards the anode. Budenstein was able to observe and photograph the discharge patterns because his samples were very thick, ranging from 0.6 μm to 3.1 mm. Further evidence of trap generation and its precedence to breakdown is presented in a very recent paper (1986) by LeBlanc et al [109]. The effect of traps on the internal fields in the oxide has been quantitatively modelled by Chen et al [115]. Nissan-Cohen et al [110] have studied the effect of continuous charge injection on the physical properties of the SiO₂. They found that the trap generation rate is proportional to the flux of the injected charge and increases exponentially with the electric field in the oxide (between 4 and 10 MV/cm). Traps are believed to be formed by broken Si-O bonds. Olivo et al [111] have presented a model for electron trapping in thin SiO₂ films. The model is based on the presence of
defective Si-O bonds in the oxide region which can form traps, in constant electric fields.

Klein [112] discusses the difference between impact ionisation and charge injection models of breakdown in solid dielectrics. He reviews breakdown data taken from numerous sources regarding different types of dielectric materials, ranging from single crystal alkali halides to amorphous SiO$_2$.

It becomes apparent that the one constant parameter in dielectric breakdown experiments is the breakdown field strength. This is of the order of 5 - 10 MV/cm, irrespective of the material. Therefore, it seems that the breakdown strength is less affected by the material composition than by weaknesses which may exist in the dielectric. Cohen [113] has shown that the electric field strength of SiO$_2$ can be improved by considering the intrinsic defect density in the film. Balk et al [114] show a reduction in the density of electron traps after post-oxidation high temperature annealing. One may conclude that a completely defect-free oxide will have an electric field breakdown strength, which could be orders of magnitude higher than those obtained at present.

All the work reviewed in this section has dealt with the breakdown of dielectrics under the influence of constant applied voltages or constant charge injection. The breakdown voltages of the dielectrics used, have been between 5 - 10 MV/cm, thereby limiting the extent of the voltage stress that can be applied. No literature on pulsed voltage breakdown experiments was found with regard to MOS structures. None of the extensive works discussed here [106][110][112][115] refer to any work on very high voltage transients. Klein [112] mentions that experiments using pulsed voltages would help to develop a more definitive model of the oxide breakdown mechanism. In this thesis, the experimental work on pulsed voltages, has looked at the effects of electric
fields in excess of 100 MV/cm. A model has been developed for the breakdown mechanism, which draws on both the impact ionization and the charge injection theories.

2.8. SUMMARY

1. A description of the major failure mechanisms in semiconductor devices has been presented, with reference to published work on the subject. The author has dealt with the topic in greater detail in a book which is to be published later this year [116].

2. Methods of obtaining higher reliability have been discussed. It is believed that the focus should be on the policy of "getting-it-right-first-time", rather than attempting to screen out defective devices after the manufacturing process has been completed.

3. Reliability estimates obtained via modelling and accelerated test methods can only be looked upon as indicators of component lifetimes. In order to perform valid accelerated tests, it is necessary to conduct a detailed study of the failure mechanisms involved and the effects of accelerated stress factors on expected degradation mechanisms.

4. A study of the physics of failure processes is therefore seen as being important in the development of high reliability components. Four areas of research were proposed. This thesis investigates the effects of high electric fields on MOS devices.

5. A review has been made of published work on MOS oxide breakdown. It shows that models of breakdown have been based on, either, the theories of trap
generation in the SiO₂ or impact ionization. Experimental work on very high voltage pulses, with the intention of studying the physics of the breakdown process, has not been found. The work presented in this thesis intends to fulfill this void.

2.9. REFERENCES


20. The reliability of N-channel silicon-gate memory and microprocessor components; Report PR-1. - NATIONAL SEMICONDUCTOR CORPORATION, California, USA, 1982.


41. SINGER P.H., Techniques of wire bond testing. - Semicon. Int., 6, pp.76-80, July 1983.


99. JENG C., RANGANATH T., HUANG C., JONES J., CHANG T., High field generation of electron traps and charge trapping in ultra-thin SiO₂. - IEDM, Dig. of Papers, pp. 7-9, 1981.


103. HEYNS M.M., DEKEERSMACHER R.F., HILLEN M.W., Trap 
generation and electron detrapping in SiO₂ during high-field stressing of metal-

104. WOLTERS D.R., VAN DER SCHOOT J.J., Dielectric breakdown in 
MOS devices. Part I: Defect related and intrinsic breakdown. - Philips J. Res., 40, 

105. WOLTERS D.R., VAN DER SCHOOT J.J., Dielectric breakdown in 
MOS devices Part II: Conditions for the intrinsic breakdown - Philips J. Res., 40, 

106. WOLTERS D.R., VAN DER SCHOOT J.J., Dielectric breakdown in 

107. THEIS T.N., DiMARIA D.J., KIRTLERY J.R., DONG D.W., Strong 
electric field heating of conduction band electrons in SiO₂ - Phys. Rev. Lett., 52, 

108. BUDENSTEIN P.P., On the mechanism of dielectric breakdown of solids. - 

109. LEBLANC A.R., ABADEER W.W., Behaviour of SiO₂ under high 

110. NISSAN-COHEN Y., SHAPPIR J., FROHMAN-
BENTCHKOWSKY D., Trap generation and occupation dynamics in SiO₂ under 

111. OLIVO P., RICCO B., SANGIORGI E., Electron trapping/detrapping 
within thin SiO₂ films in the high tunnelling regime. - J. Appl. Phys., 54, pp. 5267-
5276, 1983.

112. KLEIN N., Mechanisms of electrical breakdown in thin insulators - an open 


CHAPTER 3

EXPERIMENTS

3.1. INTRODUCTION

Experimental work was focused on the influence of high electric fields on thin MOS gate oxides. Continuous and pulsed voltages were used, simulating EOS and ESD conditions respectively.

The devices were small dimension, discrete n-channel and p-channel MOS transistors with no input protection at the gate. The gate dimensions ranged from 1\mu m \times 1\mu m to 100\mu m \times 100\mu m. Separate gate oxide capacitors with the same structural properties as the transistors were also available on the same chips. These devices provided an ideal base for an investigation into the physics of the breakdown mechanisms of MOS devices.

3.2. THE MOS STRUCTURES

3.2.1. Introduction

The MOS devices were resident on two types of silicon wafers. Both wafer types were manufactured by Plessey Research (Caswell) Ltd., as part of their NMOS and CMOS small geometry process characterisation [2]. The use of process characterisation wafers enables the study to be based on possible production situations. This is an advantage over working with idealised wafers manufactured especially for the purposes of this research.
Figure 3.1 Cross section of transistor on wafer no: 1.

Figure 3.2 An E-Mode p-Si MOS Capacitor, indicating the p'-type Boron-implant at the Semiconductor Surface.
Fig. 3.3 - A map of a block of transistors on wafer no 1. The pattern is repeated about 100 times through the wafer.
FIGURE 3.4. A map of a block of transistors on wafer no: 2.
FIGURE 3.5. (a) Photograph of Wafer No: 1.

Diameter of wafer = 3 inches.

FIGURE 3.5. (b) Photograph of Wafer No: 2.

Diameter of wafer = 3 inches.
3.2.2. The NMOS process characterisation wafers (Wafer type No.1)

The 3-inch wafers consisted of both enhancement-mode (E-mode) and depletion-mode (D-mode) transistors of varying dimensions. A schematic cross-sectional view of an E-mode transistor is shown in Figure 3.1. MOS gate oxide capacitors were fabricated on the same chips as the transistors. A cross-sectional view of a capacitor simulating the gate oxide structure of an E-mode NMOS transistor is schematically represented in Figure 3.2.

The devices were manufactured on 3-inch p-type wafers, of resistivity 20 Ω-cm, associated with a substrate dopant concentration of 6.6 x 10^{14} cm^{-3} [3]. Preferential doping in the channel was obtained by ion implantation of boron to a density of 4 x 10^{15} cm^{-3} holes, for E-mode devices. D-mode devices had a further phosphorus implant to give a net charge density of 4 x 10^{15} cm^{-3} electrons. Source/drain doping was by arsenic implantation to a charge density of 2 x 10^{20} cm^{-3} electrons.

The gate oxide was thermally grown at 950 °C, to a thickness of 425 Å (42.5 nm), and is uniform throughout the wafer. The field oxide was 0.6μm thick.

Wafer passivation was provided by a 1.5μm thick layer of silicon glass. Electrical isolation was provided by the implant of boron ions.

The NMOS transistors were laid out as shown in Figure 3.3. Rows 1 to 11 shown here, represent 88 E-mode devices on the chip. Rows 12 to 22 repeated the pattern, to provide 176 E-mode transistors altogether. Rows 23 to 33 consisted of D-mode NMOS transistors. The layout of these transistors is identical to those in rows 1 to 11. In addition, one E-mode capacitor and one D-mode capacitor were on each chip.
The capacitors had a gate area of $49 \times 10^3 \, \mu m^2$ with a peripheral length of $910 \, \mu m$. n+ polysilicon gates were deposited using chemical vapour deposition (CVD) techniques and doped with arsenic ions to a density of approximately $1 \times 10^{21} \, cm^{-3}$.

The surface state density at the Si-SiO$_2$ interface was estimated by the manufacturers to be of the order of $1 \times 10^{11} \, cm^{-2}$ [4].

The mask described for this chip covered a surface area of $6 \, mm^2$ and was repeated about 100 times across each wafer. 4 wafers of this type were provided for the experimentation.

Appendix 1 lists the principal specifications of this wafer.

3.2.3. The CMOS process characterisation wafers (Wafer type No.2)

E-mode PMOS transistors were fabricated on a 3-inch wafer. The n-type substrate was doped to a concentration of $2 \times 10^{14} \, cm^{-3}$, associated with a resistivity of $25 \, \Omega \cdot cm$ [4].

The E-mode PMOS transistors consist of boron source/drain diffusions with dopant densities of $5 \times 10^{19} \, cm^{-3}$ holes. The channel was preferentially doped to a concentration of $1 \times 10^{16} \, cm^{-3}$ with n+ (As) ions. E-mode transistors were fabricated in a p-well which was doped to approximately $2 \times 10^{15} \, cm^{-3}$. The channel was preferentially doped using boron to a concentration of $1 \times 10^{16} \, cm^{-3}$.

The gate oxide thickness was uniform throughout the wafer at $320 \, \AA$ (32nm). The field oxide was 720 nm thick.
The PMOS transistors were laid out as shown in Figure 3.4. Each chip contained four rows of transistors with eight devices per row. NMOS transistors on the same chip had the identical layout. In addition, one E-mode NMOS capacitor (p+-Si surface) and one E-mode PMOS capacitor (n+-Si surface) were available on each chip. The capacitors had gate areas of $4.9 \times 10^3 \mu m^2$ with a peripheral length of 280 $\mu m$. Schematic cross-sectional views of the transistors and capacitors are the same as those presented in Figures 3.1. and 3.2.

All devices had n+ polysilicon gates doped to approximately $1 \times 10^{21} \text{cm}^{-3}$. The surface state density at the Si-SiO$_2$ interface was estimated to be of the order of $1 \times 10^{11} \text{cm}^{-2}$ [4].

The mask described for this chip covered a surface area of 6 mm$^2$ and was repeated about 100 times across the wafer. Two wafers of this type were provided for experiments.

Appendix 1 lists the principal specifications of this wafer.

3.2.4. Past history of the samples

The wafers had not been subjected to any post-manufacture lifetesting before being used in the experiments. Electrical functional tests, however, had been carried out to evaluate the process employed in fabrication. These tests measured the threshold voltages, gain and the drain-source currents under specified bias conditions. The electrical dimensions of the channel were calculated, giving values for the change in channel length ($\Delta L$) and the change in channel width ($\Delta W$) caused by the fabrication process. $\Delta L$ was the result of lateral diffusion of source/drain dopants, while $\Delta W$ was due to undercutting of the field oxide.
The tests were non-destructive and did not stress the devices in any way.

Figures 3.5 a) and 3.5 b) are photographs of wafer type no.1 and wafer type no.2, respectively.

3.3. APPARATUS

3.3.1. Introduction

A schematic diagram of the apparatus used in the pulsed voltage experiments is shown in Figure 3.6. Figure 3.7. shows a schematic of the apparatus used in the continuous voltage experiments.

3.3.2. The wafer microprober

The wafer microprober is shown in photographs in Figure 3.8.. The wafer was placed on a smooth metal chuck and held in place by vacuum suction. The chuck could be electrically grounded if required. Contact with the devices was made using a set of adjustable probes with 20μm diameter tips. Electrical connections to and from the probes were made using low-capacitance screened leads. A stereo optical microscope incorporated in the microprober enabled visual analysis to be made. The magnification capabilities of the microscope range from 50X to 3000X. Illumination was available through the lens itself and could be turned-off when not required.

In addition to the chuck incorporated in the microprober, a heater chuck was built. This chuck used the heat generated by four 1 Ω resistors, each capable of dissipating 4W. A current of 1A was provided to these resistors using the mains voltage supply.
Figure 3.6 Schematic of the apparatus used in the pulsed voltage experiments.

Figure 3.7. Diagram of the circuit used in continuous voltage stress investigations.
FIGURE 3.8: Photographs of the wafer microprober
R1 = 1 M Ω
R2 = 1.5 k Ω
C1 = 100 pF
S = Two-pole relay
DUT = Device Under Test

FIGURE 3.19  Circuit diagram of the pulse generator (after MIL-STD-883 C).

FIGURE 3.19. Schematic illustration of the fast voltage pulse
Fig 3.11 - Form of the continuous voltage stress as applied to MOS structures.
and a step-down transformer. Temperature control was achieved through a thermocouple and a ready-made thermostat. The thermostat was calibrated for the heated chuck using a Noronix digital thermometer.

3.3.3. The pulsed voltage generator

Voltage pulses were generated using the circuit shown in Figure 3.9. This circuit is based on the "human body" model pulse generator described by MIL-STD-883C [1].

A 100 pF capacitor was charged to the required voltage from a d.c. power supply. It was then discharged through a 1.5 kΩ resistor into the device under test (DUT). A two-pole mercury-wetted relay was used to perform the change-over. This relay had negligible contact bounce which was essential in the discharge circuit. All the components were required to operate at high voltages (upto 1 kV). Hence the capacitor (C1) and resistor (R2) were designed and constructed in the laboratory using thick-film techniques. The actual capacitance and resistance values measured for C1 and R2 were 101.7 pF and 1.55 kΩ, respectively, (at a measuring frequency of 1 kHz on the LCR meter described in section 3.4.5.). R2 has a parallel capacitance of less than $10^{-14}$ pF and a series inductance of less than $10^{-9}$ H. The parallel resistance of C2 was in excess of $10^{15}$ Ω.

The pulse shape obtained is shown in Figure 3.10, for a discharge into a low capacitance oscilloscope probe. The resistance of the probe was 10 MΩ, while the capacitance was 2 pF. The decay time of the pulse was governed by the impedance of the probe. The risetime obtained is shown in the figure as being less than 500 ns. More detailed analysis of the pulse risetime has shown it to be less than 100 ns. Limitations in accurately defining the risetime, were due to the inadequacies of the available oscilloscopes.
3.3.4. Continuous voltage generation

Continuous voltages were generated using a Keithley 230 Programmable Voltage Source. The instrument was programmed to generate a d.c. voltage for a fixed time period ($\Delta t$) after which the voltage was raised by a fixed step value ($\Delta V$). A ramped voltage as shown in Figure 3.11 was obtained. The voltage source had a built-in current limiter which was set to 2 mA, eliminating the need to provide short-circuit protection.

3.4. MEASUREMENTS

3.4.1. Introduction

Transistor properties were measured using a curve-tracer (C-T) to monitor the current-voltage characteristics of the devices. Capacitance and resistance properties were measured using a Wayne-Kerr 4210 LCR meter. I-V characteristics were also studied using the Keithley 230 Voltage Source in circuit with the Keithley 617 Programmable Electrometer. By individually plotting the I-V characteristics of selected transistors, it was possible to confirm that the C-T technique and the manual techniques gave the same results.

3.4.2. The curve-tracer

A Telequipment CT71 curve-tracer was used. Drain-source current ($I_{DS}$) measurements could be made down to 10 $\mu$A/cm with 12 gate-source voltage ($V_{GS}$) steps. Step amplitudes ranged from $\pm 0.1$ V to $\pm 2$ V. Drain-source voltages ($V_{DS}$) down to a minimum of 0.1 V/cm, in both positive and negative directions, could be
made. An output power limiter was provided to protect sensitive devices from being accidentally overstressed.

3.4.3. Voltage measurements

The Keithley 230 Voltage Source had a digital readout accurate to 4 decimal places at ±0.05%. For other voltage measurements the Keithley 617 Electrometer was used. This instrument had a built-in voltage source with an output ranging from -100 V to +100 V in 50 mV steps. Again, a digital readout indicated voltage magnitudes.

The electrometer had a voltage measuring capability from 10μV to 200 V on four voltage ranges, with an accuracy of ±0.05%.

3.4.4. Current measurements

The Keithley 617 electrometer was used for all current measurements. Resolution ranged from 10⁻¹⁵ A to 20 mA over 11 voltage ranges. An accuracy of ±0.25% is quoted for the 2 nA range.

Although a 10⁻¹⁵ A resolution is claimed, measurements below 10⁻¹² A levels were not considered reliable because of problems with background noise. The instrument included a "suppress" facility which enabled any steady background noise to be zeroed before measurements were taken. This facility was also used to eliminate any self-generated currents caused by flexing of the coaxial cables.

3.4.5. The Wayne-Kerr LCR meter
Resistance and capacitance measurements were made using the Wayne-Kerr 4210 LCR meter. Resistance could be measured down to 0.01 Ω, with an accuracy of ±1.0%. The maximum resistance which could be measured was $10^9$ Ω. Capacitance could be measured down to 0.1 pF with a quoted accuracy of ±1.0%.

The LCR meter also include a "trim" facility, similar to the "suppress" facility of the electrometer. This enabled any stray resistances or capacitances to be eliminated from the measurements. All experimental readings were taken at the 1 kHz measurement frequency at a voltage level of 250 mV.

3.4.6. Temperature

A Noronix NTD30 electronic thermometer with an accuracy of ±1 °C was used.

3.4.7. Capacitance-voltage (C-V) measurements

A useful parameter of an MOS capacitor is the variation of capacitance with applied gate voltage. Measurements, in these experiments, were focussed on the relative shift in the C-V curve caused by applied voltage stress. A simple C-V meter, capable of fulfilling these functions, was designed and constructed.

A circuit diagram of the C-V meter is shown in Figure 3.12. The core of the circuit was a Burr-Brown 3430 Electrometer Amplifier [5]. The amplifier had an open loop gain of 100 dB and an input bias current of ±0.01 pA. The input impedance of the device was $10^{14}$ Ω. Supply voltages of +15V and -15V were required. A 100 kΩ variable resistor between V+ and V-, connected to the trim input of the amplifier, enabled the offset to be adjusted. The warm-up drift for the 3430 was specified as 75 μV over 15 minutes, and needed to be offset for very small voltage measurements.
FIGURE 3.12 - Circuit diagram of the Capacitance-Voltage meter.

FIGURE 3.13 - Resistor network used in C-V meter.
resistor \( R \) was required to be of the order of \( 10^{12} \) \( \Omega \) to obtain output voltages of the order of 1 V, with input currents of about \( 10^{-12} \) A. Metal film resistors have the best temperature stability and are accurate to 0.1\%, but were only available in values up to \( 10^7 \) \( \Omega \). Carbon resistors were available in values up to \( 10^{14} \) \( \Omega \) but have relatively large temperature coefficients (1000 to 5000 ppm/°C compared with 100 ppm/°C) and tolerances of 1\% to 10\%. They also have comparatively poor long term stability. The problem was overcome by using a resistor network for \( R \), as shown in Figure 3.13. \( R_1 \) and \( R_2 \) were metal film resistors where \( R_1 = 100 \) M\( \Omega \), and \( R_2 = 10 \) M\( \Omega \). \( R_3 \) was a 10k\( \Omega \) carbon resistor. The overall resistance is given by,

\[
R = R_1 \left[ 1 + \frac{R_2}{R_3} + \frac{R_2}{R_1} \right] = 10^{11} \Omega
\]

Unfortunately, \( R_2 \) and \( R_3 \) amplified the voltage noise and increased the offset by the factor \( \{(R_2 + R_3)/R_3\} = 10^3 \). Also, the loop gain of the circuit was reduced by the factor \( (R_3/(R_2 + R_3)) = 10^{-3} \). However, the gain was sufficiently high to be able to tolerate these factors.

1 mF decoupling capacitors were used at the positive and negative supply rails.

A Burr-Brown 2800 MC connector was used to mount the amplifier because of the high insulating properties of the teflon insulate pin-jacks. Wiring in the circuit was kept rigid and as short as possible to prevent extraneous voltage generation or stray capacitance. Shielded coaxial cables were used for the input and output connections. In order to minimise self-generated cable noise, a special low-noise cable was used. Shielding for the circuit was provided by encasing it in a grounded aluminium box.
The MOS gate oxide capacitor was supplied with a slowly varying ramped input signal of the order of $10^{-2}$ Hz. The current through the capacitor was converted to a voltage and amplified by the electrometer. The output from the electrometer was fed into the Y-axis of an X-Y plotter. The X-axis of the plotter was fed directly from the low frequency signal source. Since the capacitance was directly proportional to the current flowing through the MOS capacitor, a C-V plot was obtained.

Tests on the C-V meter showed that relative shifts and variations in C-V curves could be satisfactorily observed. However, because of the input noise levels and the gain factors of both the electrometer and the X-Y plotter, absolute values of capacitance could not be determined.

### 3.5 EXPERIMENTAL PROCEDURE

#### 3.5.1 Tests on the wafers

##### 3.5.1.1 Homogeneity

The wafers were tested for variations between device properties from chip-to-chip and wafer-to-wafer. Transistor I-V characteristics of two sets of devices on different chips on a wafer were compared. The chips were selected at different geographical locations on the wafer. Comparisons were then made between these chips and chips on another wafer. The gate leakage currents of MOS gate oxide capacitors were also compared between the devices on these same chips.

Possible variations in oxide breakdown strengths, with respect to the position of the chip on the wafer and between wafers, were investigated. The on-chip gate oxide capacitors were used in these tests.
The homogeneity tests showed that the variation of electrical characteristics between identical devices was not more than 5%. The oxide breakdown strength was found to be consistent from chip-to-chip and between wafers.

It was concluded that the wafers in each process characterisation type were homogeneous.

3.5.1.2. The effects of heat on the device properties

Wafers were tested to investigate the effect of temperature on the devices. The I-V characteristics of devices situated at different positions on the wafer were recorded. The wafer was then heated to 70°C, 110°C, 150°C and 200°C and maintained at each temperature for approximately 15 minutes. After each temperature stress, the device characteristics were recorded at room temperature. Comparisons with the I-V characteristics made before the wafer was heated, showed no changes.

This set of tests also showed that repeated heating and cooling did not affect the electrical properties of the devices.

The oxide breakdown strengths of capacitors which had only been subjected to thermal stress, were unaffected.

3.5.1.3. High temperature reverse bias (HTRB) screening

A voltage of -12 V was applied to the gates of the devices at a temperature of 150°C, for approximately 5 minutes. This had the effect of sweeping ionic impurities (e.g. Na⁺, K⁺) out of the oxide.
The screen did not affect the I-V characteristics of the transistors or the R and C properties of the capacitors. However, capacitors used in both the pulsed and the continuous voltage experiments were HTRB screened in order to minimise any effects due to ionic impurities.

3.5.2. Pulsed voltage stress experiments on E-mode NMOS transistors

3.5.2.1. Preliminary experiments

A number of E-mode NMOS transistors on wafer type no.1, were subjected to single voltage pulses ranging from 50 V to 3 kV. The purpose of this experiment was to determine a suitable voltage range for a study of the oxide breakdown mechanisms under pulsed conditions. The ideal voltage range was considered to be one in which a steady increase in catastrophic failures would be observed as the applied voltage was increased, without incurring 100% failures. The pulses were applied to the gate, while the substrate was grounded.

Voltages ranging between 50 V and 300 V were found to be most suitable for the investigation. Beyond 300 V, more devices showed catastrophic failures. Almost all the devices tested were found to have been destroyed after application of a pulse of 1 kV.

The I-V characteristics of all the devices to be used in the experiments were then recorded. It was necessary to ensure that devices damaged accidentally, either in the measuring process or by spurious voltage spikes (ESD or otherwise), were identified. Therefore, any devices which showed $I_{DS}$ vs. $V_{DS}$ characteristics which
differed by more than 5% from that of known good device were excluded from the experiments.

3.5.2.2. Temperature range

The selected temperature range was required to encompass the standard maximum operating temperature for commercial devices (70°C) and military devices (125°C). Additional measurements at room temperature acted as a control group, while high temperature effects were monitored with a group at 200°C. The five selected temperatures were therefore, 25°C, 70°C, 110°C, 150°C and 200°C.

3.5.2.3. The number of devices

Each chip consists of devices of assorted dimensions and only 2 devices on any one chip had the same dimensions. Therefore, it was decided to subject all devices on a chip to the same voltage pulse for a given temperature. Hence, at 70°C, 176 devices were subjected to 50 V, 176 to 100 V and so on. The number of devices used in the experiment was 176 x 5 x 5 for the 5 voltages and the 5 temperatures. This amounted to a total of approximately 4400 devices.

3.5.2.4. Experimental method

1. The characteristics of all the transistors on 25 chips were recorded.
2. The wafer (No.1) was brought to room temperature (25°C).
3. All devices on a single chip were subjected to a single pulse of 50 V.
4. The experiment was repeated at 100 V, 150 V, 200 V and 250 V, on four different chips, all on the same wafer.
5. Steps 3 and 4 were repeated at 70°C, 110°C, 150°C and 200°C on different chips.
6. The wafer was returned to room temperature.

7. The device I-V characteristics were again recorded, emphasis being laid on those devices which showed deviations from their original characteristics.

3.5.2.5. General

The experiments covered both the voltage and temperature sensitivity of E-mode NMOS transistors to high voltage pulse conditions.

3.5.3. Pulsed voltage stress experiments on D-mode NMOS transistors

Due to the large amount of data collected from the experiments on E-mode NMOS transistors, the experiments on D-mode devices could be limited to the effects of high voltage pulses.

28 devices were used in this experiment. They were subjected to single pulses of 50 V, 100 V, 150 V and 250 V. The 200 V stress was omitted after preliminary experiments, on devices selected at random, indicated the trend of the voltage dependence.

3.5.4. Pulsed voltage stress experiments on E-mode PMOS transistors

Investigations into the pulsed voltage sensitivity of PMOS transistors also drew on the data accumulated in E-mode NMOS transistor experiments. As with the D-mode NMOS devices, experiments were limited to the voltage effects. A total of 42 transistors of various dimensions were used. All the devices were on the CMOS process characterisation wafer (type No.2).
Preliminary experiments on devices selected at random, showed that degradation occurred with single pulses ranging between 50 V and 1 kV. In order to investigate in more detail the effect of the voltage pulse on I-V characteristics, one group of 21 devices was subjected to 250 V, and a second group of 21 devices was subjected to 350 V pulses.

3.5.5. **Pulsed voltage stress experiments on p-Si capacitors**

p-Si (Enhancement-type) MOS capacitors on the NMOS process characterisation wafers, were subjected to single positive and negative voltage pulses. Only one capacitor was available on each chip. This meant that the numbers available for experimentation were limited.

The capacitance and resistance of each MOS capacitor used in the experiment was recorded, using the LCR meter. Groups of 10 devices were subjected to single positive voltage pulses of 75 V, 150 V, 175 V and 200 V. Their capacitances and resistances were then remeasured.

Another 3 groups, each containing 10 devices, were subjected to single negative polarity pulses of -50 V, -75 V and -100 V.

C-V curves of the devices were made after application of the pulses.

Capacitors which had not been HTRB screened were used for the bulk of the experimentation. However, a small group of 30 capacitors which had been HTRB screened were subjected to positive pulses.
3.5.6. **Pulsed voltage stress experiments on n-Si capacitors**

n-Si (D-type) MOS capacitors from wafer type No.1, were subjected to single voltage pulses of both positive and negative polarities. The pattern of the experiments was similar to that for the p-Si capacitors. All the capacitors were on the same wafer; none of them had been subjected to HTRB screens.

3.5.7. **Pulsed voltage stress experiments on n+-Si capacitors**

n+-Si (E-type) MOS capacitors from the CMOS process characterisation wafers, were subjected to single, positive and negative voltage pulses. The pattern of the experiments was similar to that described above for the other capacitor structures. However, the voltage range was different. Positive pulses were at +30 V, +50 V and +70 V; negative pulses were at -30 V, -50 V and -60 V. The capacitors were selected from both available wafers of this type, and had not been HTRB screened.

3.5.8. **Sequential pulsing of E-mode NMOS transistors**

168 transistors from the wafer type No.1, were subjected to pulse sequences. The pulses were applied at about 60 second intervals, by manually operating the relay control switch of the pulse generator.

The first experiment investigated the effect of increasing the number of pulses at different stress voltages. Groups of 24 devices were each subjected to sequences of 25, 50 and 100 pulses. In each group, 8 devices were subjected to 50 V pulses, 8 devices to 150 V pulses, and 8 devices to 250 V pulses. The device I-V characteristics were recorded at the end of each complete sequence. The group subjected to 25
pulses, had device characteristics monitored after 8 pulses, 10 pulses, 15 pulses and 20 pulses in order to observe trends.

A second experiment was carried on 4 groups of 24 transistors. These devices were subjected to a constant number of pulses, while the applied voltage was varied from group to group. 25 pulses were applied at 350 V, 525 V, 700 V and 1 kV. The device I-V characteristics were recorded at the end of each sequence.

3.5.9. Sequential pulsing of p-Si capacitors

2 groups of 10 p-Si capacitors from wafer type no.1, were subjected to pulse sequences. All the capacitors had been subjected to HTRB screens. The purpose of the experiment was to investigate the variations in capacitance and resistance after each set of pulses. One group of devices was stressed at 150 V and a second group at 200 V. The capacitance and resistance values were measured before application of the pulse. A single pulse was applied and the C and R values were measured again. After that, C and R values were measured after each set 5 pulses.

The group stressed at 150 V was subjected to 3 sets of 5 pulses, totalling 15 pulses in all. The group at 200 V was subjected to 5 sets of 5 pulses, totalling 25 pulses.

3.5.10. Continuous voltage stress experiments on p-Si capacitors

3.5.10.1. Temperature sensitivity of the oxide breakdown voltages

The oxide breakdown threshold as a function of the applied temperature, was investigated at 25°C, 75°C, 125°C, 150°C and 200°C, under continuous voltage conditions.
25 capacitors from wafer type No.1, were used in this experiment, in 5 groups of 5. Each group was subjected to a d.c. voltage, which was gradually increased from zero volts until breakdown, at a specific temperature. The voltages were raised in 1 V steps of 180 seconds duration each. Breakdown was defined at the voltage at which an extremely rapid increase in leakage current was observed. This was confirmed by C and R measurements made after completion of the experiment.

All the capacitors had been HTRB screened.

3.5.10.2. Gate leakage current experiments

2 groups of 5 devices from wafer type No.1, were subjected to both positive and negative polarity voltage stress. The gate leakage current was measured using the Keithley 617 Electrometer. One group of devices was stressed from zero volts to the positive voltage breakdown threshold in steps of 1 V. Leakage current measurements were made at each voltage step. The second group of devices was stressed from zero volts to the negative breakdown threshold in steps of 1 V. Leakage current measurements were made as before.

The C-V curves of devices subjected to constant voltage stress in the vicinity of the breakdown threshold were recorded. The curves were made after application of continuous voltages of 28 V to 50 V in the positive direction, and 26 V to 38 V in the negative direction. The stress levels were at 2 V intervals (e.g. 26 V, 28 V, 30 V, 32 V etc.) on different capacitors.

All the capacitors had been subjected to HTRB screens.
3.5.11. Continuous voltage stress experiments on n-Si capacitors

2 groups of 5 devices from wafer type No.1, were subjected to constant applied voltages. The two groups were stressed from zero volts to the positive and negative breakdown thresholds as described in the previous section. All devices had been HTRB screened.

3.5.12. Continuous stress experiments on n+-Si capacitors

The experiments used 2 groups of 5 devices each from wafer type No.2. All devices had been HTRB screened. The devices were stressed from zero volts in both the positive and negative directions as described in section 3.5.10.

3.6. SUMMARY

1. Two wafer types were used in these experiments. One wafer type characterised an NMOS fabrication process, and the other characterised a CMOS fabrication process. Each wafer contained MOS transistors and MOS gate oxide capacitors.

2. The wafers were tested for consistency and homogeneity from chip-to-chip and wafer-to-wafer.

3. The apparatus and the experimental procedure has been described. Table 3.1 summarises the experiments, and the associated sections in this chapter.
Table 3.1

Summary of experiments described in this chapter. V indicates experiments on voltage dependence, T indicates experiments on temperature dependence.

<table>
<thead>
<tr>
<th>Device Type</th>
<th>Single Pulsed Voltages</th>
<th>Sequential Pulsed Voltages</th>
<th>Continuous Voltages</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>V</td>
<td>T</td>
<td>V</td>
</tr>
<tr>
<td>E-mode NMOST</td>
<td>3.5.2</td>
<td>3.5.2</td>
<td>3.5.8</td>
</tr>
<tr>
<td>D-mode NMOST</td>
<td>3.5.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-mode PMOST</td>
<td>3.5.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p-Si Capacitors</td>
<td>3.5.5</td>
<td>3.5.9</td>
<td>3.5.10.2</td>
</tr>
<tr>
<td>n-Si Capacitors</td>
<td>3.5.6</td>
<td></td>
<td>3.5.11</td>
</tr>
<tr>
<td>n+-Si Capacitors</td>
<td>3.5.7</td>
<td></td>
<td>3.5.12</td>
</tr>
</tbody>
</table>

3.7. REFERENCES


CHAPTER 4

RESULTS

4.1. INTRODUCTION

The three main sections in this chapter present the results of the experiments conducted. Single pulse conditions, pulse sequences and continuous voltage conditions make up the main sections. Experiments on different device structures are described in subsections.

4.2 PULSED VOLTAGE STRESS EXPERIMENTS

4.2.1 E-mode NMOS transistors

4.2.1.1. Characteristics

A comparison was made of the $I_{DS}$ vs. $V_{DS}$ characteristics before and after application of a single high voltage pulse. It was determined that these characteristics could be classified into one of the following five categories.

i) Category 1 - no change. All parameters remain unaffected. The transistor characteristics shown in Figure 4.1 was observed.

ii) Category 2 - degradation. The characteristics of a degraded device with the same dimensions as the one used in Figure 4.1., is shown in Figure 4.2.. A comparison between the two characteristics shows that significant degradation has
**FIGURE 4.1** Category 1, 'No Change' $I_{DS}$ VS $V_{DS}$ characteristic

**FIGURE 4.2** Category 2, 'Degradation'

**FIGURE 4.3** Category 3, Degradation with negative $gm$ at high $V_{gs}$. 
Figure 4.4: Category 4, Resistive characteristics

Figure 4.5: Category 5, Catastrophic Failure
FIG 4.6  VOLTAGE DEPENDENCE OF DEVICE CHARACTERISTICS

FIG 4.7  TEMPERATURE DEPENDENCE OF DEVICE CHARACTERISTICS

FIG 4.8  DIMENSIONAL DEPENDENCE OF 'NO CHANGE' CATEGORY, w=102.9 \( \mu \text{m} \)
occurred. Measurements of the gate-drain resistance \( R_{GD} = 65 \text{ kΩ} \) and the gate-source resistance \( R_{GS} = 56 \text{ kΩ} \), indicate a small leakage current through the oxide. Both the saturation and the non-saturation values of the drain-source conductance, \( g_{ds(sat)} \) and \( g_{ds(non-sat)} \) have decreased. The extent of these changes may vary from a few % to almost 75%.

iii) Category 3 - degradation with negative \( g_m \) at high \( V_{GS} \). The characteristics shown in Figure 4.3. exhibit a degradation with a drain-source current which begins to decrease as \( V_{GS} \) is increased beyond 3 V.

iv) Category 4 - resistive characteristics. These characteristics as shown in Figure 4.4., are another form of degradation. When the drain and source contacts are interchanged, the characteristics resume degradation type effects as in Category 2 (Figure 4.2.). \( R_{GD} = 0.31 \text{ kΩ} \) and is very much lower than \( R_{GS} = 16.3 \text{ kΩ} \). The gate-drain current, \( I_{GD} \), measured with the gate at +4 V and the drain at 0 V, was 12.5 mA. The gate-source current, \( I \), under similar conditions was 1.0 mA.

v) Category 5 - catastrophic failure. The characteristics are shown in Figure 4.5.. \( I_{DS} \) is approximately zero, for all values of \( V_{GS} \), and for all \( V_{DS} \), up to avalanche breakdown.

4.2.1.2. Device sensitivity to the magnitude of the pulsed voltage

Figure 4.6. shows the percentage distributions of devices showing "no change", "degradation" and "catastrophic failure" characteristics, as a function of the applied voltage. It is seen, that at 50 V, over 70% of the devices were not affected by the stress. This number steadily decreased, until at 250 V, only about 7% of the devices showed no changes. The degradation characteristics increases over the same range,
from just over 20% at 50 V, to about 80% at 250 V. The percentage number of devices exhibiting catastrophic failures increase very gradually, from 5% at 50 V, to around 12% at 250 V.

4.2.1.3. Temperature sensitivity of pulsed voltage damage

The results of the temperature experiments on E-mode NMOS transistors are summarised in Figure 4.7. The curves indicate the percentage number of devices showing damage as a function of temperature.

4.2.1.4. The affect of the transistor gate dimensions on the degradation characteristics

The E-mode transistors were of many different dimensions. It was, therefore, possible to make an analysis of the relationship between the gate size and the sensitivity of the devices to the stress. However, at very small dimensions (< 5 μm), factors extraneous to the size of the gate may influence the results. Lateral diffusion of the drain/source and field oxide undercutting of the gate area, will affect the electrical properties of the device. These effects are expected to be concentrated at the gate-drain and the gate-source regions. Therefore, in Figure 4.8., the percentage number of devices showing "no change" is plotted as a function of channel length for a specific gate width. The gate width of 102.9 μm, and the gate lengths of 6.3, 10.2, 20.7, and 102.9 μm, ensure that the degradation trends observed are solely due to the effects of the gate area. The plots show distributions for single pulses of 150 V and 200 V.

4.2.2. D-mode NMOS transistors
The characteristics of an undamaged device are shown in Figures 4.9. and 4.10. for $V_{GS}>0$ and $V_{GS}<0$ respectively. The characteristics observed after application of a single 100 V pulse are shown in Figures 4.11 and 4.12 for $V_{GS}>0$ and $V_{GS}<0$ respectively.

The characteristics for $V_{GS}>0$ show similar degradation to those exhibited by E-mode NMOS devices. A difference does exist in $g_{ds(sat)}$ which shows an increase, compared to the decrease observed in damaged E-mode transistors.

For $V_{GS}<0$, the characteristics are quite different. As $V_{GS}$ becomes more negative, it is seen that $I_{DS}$ first decreases for a given $V_{DS}$ as for an undamaged device, but an increase in channel conductance ($g_{ds}$) is observed. As $g_{ds}$ gets more negative, $I_{DS}$ increases at low values of $V_{DS}$ ($<2$ V). At higher $V_{DS}$ values, between 2 V and 10 V, $I_{DS}$ remains almost constant.

4.2.3. E-mode PMOS transistors

The characteristics of an undamaged device is shown in Figure 4.13. The characteristics obtained after application of a 350 V pulse is shown in Figure 4.14. The observed degradation is identical to that shown by E-mode NMOS transistors.

4.2.4. MOS capacitor structures

The results of the experiments on the three capacitor structures are summarised in Table 4.1. These devices had not been subjected to HTRB screens.

The table shows the magnitude of the applied pulse voltage, and typical values of $R$ and $C$, before and after stress was applied. $R$ and $C$ values were measured at 1 kHz
FIGURE 4.9 I-V characteristics of a D-mode transistor before application of an ESD pulse, $V_{GS} > 0$.

FIGURE 4.10 I-V characteristics of a D-mode NMOS transistor before application of an ESD pulse, $V_{GS} < 0$.

FIGURE 4.11 I-V characteristics of a D-mode NMOS transistor after application of an ESD pulse, $V_{GS} > 0$.

FIGURE 4.12 I-V characteristics of a D-mode NMOS transistor after application of an ESD pulse, $V_{GS} < 0$. 
FIGURE 4.13. I-V characteristics of an E-mode PMOS transistor before application of an ESD pulse.

FIGURE 4.14. I-V characteristics of an E-mode PMOS transistor after application of an ESD pulse @ 350V.
RESULTS OF ESD EXPERIMENTS USING POSITIVE AND NEGATIVE VOLTAGE PULSES ON E-TYPE AND D-TYPE NMOS AND E-TYPE DMOS CAPACITORS (NO HTRB SCREENING)

<table>
<thead>
<tr>
<th>CAPACITOR TYPE</th>
<th>NO: OF DEVICES</th>
<th>APPLIED VOLTAGE (V)</th>
<th>TYPICAL CAPACITANCE (pF)</th>
<th>TYPICAL RESISTANCE (kΩ)</th>
<th>COMMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>before</td>
<td>after</td>
<td>before</td>
<td>after</td>
</tr>
<tr>
<td>E-TYPE NMOS</td>
<td>10</td>
<td>+150</td>
<td>20.0</td>
<td>20.0</td>
<td>O/C</td>
</tr>
<tr>
<td>E-TYPE NMOS</td>
<td>10</td>
<td>+200</td>
<td>20.0</td>
<td>16.5</td>
<td>O/C</td>
</tr>
<tr>
<td>E-TYPE NMOS</td>
<td>10</td>
<td>−75V</td>
<td>20.0</td>
<td>20.0</td>
<td>O/C</td>
</tr>
<tr>
<td>E-TYPE NMOS</td>
<td>10</td>
<td>−100V</td>
<td>20.0</td>
<td>17.0</td>
<td>O/C</td>
</tr>
<tr>
<td>D-TYPE NMOS</td>
<td>10</td>
<td>+150</td>
<td>13.0</td>
<td>13.0</td>
<td>O/C</td>
</tr>
<tr>
<td>D-TYPE NMOS</td>
<td>10</td>
<td>+175</td>
<td>13.0</td>
<td>9.0</td>
<td>O/C</td>
</tr>
<tr>
<td>D-TYPE NMOS</td>
<td>10</td>
<td>−75V</td>
<td>13.0</td>
<td>13.0</td>
<td>O/C</td>
</tr>
<tr>
<td>D-TYPE NMOS</td>
<td>10</td>
<td>−100V</td>
<td>13.0</td>
<td>9.0</td>
<td>O/C</td>
</tr>
<tr>
<td>E-TYPE PMOS</td>
<td>10</td>
<td>+50V</td>
<td>11.5</td>
<td>11.5</td>
<td>O/C</td>
</tr>
<tr>
<td>E-TYPE PMOS</td>
<td>10</td>
<td>+75V</td>
<td>11.5</td>
<td>7.0</td>
<td>O/C</td>
</tr>
<tr>
<td>E-TYPE PMOS</td>
<td>10</td>
<td>−50V</td>
<td>11.0</td>
<td>11.0</td>
<td>O/C</td>
</tr>
<tr>
<td>E-TYPE PMOS</td>
<td>10</td>
<td>−60V</td>
<td>11.5</td>
<td>6.0</td>
<td>O/C</td>
</tr>
</tbody>
</table>

TABLE 4.1
FIGURE 4.15. C-V curve on E-mode p-Si MOS capacitor, before application of an ESD pulse.

FIGURE 4.16. C-V curve of E-mode p-Si MOS capacitor, after application of a 150 V ESD pulse.
Figure 4.13 C-V Curve of an Undamaged E-Mode n-Si MOS Capacitor

Figure 4.18 C-V Curve of an E-Mode n-Si Capacitor after application of an ESD Pulse of +50V
ABSOLUTE NUMBER OF DEVICES SHOWING CHARACTERISTICS DEPICTED,
AS A FUNCTION OF THE NUMBER OF APPLIED ESD PULSES,
AND THE APPLIED ESD VOLTAGE

<table>
<thead>
<tr>
<th>VOLTS</th>
<th>PULSES</th>
<th>CATASTROPHIC FAILURES</th>
<th>REDUCED CHARACTERISTIC</th>
<th>NO CHANGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>25</td>
<td>0</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>150</td>
<td>25</td>
<td>1</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>1</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>200</td>
<td>25</td>
<td>1</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>1</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>7</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

TABLE 2

THE EFFECT OF INCREASING THE APPLIED ESD VOLTAGE AT A CONSTANT 25 PULSES
ABSOLUTE NUMBER OF DEVICES.
EXHIBITING THE GIVEN CHARACTERISTICS ARE SHOWN

<table>
<thead>
<tr>
<th>VOLTS</th>
<th>CATASTROPHIC FAILURES</th>
<th>LINEAR I-V CHARACTERISTICS</th>
<th>REDUCED CHARACTERISTICS</th>
</tr>
</thead>
<tbody>
<tr>
<td>350</td>
<td>10</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>525</td>
<td>16</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>720</td>
<td>16</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>1 kV</td>
<td>24</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

TABLE 4.3.
<table>
<thead>
<tr>
<th>No. of Devices</th>
<th>Pulsed Voltage (V)</th>
<th>Resistance (kΩ)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>150</td>
<td>o/c</td>
</tr>
<tr>
<td>10</td>
<td>200</td>
<td>150</td>
</tr>
</tbody>
</table>

TABLE 4.4

Results for sequentially pulsed p-St capacitors
on the LCR meter. However, some capacitors showed resistances as low as 1 kΩ, with an associated capacitance of 0 pF, after damage. The upper limit of resistance after damage was around 300 kΩ for devices on wafer type No.1. Some devices on wafer type No.2 had resistances as high as 1 MΩ.

The positive voltage breakdown threshold of p-Si (E-type NMOS) Capacitors was +200 V, while the negative voltage threshold was -100 V. With n-Si (D-type NMOS) capacitors, the breakdown thresholds were +175 V and -100 V. It is observed that the negative voltage threshold is not affected by the type of device, but the positive threshold is dependent on the dopant at the Si surface. The n+-Si (E-type PMOS) capacitors showed almost no polarity dependence.

Figures 4.15 and 4.16 show two C-V curves made before and after application of a +150 V pulse to a p-Si capacitor. There is very little variation between the curves. The device itself showed no decrease in parallel resistance, indicating that no damage had occurred. Figures 4.17 and 4.18 show two C-V curves made before and after applying a +50 V pulse to an n+-Si capacitor. A significant shift of the capacitance minima, in the negative voltage direction, is observed.

4.3 SEQUENTIAL PULSING

4.3.1. E-mode NMOS transistors

The results of sequential pulsing experiments on E-mode NMOS transistors are presented in Tables 4.2 and 4.3. Table 4.2 shows the effect of increasing the number of pulses at different voltages. The absolute number of devices with characteristics in each category are given. Table 4.3. shows the effect of increasing
the applied voltage for a fixed number of pulses. Again, the absolute number of devices, exhibiting characteristics in each category, is given in the table.

4.3.2. p-Si MOS capacitors

The results for p-Si MOS capacitors are presented in Table 4.4. R and C values are shown, after pulsed voltage stresses of 150 V and 200 V, as a function of the number of applied pulses. Resistances are observed to decrease as the number of pulses is increased.

4.4. CONTINUOUS VOLTAGE STRESS EXPERIMENTS

4.4.1. p-Si capacitors (E-mode NMOS)

4.4.1.1. Gate oxide leakage currents

The gate oxide leakage current (I_{Gleak}) has been plotted as a function of the applied voltage in Figure 4.19. It is important to note the rate at which I_{Gleak} varies with voltage, as a function of polarity. The I-V curve is asymmetrical for p-Si capacitors. A lower I_{Gleak} is observed for positive voltages compared with negative voltages.

A C-V curve made after a +36 V stress is shown in Figure 4.20. Figure 4.21. shows a C-V curve of a device stressed at -30 V. Figure 4.22 shows how the capacitance minima changes (ΔC) as a function of a positive voltage stress ranging from +26 V to +48 V.

4.4.1.2. Temperature effects
FIGURE 4.19. Graph of the oxide leakage current ($I_{\text{leak}}$) as a function of the applied stress voltage for an E-mode p-Si MOS capacitor.

**FIG. 4.20.** C-V curve of E-mode p-Si MOS capacitor after continuous positive voltage stressing at +36 V.

**FIG. 4.21.** C-V curve of E-mode p-Si MOS capacitor after continuous negative voltage stressing at -30 V.
\[ \Delta C = C - C_{\text{stress}} \]

**Figure 4.22** Graph showing the change in C-V curve minima \( \Delta V_c \) as a function of the stress voltage for an E-mode p-Si MOS capacitor.

**Figure 4.23** The temperature dependence of the oxide breakdown threshold under continuous voltage stress for MOS capacitors.
FIGURE 4.24  Graph of the oxide leakage current ($I_{\text{leak}}$) as a function of the applied stress voltage for a P-mode $n$-Si MOS capacitor.

FIGURE 4.25. Graph of Oxide Leakage Current ($I_{\text{leak}}$) as a Function of the Applied Stress Voltage for an E-Mode $n$-Si capacitor.
The effect of temperature on the gate oxide breakdown strength of a p-Si capacitor is shown in Figure 4.23. The breakdown voltage threshold decreases from +36 V at room temperature (25°C) to +28 V at 200°C.

4.4.2. n-Si capacitors (D-mode NMOS)

Figure 4.24. shows the variation of $I_{\text{Gleak}}$ with voltage for an n-Si capacitor. The curve is still asymmetrical, but the disparity between $I_{\text{Gleak}}$ at positive and negative voltages is not as large as that observed with p-Si capacitors.

4.4.3. n+-Si capacitors (E-mode PMOS)

Figure 4.25. shows $I_{\text{Gleak}}$ as a function of applied voltage for an n+-Si capacitor. The curve is seen to be fairly symmetrical about the origin, for positive and negative voltages and currents.

4.5 SUMMARY

1. The results of single pulse voltage stress experiments, pulse sequence experiments and continuous voltage experiments have been presented. E-mode and D-mode NMOS, E-mode PMOS transistors, p-Si, n-Si and n+-Si capacitors, have been reported in separate subsections, for each experimental condition.

2. The pulsed voltage experiments on E-mode NMOS transistors, have been evaluated by considering the percentage distributions of the device I-V characteristics, after pulsing. The characteristics are summarised in three categories, a) no change, b) degraded, c) catastrophic failure.
3. The distributions indicate that the pulsed breakdown is voltage sensitive, but is not affected by temperature. It is also dependent on the surface area of the gate oxide.

4. D-mode NMOS characteristics when degraded were observed to be different to those of degraded E-mode NMOS transistors.

5. E-mode PMOS transistors showed degradation identical to that of E-mode NMOS transistors.

6. The pulsed voltage breakdown of capacitor structures was shown to be polarity dependent. The breakdown threshold is higher for positive voltages. The dopant concentration of the Si surface influences breakdown, more n-type surfaces showing lower breakdown thresholds.

7. Sequential pulsing was found to increase the degradation observed in both transistors and capacitors.

8. Constant applied voltage stress on capacitor structures, showed that breakdown was dependent on the polarity of the stress voltage and the dopant concentration at the Si surface. More n-type surfaces showed a higher leakage current, while devices were more affected by negative voltages.
CHAPTER 5

DISCUSSION

5.1. INTRODUCTION

In this chapter, an analysis is made of the experimental results. The mechanisms and modelling of oxide breakdown are considered in Chapter 6.

5.2. DAMAGE IN MOS STRUCTURES

The I-V characteristics indicate that damage in the transistors manifests itself as various levels of degradation, i.e. $I_{DS}$ decreases for a given $V_{DS}$. Degradation can range from minor (1% of the undamaged $I_{DS}$), to major ($I_{DS}=0$), implying catastrophic failure.

In E-mode NMOS transistors, the decrease in $I_{DS}$, and the corresponding decrease in $g_{ds}$, indicates that the n-type channel had not reached the level of inversion expected in an undamaged device. This is due to the formation of a leakage path through the gate oxide, which prevents electrons from accumulating at the semiconductor surface under positive bias conditions.

In D-mode NMOS transistors, the leakage path prevents the surface from depleting under negative bias conditions. This is confirmed by the increase in $g_{ds}$, observed in these devices, after degradation.
In E-mode PMOS transistors, the inversion channel is affected by oxide leakage. Hence degradation characteristics, of the same kind as those obtained with E-mode NMOS transistors, are observed.

R and C measurements on damaged MOS capacitors, showed that the leakage resistance dropped by orders of magnitude, \( \approx 10^{14} \) ohms to \( 10^6 \) ohms. The capacitance was observed to decrease by various degrees, ranging from 1\% to 100\%. Oxide breakdown is, therefore, confirmed to have taken place.

Detailed failure analysis of the damaged devices was not performed. This would have involved sectioning the wafer and dissecting the chips to sizes suitable to surface analysis equipment. As a result, other devices on the wafers, which were required for further experiments, would have been destroyed.

5.3 PULSED VOLTAGE STRESS EXPERIMENTS

5.3.1 E-mode NMOS transistors

5.3.1.1 Analysis of I-V characteristics - voltage effects

Category 1 - The unchanged characteristics indicate that the oxide has not been damaged. As the magnitude of the pulsed voltage increases, the number of undamaged devices decreases, from 70\% at 50 V to 10\% at 250 V. A strong voltage dependence of the damage mechanism is therefore indicated.

Category 2 - This category of damage has serious implications in the applications of MOS devices. The characteristics show such devices could still appear as operational transistors, especially in digital circuitry. They form part of the "walking wounded"
which could contribute to latent failures. Failures of this type are extremely hazardous since they occur during the working life of a device. This is supported by the results of experiments with sequential pulses.

Category 3 - In this case, as $V_{GS}$ increases beyond a certain voltage, $I_{DS}$ begins to decrease for a given $V_{DS}$. This effect is the result of deep depletion taking place at the semiconductor surface; i.e. when the depletion layer width becomes wider than in thermal equilibrium [1]. The oxide is once again damaged and becomes a weak conductor; however, the leakage resistance is less than in the previous category. Consequently, a larger number of electrons are conducted out of the semiconductor surface, through the oxide. The number of ionised acceptors ($N_A$) at the surface increases, in order to support charge neutrality, and this causes the depletion layer to widen. The result is an increase in the threshold voltage of the transistor, since [2] - [4],

$$V_T = \phi_{ms} + 2\phi_F - \frac{Q_{ss}}{C_{ox}} + \frac{(4\varepsilon_s \phi F N_A)^{\frac{1}{2}}}{C_{ox}} - \frac{Q_{off}}{C_{ox}}$$

The transistor mutual transconductance is given by [2] - [4],

$$g_{m \text{ sat}} = \frac{\Delta I_{D \text{ sat}}}{\Delta V_{GS}}$$

and,

$$g_{m \text{ sat}} = C_{ox} \mu_n \left( \frac{W}{L} \right) (V_{GS} - V_T)$$

The symbols have their usual meanings.
Therefore, as $V_T$ becomes greater than $V_{GS}$, $g_{msat}$ goes negative, giving rise to the characteristics observed in this category. At acceptor concentrations of $N_A = 10^{17}$ cm$^{-3}$, a $V_T$ of $\approx 5$ V can be expected [2].

Category 4 - The resistive characteristics are another form of degradation. The restoration of standard I-V characteristics when the source and drain contacts to the CT are reversed, indicates that the oxide leakage path is between gate and drain. Drain lateral diffusion under the gate is approximately 0.5 μm and oxide breakdown can take place directly between the gate and drain. The CT measures $I_{DS}$ at the drain terminal, and, will only see the gate oxide resistance as current flows from the gate to the drain. Measurements using separate multimeters at the drain and the source of a biased transistor, have confirmed that this is true. Since oxide breakdown between the gate and source/drain is not significantly different from breakdown directly over the channel, all three types of breakdown described by categories 2 to 4 will be classified as degradation. This degradation damage, in transistors can be compared to the MOS capacitor damage, where the leakage resistance ranges from 50 kΩ to 1 MΩ, with a capacitance which is still significant ($\approx 50\%$ of its original value).

Category 5 - These characteristics are the result of serious oxide breakdown. The extent of the damage prevents an inversion layer from forming at any $V_{GS}$; the device is, therefore, inoperative. A comparison with MOS capacitors can be made, where the measured resistance ranges from 100 Ω to 1 kΩ and the gate capacitance is zero.

5 3 1.2. Temperature effects on oxide breakdown

No distinct trend was observed for the temperature dependence of any of the categories. It is therefore concluded, that the pulsed voltage breakdown sensitivity of MOS devices is temperature independent.
Other workers have observed an ESD sensitivity for NMOS LSI devices which is affected by temperature [5]. The cause of failure was traced to a diode in the protection circuit which was influenced by the applied temperature. Relating this finding with the results for discrete MOS structures, as described in this thesis, implies that it is possible for input protection circuitry to introduce a temperature dependence.

5.3.1.3. Dimensional dependence of oxide breakdown

The increase in the percentage number of devices showing degradation as the gate area is increased, supports the results of Wolters et al [6], for continuous voltage breakdown (N.B. The results on dimensional dependence presented in this thesis were first published in August 1984 [7]).

The presence of defects in an oxide in the form of impurities or dangling Si or O bonds, will weaken the dielectric strength of the oxide [8][9]. Such defects have been termed intrinsic. Extrinsic defects are introduced in the manufacturing process, and take the form of pinholes through the oxide, and, surface non-uniformities at the interfaces. The density of such defects is expected to be very small (=1 cm⁻²) in a production assessment wafer. A large sample size, such as that used in the NMOS transistor experiments, minimises the possibility of extrinsic defects influencing the results.

It is reasonable to assume that intrinsic defects are randomly distributed across the oxide area, A, which is defined by the gate width and the gate length. Wolters assumes that the probability that m defects, failing at or below a field strength F, being present in a certain area is given by the Poisson distribution.
The probability of having one or more defects in an oxide failing at a field strength F, is given by,

\[ P(F) = 1 - \exp\left\{ -A.D(F) \right\} \] ............ Eq(5.4)

Therefore, as A increases, P(F) gets larger. Eq(5.4) holds for large gate areas and uniform electric fields. The effects of source/drain lateral diffusion and field oxide undercutting which will be found in very small dimension devices, are not included in the equation.

The gate to drain/source breakdown would be expected to become more significant as the ratio of the effective channel length to the length of the lateral diffusion gets smaller. Figure 5.1 shows the % number of devices which exhibited gate-source or gate-drain breakdown as a function of the channel length at a constant gate width of 3.2 μm. The devices were all stressed at 200 V. Hence, it can be seen that as the length approaches \( \Delta L(=0.5 \mu m) \) the number of breakdowns increases. The implications of this result with respect to very small devices are discussed in Appendix 3.

5.3.2. D-mode NMOS transistors

The degradation characteristics of depletion-mode devices were different to those of the E-mode NMOS transistors because of the implanted inversion layer. The higher \( g_{ds\ sat} \) observed in damaged devices with negative \( V_{GS} \), indicates that the surface was not depleted to the extent of an undamaged device. As \( V_{GS} \) is made more negative, a constant \( g_{ds} \) is observed, with \( I_{DS} \) increasing in the positive direction. This implies
FIG. 5.1 % DEVICES SHOWING GATE TO SOURCE/DRAIN DAMAGE AT 200 V

% DEVICES

\[ w = 3.2 \mu m \]

CHANNEL LENGTH, \( \mu m \)
that the increase in $|V_{GS}|$ does not affect the inversion layer, but instead begins to aid conduction through the oxide.

However, there is the possibility that if the point of breakdown is closer to the source, then a positive $V_{DS}$ will begin to cause pinch-off in the inversion layer at low $|V_{GS}|$. This is because the lateral voltage in the channel, i.e. $V_{DS \text{ eff}}$, is greater than $V_{DS}$ because of the added effect of $V_{GS}$. $I_{DS}$ therefore, increases as $V_{DS}$ goes from 0 V, and then decreases as pinch-off sets in. The current, preferring the path of lower resistance, travels from the source contact to the gate. For higher $V_{GS}$, the circuit becomes a current divider, and, a steady $I_{DS}$ will be observed. From the type of I-V characteristic observed, it is possible to identify the possible location of the damage; i.e. whether damage is closer to the source or the drain.

The experiments on a small sample of D-mode NMOS transistors (28 devices), showed that the extent of the degradation increased as the magnitude of the applied voltage pulse was increased. All devices which were stressed showed degradation.

5.3.3. D-mode PMOS transistors

D-mode PMOS transistors were not available for experimentation. However, it is expected that the shape of the I-V characteristics after damage, will be similar to that of the D-mode NMOS transistors. The only consideration here, is that the operating voltage polarity of a PMOS transistor, is opposite to that of the equivalent NMOS device. Electrons in PMOS devices will flow in the opposite direction to those in NMOS devices, but this should have no significant effect on the characteristics obtained.

5.3.5. MOS capacitor structures
The results for the capacitor structures have been related to the transistor I-V characteristics in Section 5.3.

The breakdown threshold for negative voltage pulses, i.e. when the n+ polysilicon gate was the injecting electrode, was not affected by the dopant concentration at the silicon surface. However, positive voltage breakdown thresholds were significantly affected by the type of silicon surface. As the surface became more n-type, the disparity between the positive and negative breakdown thresholds becomes less. This indicates that the breakdown threshold is influenced by the electrode which injects electrons into the SiO₂. Charge injection and its influence on breakdown has been discussed by other workers [10][11], although it is a relatively new concept with regard to the mechanism of oxide breakdown [12]. Earlier work always considered breakdown to be a field-dependent phenomenon [13].

C-V curves made on pulsed MOS capacitors would indicate whether any new electron oxide traps have been generated in the). No shift in the C-V curve of a p-Si MOS capacitor was observed after a 150 V pulse was applied. It can, therefore, be concluded that the pulse did not generate any new electron traps in the p-Si capacitor. However, the capacitance minima of n+-Si capacitors showed a significant shift in the negative voltage direction [14], after being subjected to a positive pulse just below the breakdown threshold. Hence, positive pulsed voltage stress forms interface states in n+-Si capacitors.

5.4. SEQUENTIAL PULSING

5.4.1. E-mode NMOS transistors
The extent of the degradation increases as the number of pulses is increases. However, devices which had not shown degradation after a single pulse, did not appear to be affected by further pulses at the same voltage. This is in agreement with Wolters' [6] assumption, for constant applied voltages, that the weakest defect determines the dielectric strength of the oxide. Therefore, in the oxides which showed no damage at a given voltage, there were no weaknesses activated by that electric field.

5.4.2. p-Si MOS capacitors

Experiments on p-Si MOS capacitors clarified the results obtained after sequentially pulsing E-mode NMOS transistors. As the number of pulses increased, the resistance got lower. This is comparable to the increase in degradation observed in the transistors (Section 5.4.1.).

It must be noted, that in both the transistors and the capacitors, the degradation was not a linear function of the applied number of pulses.

5.5. CONTINUOUS VOLTAGE STRESS EXPERIMENTS

5.5.1. Breakdown thresholds

The breakdown threshold of the three MOS capacitor structures were approximately ±35 V. This is significantly different to the pulsed voltage breakdown thresholds, implying that two different mechanisms are involved in the two cases. The continuous voltage breakdown thresholds were shown to decrease as the temperature was raised. Pulsed breakdown, in comparison, was not temperature dependent, adding further evidence to the case for two different mechanisms of breakdown.
5.5.2. Polarity

The polarity of the constant applied voltage did influence the breakdown threshold, and, was related to the type of capacitor. For all three types of capacitor, the negative voltage breakdown thresholds were very abrupt. However, in the positive direction, the rate of increase of gate leakage current with applied voltage, was greater for more n-type Si surfaces. These results are comparable to the threshold voltage differences under pulsed conditions, and, supports the charge injection model of breakdown.

5.5.3. C-V curves

Changes in C-V curves of the MOS capacitors began at ±30 V, which is in the voltage region where the equivalent I-V curves began to change. Therefore, shifts in the C-V curves can be considered to indicate the onset of oxide damage.

The shift of the curves in the negative direction, after the application of negative voltage stress, implies that a positive charge has been formed at the Si-SiO₂ interface. This can be attributed to the emission of electrons from surface trap states and into the Si under the influence of the large negative voltage. A number of uncompensated, positively charged states, are formed at the interface [15]. The applied gate voltage under normal operating conditions is supplemented by these charges, causing depletion, and subsequently, inversion to occur at less positive voltages.

After a positive voltage stress, the C-V curves become distorted, although there is negligible shift along the negative voltage axis. In this case, the semiconductor surface is the cathode. Any unoccupied interface states are filled by electrons injected from the surface. In addition, a number of electrons will be trapped at newly formed
states in the oxide within close proximity (≈20 Å) of the surface. The result is that the fixed charge concentration at the surface increases, requiring that the voltage required for inversion is greater. This broadens the C-V curve. The space-charge capacitance ($C_{sp}$) will also rise because of the fixed charge, causing the minimum overall capacitance to increase.

5.6. GENERAL

A distinct difference exists between the charge supplied by a single pulsed voltage and a constant applied voltage. In the first case, a fixed amount of charge is deposited on the contact electrode. This charge will decay according to the properties of the gate oxide capacitor. In contrast, a constant applied voltage will continuously supply charge to the contact electrode, at a constant energy level.

5.7 SUMMARY

1. Damage in MOS structures is due to electron conduction paths in the gate oxide, as a result of the voltage stress. The presence of oxide defects will enhance the prospect of damage.

2. Analysis of the I-V characteristics of MOS transistors showed that as the magnitude of the pulsed voltage increases, devices are more likely to be damaged.

3. Devices can still be considered to be functional even though their I-V characteristics show degradation, especially in digital applications.
4. The application of sequential pulses showed that degraded devices would be highly prone to subsequent failure. Such devices would form part of the "walking-wounded" once operational and can give rise to latent failures.

5. The dimensional dependence of oxide breakdown supports the assumption that intrinsic defects, such as those introduced by weak Si-O bonds or impurities, increase the sensitivity of the SiO₂ to breakdown.

6. The temperature independence of pulsed voltage damage when compared with the temperature dependence of continuous voltage damage, indicates that two different breakdown mechanisms are involved.

7. This is further highlighted by the large difference between the breakdown thresholds under pulsed and continuous voltage conditions. Pulsed breakdown thresholds are a factor of 5 greater than the continuous voltage thresholds.

8. The observation of a significant polarity dependence for both pulsed and continuous voltage breakdown of MOS capacitors, is related to the effect of the dopant concentration at the Si surface. This indicates that the breakdown mechanisms are strongly influenced by the charge injection process at the electrodes.

9. C-V curves on MOS capacitors show that the injection of charge at the Si-SiO₂ interface does take place. The effects are observed just prior to the onset of oxide breakdown.

5.8. REFERENCES


CHAPTER 6

MODELLING OXIDE BREAKDOWN MECHANISMS
IN MOS STRUCTURES

6.1. INTRODUCTION

A model of the SiO₂ breakdown mechanisms, under pulsed and continuous voltage conditions, is presented in this chapter. The model is founded on the experimental work discussed in the previous chapters.

In Chapter 4, it was observed that MOS capacitor structures with three different concentrations of electrons at the surface, when in thermal equilibrium, had different rates of breakdown. Devices with more n-type doping showed more defined breakdown thresholds. Under pulsed conditions, the positive breakdown thresholds were greater than the negative thresholds, for more p-type devices. Therefore, it was concluded that the injection of electrons from the cathode into the SiO₂ was responsible for breakdown.

Previous workers have observed that MOS structures with n-type surfaces had lower breakdown thresholds than p-type surfaces [61] - [63]. There is also evidence to show that the breakdown threshold with an n+ polysilicon gate as the cathode, is lower than when a less n-type surface is the cathode [51]. This disparity between the breakdown thresholds of the p and n-type structures has been attributed to intrinsically higher defect densities in films grown on different wafer types [62]. However, this explanation is not consistent with both the pulsed and continuous voltage observations which indicate that the injected charge must be considered. In addition, numerous
papers on oxide breakdown and charge injection, have observed differences in the behaviour of SiO₂ films on p and n-type Si, but have not analysed these effects [24][51][53][56][64]. It is significant that, in every case, the n-type structure has a lower breakdown threshold or shows a higher leakage current.

MOS charge injection theory, and electron conduction processes are described in this chapter, and adapted to explain the phenomenon associated with oxide breakdown in MOS structures [67][70]. Two different breakdown mechanisms are proposed for continuous and pulsed voltage breakdown.

6.2. THE MOS STRUCTURE

6.2.1. Introduction

This section (6.2) outlines the principal theory of MOS structures, paving the way for the adaptation of dielectric breakdown theory to these structures in later sections.

6.2.2. Energy band structure

The energy band structure of an MOS capacitor is shown in Figure 6.1. [1][2]. Si has a band gap of 1.12 eV which is small compared to the forbidden gap in SiO₂ of 8.8 eV. It must be noted, however, that in SiO₂, it is not strictly correct to refer to a forbidden (or band) gap, since it is an amorphous material and therefore has localised rather than continuous energy levels [3].

The electron affinity (χ) of SiO₂ is 0.9 V, i.e. the energy required to remove an electron from the bottom of the conduction band is 0.9 eV. χ is ≈4.15 V in Si. The
FIGURE 6.1. Energy band-diagram of MOS structure (n polySi-SiO₂–pSi).
FIGURE 6.2. a) n-Si MOS in accumulation ($V_G > 0$)

FIGURE 6.2. (b) n-Si MOS in inversion ($V_G < 0$)
position of the Fermi level in the Si is dependent on the impurity concentration in the material. A p-type material will have an effective, or quasi, Fermi level ($E_{FP}$) below the intrinsic Fermi level ($E_{FI}$) of undoped Si. An n-type material will have $E_{FN} > E_{FI}$ [4]. The n+ polysilicon gate has a Fermi-level which coincides with the bottom of the Si conduction band [2][5].

From the energy level differences between the bottom of the SiO$_2$ conduction band and the Si conduction bands, the energy required to get an electron from the Si into the lowest unoccupied states in the SiO$_2$ is $\approx 3.15$ eV.

6.2.3. Band-bending

Upon application of an electric field across the capacitor, the energy levels in the conduction band and the valence band are shifted in order to accommodate the charges at the Si surface. Energy diagrams for an n-type MOS system in accumulation (Figure 6.2 (a)) and in depletion/inversion (Figure 6.2 (b)) show the various potential and the surface charge concentrations [1].

The potential $\phi(x)$ is defined by,

$$q\phi(x) = E_{FN} - E_i(x)$$  \hspace{1cm} \text{Eq (6.1)}

where $E_i(x)$ is the intrinsic Fermi level.

As $x \to \infty$, i.e. deep in the bulk of the Si, $\phi(x)$ becomes the bulk potential $\phi_B$. At the silicon surface ($x = 0$), $\phi(x)$ is the surface potential $\phi_S$. The band-bending $\phi(x)$ is then defined as
\[ \varphi(x) = \phi(x) - \phi_B \quad \text{Eq (6.2)} \]

\( \varphi(x) \) is therefore, representative of the potential at any point in the depletion region, with respect to its value in the bulk. The total bend-bending, \( \varphi_s \), is given by the total potential difference between the silicon surface (\( \varphi_s \)) and the bulk (\( \varphi_B \)).

\[ \varphi_s = \varphi_s - \varphi_B \quad \text{Eq (6.3)} \]

In order to calculate the surface potential as a function of the distance, \( x \), from the Si-SiO\(_2\) interface (\( x = 0 \)) to the Si bulk, it is necessary to solve Poisson's equation in one dimension.

\[ \frac{d^2 \varphi(x)}{dx^2} = -\frac{\rho(x)}{\varepsilon_s} \quad \text{Eq (6.4)} \]

\( \rho(x) \) is the charge density (coul cm\(^{-3}\)) and \( \varepsilon_s = 1.04 \times 10^{-12} \text{ Fcm}^{-1} \) is the permittivity of silicon.

The solutions of this equation have been analysed by Kingston and Neustadter [6], Garrett and Brattain [7], Seiwatz and Green [8] among others [9][10][11] and will not be described here. However, the solutions are complicated and will be summarised as follows:

1. The relationship between the electric field (\( E = -d\varphi/dx \)) and the potential is given for a p-type material as,

\[ E = \pm \frac{\sqrt{2} kT}{qL_D} \text{F} \left[ \beta \varphi \left( \frac{n_{po}}{P_{po}} \right) \right] \quad \text{Eq (6.5)} \]
where $L_D$ is the effective Debye length for holes.

$$L_D = \sqrt{\frac{kT e_s}{p_{po} q^2}} \quad \text{Eq (6.6)}$$

and

$$\beta = \frac{q}{kT} \quad \text{Eq (6.7)}$$

$p_{po}$ and $n_{po}$ are the equilibrium densities of electrons and holes respectively, in the bulk of the semiconductor.

$$F\left[ \beta \phi, \frac{n_{po}}{p_{po}} \right] = \left[ (e^{-\beta \phi} + \beta \phi - 1) + \frac{n_{po}}{p_{po}} (e^{\beta \phi} - \beta \phi - 1) \right]^{1/2} \quad \text{Eq (6.8)}$$

for a non-degenerate surface. A degenerate surface will require adjustment as described by Seiwatz and Green [8] and Many et al [12].

2. The potential as a function of distance is given by

$$\frac{x}{\beta L_D} = \frac{1}{\sqrt{2}} \int_{\beta \phi_s}^{\beta \phi} \frac{d(\beta \phi)}{F\left[ \beta \phi, \frac{n_{po}}{p_{po}} \right]} \quad \text{Eq (6.9)}$$
which has to be solved numerically.

3. The surface charge per unit area is given by,

\[
Q_s = \pm \sqrt{2} \frac{e_s kT}{qL_D} F(\beta \varphi_s, \frac{n_{po}}{p_{po}}) \tag{Eq (6.10)}
\]

Eq (6.8), (6.9) and (6.10) are represented graphically in Figures 6.3, 6.4, and 6.5 respectively, for easy reference. Using these equations and the equations for surface degeneracy [8][12], the maximum band-bending at the surface of an MOS system can be calculated.

For example, with a positive applied voltage of 40 V across an oxide of thickness 400 Å,

a) a p-type material with a bulk dopant concentration \((N_A)\) of \(4 \times 10^{15} \text{ cm}^{-3}\) has a \(\varphi_s = 0.98 \text{ V}\) and is in strong inversion.

b) an n-type material with a bulk dopant concentration \((N_D)\) of \(4 \times 10^{15} \text{ cm}^{-3}\) has a \(\varphi_s = 0.33 \text{ V}\) and is in strong accumulation.

6.2.4. Majority and minority carrier response times [1]

In accumulation and depletion, the charge in the silicon surface contributes to the MOS capacitance. Majority carrier response times to an applied gate voltage are given by [1],
FIG. 6.3 GRAPH OF F-FUNCTION VS SURFACE POTENTIAL FOR p-TYPE Si

FIG. 6.4 BAND-BENDING VS DISTANCE FROM INTERFACE FOR n-Si

FIG 6.5 GRAPH OF SURFACE CHARGE DENSITY VS SURFACE POTENTIAL
where \( n \) is the number of carriers per unit volume. For \( n \approx 10^{16} \text{cm}^{-3} \), \( \tau_{\text{maj}} \approx 10^{-12} \) secs. The majority carrier response can be considered instantaneous for the purposes of the analysis presented here.

In inversion, the minority carrier response time is important. Response times in Si at room temperature are very slow, typically 0.01 - 1 sec. in strong inversion. The exact value is dependent upon the impurity concentration, the density of conduction band electrons or valence band holes and the electron/hole lifetimes [1]. However, the response time is very long and will influence the charge concentrations at the silicon surface, when pulses with risetimes of the order of \( 10^{-6} \) secs. are applied. While the risetime of the pulse should have no effect on the inversion layer, the decay time, which is proportional to the system RC time constant, will result in bringing minority carriers to the silicon surface. The resistance of a typical MOS gate oxide capacitor is \( \approx 10^{16} \Omega \) while \( C = 10^{-16} \text{F} \), for a gate area of \( 1 \mu \text{m}^2 \). The decay time is 1 sec. which is well within the minimum minority carrier response time.

6.3. PROPERTIES OF THE SILICON DIOXIDE DIELECTRIC

6.3.1. Structure of silicon dioxide

Thermal SiO\(_2\) is amorphous in structure, i.e. long range order is absent, the array of equilibrium atomic positions being strongly disordered [13]. This disorder gives rise to localised energy levels within the forbidden gap due to the tails of energy states emerging from the conduction and valence band edges and stretching into the
forbidden gap (see Figure 6.6) [14][16][17]. However, these localised energy levels are not considered to be sufficiently deep or dense to be important to the electronic conduction processes in SiO₂ [14][17]-[20].

SiO₂ is formed by a sharing of valence-electrons between silicon and oxygen to form four silicon-oxygen bonds. Each bond is largely covalent with a small ionic component [1]. The bonding energy between silicon and oxygen is \( \approx 6 \text{ eV} \) [15].

6.3.2. Oxide defects

6.3.2.1. The Si-SiO₂ interface [1][21]

Chemical inhomogeneities such as stretched, bent or broken bonds, stoichiometric conditions as well as ionic impurities at the Si-SiO₂ interface, will result in the formation of traps, which, by capturing charge carriers, can affect device performance. Additionally, these traps can also provide the first stages of an electronic conduction process in the SiO₂. Interface traps characteristic of thermal oxidation are donor type, and, under the influence of a suitable electric field, will release electrons to form a positive charge distribution at the interface. They have energies ranging all the way across the Si forbidden gap. If the silicon surface is p-type, the positive charge causes inversion to occur at more negative gate voltages. This can be observed using a C-V plot as shown in Chapter 4.

6.3.2.2. Bulk oxide traps [1][22]-[27]

Bulk oxide traps are associated with defects in the SiO₂ such as impurities and broken bonds. All thermally grown SiO₂ films contain electron traps related to the presence of water (H₂O) in the film [22][23]. It has also been shown that ionic impurities such
FIGURE 6.6. Schematic illustration of the forbidden gap, depicting the presence of localised energy states.
as Na⁺ exist as traps in the SiO₂. The energy levels of the traps with respect to the conduction band edge (the trap depth) have been reported as 1.9, 2.0, 2.13, 2.35 and 2.37 eV, with a spatial distribution which is uniform through the oxide [24]-[26]. Ion implantation of boron, arsenic and phosphorus is also a source of defects, creating numerous traps resulting from exchange of energy between the ions and the SiO₂ lattice. Trap densities are proportional to the implantation energy for a given number of implanted ions [1][27].

6.4. ELECTRONIC CONDUCTION PROCESSES IN SiO₂

6.4.1. Hopping conduction [14][28]-[32]

The oxide traps and donor sites described in the previous section, will appear as coulombic potential wells in the SiO₂ band-diagram, once charged. A trap site having captured an electron becomes negatively charged and the energy diagram takes the form shown in Figure 6.7 [29]. A donor site having emitted an electron becomes positively charged and the energy diagram takes the form shown in Figure 6.8 [32]. The difference in the two energy diagrams is that the former has an external potential well which is assumed to be coulombic, while the inner potential well is slightly steeper in order to maintain the trapping level. However, the effect of a uniform external field, in either of the two cases, would be to lower the potential barrier. This is due to the density of ionized centres becoming sufficiently large for their potential wells to begin to interact with each other. For a donor centre, the effect can be described by Figure 6.9 [14]. The potential energy of the well $\phi(r)$ then becomes,

$$\phi(r) = - \frac{e^2}{4\pi\epsilon r} - eFr \cos\theta$$

Eq (6.12)
where $F$ is the external field, $\theta$ is the angle between the radius vector $r$ and the electric field. The zero of energy is taken to be the bottom of the conduction band [32].

The result is that electrons are able to move through the oxide by hopping from well to well in a manner analogous to that described by Schottky emission. This is known as the Poole-Frenkel effect [33].

It is possible, to picture the mechanism of hopping transitions between traps in a high electric field, in a qualitative manner, using Figure 6.10 [30]. The electric field enables carriers to raise their position relative to the conduction band edge and so to 'move' to another trap site.

A finite temperature dependence does exist for this mechanism. Electrons gain thermal energy from the lattice vibrations to enable them to escape from the coulombic potential wells. Hence, it is possible for conduction to occur at lower electric fields as the temperature of the sample is increased [29].

Trap relaxation times are expected to be of the order of milliseconds ($10^{-3}$ secs) if not longer, and are dependent on the energetic distances between neighbouring traps [30].

6.4.2. Impact ionization in $\text{SiO}_2$

Impact ionization in solids has been extensively studied by Fröhlich [37][38], Seitz [39], Callen [40] and, more recently, O'Dwyer [35].

Since the forbidden gap in $\text{SiO}_2$ is very large (≈8eV), the conduction band can be considered to be almost devoid of free electrons. However, under the influence of a large electric field, the conduction band edge is sufficiently deformed to enable
FIGURE 6.7. A negatively charged electron trap site

FIGURE 6.8. A positively charged donor site

FIGURE 6.9. A positively charged donor site under the influence of an external electric field

FIGURE 6.10. Schematic illustration of electron transitions between full and empty traps under the influence of a high electric field. (ref. 14)
FIGURE 6.11. Schematic illustration of the impact ionization process in SiO₂ (ref 46)
injection of electrons from the conduction band of the silicon into the conduction band of the silicon dioxide, either by means of Schottky emission or by Fowler-Nordheim tunnelling [33]-[35]. Once in the conduction band of the SiO₂, these free electrons gain energy from the electric field and upon collision with the lattice, until there are enough high energy electrons to cause impact ionization of electrons from the valence band, leaving behind low energy holes. A distortion of the electric field occurs because of the positive charge contributed by the build-up of residual holes, resulting in a corresponding increase in the injection current. Hence, an increase in the rate of impact ionization occurs due to the higher field and current [46]. The process is schematically depicted in Figure 6.11.

The threshold energy for ionization, \( E_i \), is given by [45][49][50],

\[
E_i = E_C \left[ \frac{2 + \frac{m_h}{m_e}}{1 + \frac{m_h}{m_e}} \right]
\]

Eq (6.13)

and lies between 11.5 eV and 12.5 eV for SiO₂. \( \frac{m_h}{m_e} \) is the effective mass ratio for SiO₂ and \( E_G \) is the band gap. The critical electric field for breakdown under continuous voltage conditions is given by [46],

\[
F_b = \left[ 1 - e^{-\gamma(1 + \gamma e^{-\gamma})} \right] \frac{E_G}{41e\lambda}
\]

Eq (6.14)

\( E_G = 9.0 \) eV is the insulator band gap, \( \lambda \) is the electron-phonon scattering length and \( \gamma = \frac{\hbar \omega}{e F \lambda} \). \( \hbar \omega \) is the phonon energy. \( \lambda \) is required to be about 1.4 Å in order to fit experimentally determined breakdown fields where \( F_b = 9.5 \) MV cm⁻¹. It is found that \( F_b \) is only dependent on \( \lambda \) and \( E_G \).
The ionization rate for amorphous SiO$_2$ with an applied field of $=10^7$ Vcm$^{-1}$, is $10^{12}$ per sec. [49]. It has been estimated that $=10^{12}$ electrons are required to cause breakdown in a solid $10^{-5}$ cm$^2$ in area and 1 cm long, which is equivalent to $=40$ generations of collision ionization [39]. If $\alpha(F_b)$ is the collision ionization rate per unit length, then [48]

$$\alpha(F_b) \cdot d = 40$$

Eq (6.15)

where $d$ is the thickness of the oxide. $\alpha(F_b)$ has been shown [45] to be $=10^4$ cm$^{-1}$ for $F_b = 10^7$ Vcm$^{-1}$, and with $d=10$ nm the time required to initiate impact ionization breakdown is $<10^{-9}$ sec. in SiO$_2$. Therefore, it is well within the risetimes of the pulses used in the experiments described in Chapter 3.

The effect of traps in the insulator on the tunnelling properties has been investigated [41]-[44]. It has been shown that the presence of a trap with the same energy as that of the incident electron causes the amplitude of the wave to resonate with the potential well created by the trap. The transmission function of such an electron will be greater than for an electron tunnelling via the usual process. Transmission of this nature is called resonance tunnelling, and electrons transmitted in this manner will increase the probability of impact ionization in the SiO$_2$ conduction band.

Traps have also been shown to have an important effect on the internal fields in the oxide [46][51]. As mentioned earlier, holes generated in impact ionization will be driven to the cathode where they can be trapped, thereby, increasing the field.

The result will be an increase in electron injection from the cathode due to the higher localized field and therefore a higher probability of impact ionization in the SiO$_2$. A quantitative model has been developed by Chen et al [51].
6.5. CHARGE INJECTION MECHANISMS

6.5.1. The Schottky effect [2][29]

The Schottky effect is the injection of charge by thermionic emission over a potential barrier, which has been reduced by the influence of an external field (Figure 6.12). The current density can be written as [2][29],

\[
J = \frac{4\pi q m^* k^2}{h^3} T^2 \exp \left[ -\frac{e \left( \frac{1}{\beta_s F^2} - \phi \right)}{kT} \right]
\]

Eq (6.16)

where \( m^* \) is the effective mass of the electron and \( \beta_s \) is the Schottky field-lowering coefficient given by,

\[
\beta_s = \sqrt{\frac{e}{4\pi \varepsilon_o \varepsilon_i}}
\]

Eq (6.17)

where \( \varepsilon_o \) is the permittivity of free space and \( \varepsilon_i \) is the relative permittivity of the insulator. \( \phi \) is the potential barrier between the cathode and the dielectric. Injection takes place over the potential hill and not through it, and is a function of the barrier lowering due to the applied field, as well as, the thermal energy of the electrons at the electrode.

6.5.2. The Poole-Frenkel effect [2][14][29][32][33]
The Poole-Frenkel effect is an extension of the Schottky effect [33], and considers the thermionic emission of trapped charge carriers in a solid under the influence of an external electric field, as described in Section 6.4.1. The current density is of the form [2][29],

\[
J = F \cdot \exp \left[ \frac{1}{e(\beta_{PF}F^2 - \phi)} \right]
\]

Eq (6.18)

where \( \beta_{PF} \) is the Poole-Frenkel field lowering coefficient given by,

\[
\beta_{PF} \approx 2\beta_S
\]

Eq (6.19)

\( \beta_S \) is as given in Eq (6.17).

The constant of proportionality will depend on the number of available donor/trap centres in the dielectric and the distance travelled by each carrier after emission [14].

6.5.3. Tunnelling in MOS structures

Electrons can also reach the SiO₂ conduction band by quantum mechanical tunnelling under the influence of strong electric fields [34]-[36][52]. Early work is based on a semiclassical framework, using the dynamics of a free-electron model of a metal-vacuum contact.

Consider the triangular potential barrier shown in Figure 6.13. The potential barrier \( V(x) \) is given by,
\[ eV(x) = -eFx \quad \text{Eq (6.20)} \]

where \( e \) is the electron charge \((1.602 \times 10^{-19} \text{ coulombs})\), \( F \) is the external electric field and \( x \) is the distance in the direction indicated in the figure. However, the external electric field induces an image-force which rounds off the top of the barrier, lowering the potential energy at the interface (as in the Schottky effect) (Figure 6.12). The potential barrier becomes [35][36],

\[ eV(x) = -eFx - \frac{e^2}{4ex} \quad \text{Eq (6.21)} \]

The tunnelling current, \( J \), is found by inserting this potential energy into Schrödinger's Equation and solving it to obtain the wave function and the transmission and the transmission probability \( D \), as described in Appendix 2 [34][52]. Hence [35][53],

\[ J = \frac{e^3F^2}{8\pi\hbar\phi} \exp \left[ -\frac{4\sqrt{2m^*} \phi^2}{3\hbar eF} \right] \quad \text{Eq (6.22)} \]

where \( \phi \) is the barrier height measured between the Fermi surface in the metal and the conduction band in the dielectric. \( m^* \) is the effective electron mass appropriate to the electrons in the dielectric.

The transmission coefficient, \( D \), is given by [34][36],
\[ D(E_x) = \frac{4 E_x^2 (\phi - E_x)}{\phi} \exp \left[ -\frac{4\alpha}{3\hbar^2} (\phi - E_x)^{\frac{3}{2}} \right] \]  

where,

\[ \alpha = \left[ \frac{8m^2}{\hbar^2} \right]^{\frac{1}{2}} = 5.16 \times 10^{-7} \text{ (eV)}^{-\frac{1}{2}} \text{ cm}^{-1} \]  

and,

\[ E_x = \frac{p_x^2}{2m} \]  

is the kinetic energy of the incident electrons.

However, when considering tunnelling through MOS structures, the above theory is inadequate. Under the influence of the high fields required to initiate tunnelling at the Si-SiO\textsubscript{2} interface, the Si band-bending will be significant and the high concentration of charge at the semiconductor surface (whether in inversion or accumulation) must be considered. Large concentrations of charge at the surface will cause energy level quantisation (subbands) to occur within the Si conduction band. Motion perpendicular to the Si-SiO\textsubscript{2} interface will, therefore, take place only at discrete energy levels. Hence, the free-electron model used by Fowler and Nordheim in their theory, cannot be used [54]-[58].

The effect of energy band quantisation at the Si-SiO\textsubscript{2} interface has been extensively studied [54]-[57]. It has been shown that the ground state in the silicon conduction
band \((E_0)\) can be separated from the next energy level by almost 0.1 eV depending on the extent of the band-bending and the dopant concentration \([55]\). In order to obtain the modified transmission coefficient and, therefore, the tunnelling current under these conditions, the basic approach described in Appendix can be used \([57][58]\). However, once energy level quantisation is introduced, and good modelling of the band-bending is included, an exact solution becomes impossible. Hence, various approximation techniques have been used, although even these can become extremely complicated. The outcome is a transmission coefficient and a tunnelling current which are dependent on the energy of the ground state subband \((E_0)\) and the depth of the silicon band-bending \([56][58][60]\). A simplified expression for the tunnelling current is obtained by considering the Si-SiO\(_2\) interface shown in Figure 6.14 as \([56][60]\),

\[
J = q\theta N_{inv} P \tag{6.26}
\]

where \(N_{inv}\) is the number of electrons per unit area at the interface, and \(\theta\) is the fraction of these that reside in the lowest subband. \(P\) is the transmission probability current per electron for the ground state. For an inversion layer density \(N_{inv} = 10^{13}\) \(\text{cm}^{-2}\) corresponding to an external field of \(10^7 \text{ Vcm}^{-1}\), and with a depletion concentration \(N_{dep} = 10^{11}\) \(\text{cm}^{-2}\) corresponding to a bulk impurity concentration of \(N_A = 10^{15}\) \(\text{cm}^{-3}\), \(J\) becomes,

\[
J = 1.44 \times 10^{-6} F_0^2 \exp \left[ \frac{-\beta(E_0)}{F_0} \right] \text{A cm}^{-2} \tag{6.27}
\]

where,

\[
\beta(E_0) = 4. \frac{\sqrt{2m_{ox}(q\phi - E_0)^3}}{3qh} \tag{6.28}
\]
FIGURE 6.12. The Schottky effect at a metal-semiconductor/dielectric interface
FIGURE 6.13. The Fowler-Nordheim triangular barrier at a metal-semiconductor/dielectric interface

FIGURE 6.14. Linear approximation of the potential barrier at the Si - SiO\textsubscript{2} interface (ref. 56)
FIGURE 6.15 Illustration of the trap states at an n-type and a p-type Si-SiO₂ interface ($V > 0$)

FIGURE 6.16 Illustration of the trap states at an n-type and p-type Si-SiO₂ interface ($V < 0$)
and $F_{ox}$ is the oxide field.

Comparing the multiplier in this expression with those of the Fowler-Nordheim expression, it has been calculated that Eq (6.27) is approximately a factor of 2.5 higher for the conditions given here [56][60].

6.6. THE OXIDE BREAKDOWN MODEL

6.6.1. Continuous voltage stress breakdown

The main features of continuous voltage breakdown were:

a) the significant temperature dependence of the breakdown threshold,

b) the asymmetrical $I_{leak}$ vs. $V_{appl}$ curves for p-type structures compared to more n-type structures,

c) the shift in the minima of the C-V curves in the negative voltage direction with negative applied voltage, and the increase in the minimum value of capacitance with a negligible shift in the minima with positive applied voltage.

The temperature effect implies that electron traps are generated within the oxide under the influence of high electric fields ($\approx 7.5$ MV cm$^{-1}$) and enhanced by temperature [69]. Emission of electrons from impurity or donor sites is, therefore concluded to be of a thermionic field emission nature governed by the Poole-Frenkel effect.
The C-V curves are typical of surface state interactions [1][2][12]. Figure 6.15 shows the effect of interface states on the energy bands of an n-type and a p-type semiconductor. A positive electric field across an n-type capacitor structure causes the bands to bend downwards. Hence, traps at the interface will be filled with electrons, while donor sites will emit electrons to traps (or empty donors) in the oxide bulk under the Poole-Frenkel effect and capture electrons from the Si surface. The trap sites become negatively charged, but the donor sites remain neutral. The increased negative charge at the surface affects the depletion layer of the n-type Si and so reduces its influence on the overall capacitance since [1][2],

\[
\frac{1}{C_s} = \frac{1}{C_{ox}} + \frac{1}{C_{sp}} \quad \text{Eq (6.29)}
\]

where \( C_{sp} \) is the capacitance of the space-charge (depletion) region. A large negative voltage will bend the bands away from the Fermi level (Figure 6.16.), causing electrons to be emitted from trap and donor sites at the interface. A positive charge is formed at the interface which, in p-type capacitors, enhances the action of the gate voltage and attracts electrons to the surface.

The result is that the onset of depletion takes place at a less positive voltage. In n-type structures, the positive charge requires that the applied voltage has to be more negative, in order to obtain inversion. Both structures show C-V curves shifted in the negative direction along the voltage axis.

The asymmetry of the \( I_{\text{leak}} \) vs. \( V_{\text{appl}} \) curves is a consequence of the interface state effect observed in the C-V curves. The n+ polysilicon gate has a Fermi level which coincides with the bottom of the conduction band, and so has a large population of free electrons when in thermal equilibrium. Although the interface between the
FIG 6.17  BAND-BENDING VS. DISTANCE FROM INTERFACE

\[ N = 4 \times 10^{15} \text{ cm}^{-3} \]
polysilicon and the SiO$_2$ is a deposited one, it can be considered to have interface states similar to those at the thermally grown Si-SiO$_2$ interface [1][65][66]. When the gate is the injecting electrode (negative voltages), the conduction bands at the polysilicon-SiO$_2$ interface bend very slightly downwards. In fact, due to the very high doping of the polysilicon gate ($\approx 10^{21}$ cm$^{-3}$), this band-bending can be considered to be negligible. Band-bending at the semiconductor surface depends on the doping concentration and is described in section 6.2.3. Figure 6.17 shows how the bands vary as a function of distance from the interface for $N_A = 10^{16}$ cm$^{-3}$ and $N_D = 10^{16}$ cm$^{-3}$, the bulk acceptor and donor densities, respectively (from Eq (6.9)).

For an n-type surface with a dopant density $N_D = 10^{16}$ cm$^{-3}$, a positive gate voltage results in an accumulation region with a maximum band-bending of 0.33 V. For a p-type surface with a dopant density $N_A = 10^{16}$ cm$^{-3}$, a positive gate voltage will result in an inversion region, the maximum band-bending being 0.97 V. The charge induced at the surface as a result of an applied gate voltage can be obtained using Gauss' Law which gives [2][67],

$$Q_s = \varepsilon_{ox} E_{ox}$$  \hspace{1cm} Eq (6.30)

where $\varepsilon_{ox}$ is the permittivity of the oxide, and $E_{ox}$ is the electric field across the oxide.

For a gate voltage of 40 V across an oxide of thickness 400 Å, the total charge is $Q_s = 3.45 \times 10^{-6}$ Coul cm$^{-2}$ indicating a surface charge density of $n_s = 2.2 \times 10^{13}$ cm$^{-2}$. $n_s$ is constant for the values assumed, and is independent of the doping type or the concentration of charge at the Si surface. However, the spatial depth of the band-bending is greater for p-type than for n-type Si. For n-type Si the electron concentration in the accumulation region can be 5 to 10 times greater than the electron concentration in a p-type inversion layer. Therefore, the number of occupied states in
the conduction band of an n-type accumulation layer is greater than that of a p-type inversion layer with the same applied voltage.

The tunnelling current density is dependent on the electron charge density by means of the supply function $N(E_x)$, since [41][52][58],

$$J \approx e \int_0^\infty D(E_x) \cdot N(E_x) \, dE_x$$

Eq (6.31)

where $E_x$ is the energy of the incident electrons, $D(E_x)$ is the transmission probability and $N(E_x)$ is the supply function, i.e. the number of electrons in the energy range $E_x$ to $E_x + dE_x$ available for tunnelling through the barrier.

$$N(E_x) = C \ln \left\{ 1 + \exp \left[ \frac{(E_F - E_x)}{kT} \right] \right\}$$

Eq (6.32)

where $C$ is a constant and the logarithmic function is based on occupation statistics [52].

The low number of available unoccupied states in the n-type conduction band in accumulation, would also indicate that the probability of reflection at the Si-SiO$_2$ interface is less, increasing $D(E_x)$ [68].

Hence, the tunnelling current for electrons at the Si-SiO$_2$ interface would be greater for an n-type accumulation surface compared to a p-type inversion surface with the same biasing conditions.
The process can be summarised as follows;

Electrons tunnel from the Si into interface states or impurity sites in the SiO₂, initiating the electronic conduction process through the oxide. The higher the concentration of electrons (N_A) in the Si, the greater the leakage, giving rise to the I_leak vs. V_apply curves observed experimentally (Chapter 4).

6.6.2. Pulsed voltage stress conditions

Pulsed voltage stress experiments described in Chapter 3, showed that the positive voltage breakdown thresholds for p-type MOS structures were almost twice as much as the negative voltage thresholds. This difference was reduced as the semiconductor was made n-type, suggesting a charge injection effect. However, the effects of pulsed voltages on the semiconductor surface must be considered.

When a large positive pulse is applied to an n-type structure, the surface is forced towards accumulation. The majority carrier movement is important in accumulation and, as discussed in Section 6.2.4., the response time is almost instantaneous (10⁻¹² secs.). Hence, the Si surface has time to reach strong inversion within the risetime of the voltage pulses used in the experiments. Electrons injected from the strongly accumulated Si surface, together with electrons from traps and impurity sites already in the oxide, tunnel through to the SiO₂ conduction band. Impact ionization (as described in Section 6.4.2.) takes place and results in oxide breakdown [71]. With a large negative voltage, the injecting electrode is the n+ polysilicon gate, with sufficient electrons to be able to match the process described for a positive pulse. The breakdown thresholds are, therefore very similar in the two cases for an n+ type surface.
For a p-type surface, a large positive voltage induces an inversion layer consisting of electrons, at the Si surface. In this case the electrons are minority carriers, and as described in Section 6.2.4. $\tau_{\text{min}} = 0.01 - 1$ sec. which is definitely not within the risetime of the pulse. Initial injection of charge into the SiO$_2$ conduction band, therefore, takes place from impurity sites and trap sites in the SiO$_2$. However, during the decay time of the pulse (dependent on the resistance and capacitance of the oxide; Section 6.2.4), the surface has time to go into strong inversion, but the electric field across the oxide would have decreased. For $\tau_{\text{min}} = 0.01$ secs., $R_{\text{ox}} = 10^{16} \, \Omega$ and $C_{\text{ox}} = 10^{-16} \, \text{F}$, the effective voltage at strong inversion is $\approx 0.9 \, V_{\text{appl}}$. In addition, the tunnelling factors which influenced d.c. breakdown thresholds would also be present here, resulting in an injection current lower than that with an n-type semiconductor. The consequence is a higher breakdown threshold under the same stress conditions. A negative voltage pulse would have a breakdown threshold similar to that of an n+-type structure because the injecting electrode would be the n+ polysilicon gate. Therefore, the positive voltage breakdown threshold of p-type structures would be higher than the negative thresholds. The more p-type the surface, the higher the expected breakdown threshold. There is, however, the consideration of oxide defects which do contribute electrons to the conduction process for the whole duration of the pulse. If there are large enough numbers of these defects, it is possible that they contribute sufficient electrons to reduce the breakdown threshold of the oxide.

6.6.3. Comparison with experimental results

6.6.3.1 Pulsed voltage stress conditions

The model explains the polarity dependence and the electrode dependence of the oxide damage. Impact ionization is a field-dependent mechanism and is not significantly
affected by temperature. The temperature independence of the results is, therefore, supported. The electric fields at which breakdown occurs, $\approx 100 \text{ MV cm}^{-1}$, are sufficiently high to provide electrons with energies of the order required to cause avalanche multiplication in the SiO$_2$ conduction band ($\approx 16 \text{ eV}$).

The presence of defects in the oxide, close to the Si-SiO$_2$ interface, deforms the SiO$_2$ conduction band and enhances the prospect of electron injection directly into the conduction band. This explains the higher breakdown sensitivity of MOS devices with larger oxide areas.

6.6.3.2. Continuous voltage stress conditions

The mechanism presented for breakdown due to constant applied voltages explains the polarity effect and the electrode effect observed. The lower breakdown voltage, compared to that under pulsed conditions, is a result of the longer time periods involved with continuous voltage stress. Sufficient charge can, therefore, be conducted via defect states in the SiO$_2$ forbidden gap to cause breakdown. When the temperature is raised, more electron traps are generated as a result of field-enhanced thermionic processes. More electron conduction takes place and a lower breakdown threshold is observed. Therefore, the temperature dependence is explained by the model.

Leakage currents measured were of the order of $10^{-9} \text{ A}$ at $\approx 35 \text{ V}$, which is comparable with a calculated current of $\approx 10^{-11} \text{ A}$ for a 400 Å MOS structure, in an electric field of $\approx 10^9 \text{ Vm}^{-1}$. This is a reasonable comparison when additional factors which contribute to the leakage current, are considered. The calculation is made using Eq (6.17) of Section 6.5.3.
6.6.4. Comparisons with existing work

Existing work on oxide breakdown is beginning to show more in favour of the charge injection model, as opposed to the field dependent model [72][73]. The experimental data presented in this thesis is consistent with the charge injection model. The mechanism of charge conduction proposed in the model for constant electric fields, in this chapter, is in agreement with that proposed by Wolters [72]. The experimental work showing the temperature dependence of the constant voltage breakdown thresholds, can be favourably compared with the results of Harari [64]. The agreement of the experimental data of the work, in this thesis, on continuous voltage breakdown, with published work, lends confidence to the results obtained under pulsed conditions.

The work on pulsed breakdown appears to be isolated; no literature was found on experimental work on MOS structures, using pulses in the $10^{-6}$ secs. to $10^{-9}$ secs. range. The model is consistent with the charge injection models considered for constant voltage breakdown, while taking into account the higher electric fields ($\sim 100$MVcm$^{-1}$) and the shorter duration of the pulses. Impact ionization has been refuted by some authors [72] for breakdown in continuous electric fields. However, the fields under pulsed conditions are very much higher. Experimental work on tunnelling through thin oxides [60], and theoretical work on high field effects at the semiconductor surface [54][57], do tend to support the model proposed for pulsed breakdown proposed in this chapter.

6.7. SUMMARY

1. The silicon MOS structure has been described with respect to its physical properties, energy bands and electronic conduction processes.
2. Qualitative models have been presented for $\text{SiO}_2$ breakdown, under continuous and pulsed conditions. These models are based on existing solid state physics theory and are supported by the experimental data in Chapter 4.

Continuous voltage conditions

3. Oxide breakdown under continuous voltage stress is the result of charge injection from the Si surface or the n+ polysilicon gate. Charge injection at the Si-$\text{SiO}_2$ interface takes place by quantum mechanical tunnelling through the potential barrier, or, by thermionic excitation over it. Electrons may tunnel directly into the $\text{SiO}_2$ conduction band or into traps and impurity sites within the $\text{SiO}_2$ bulk/interface. C-V plots suggest that tunnelling to trap/impurity sites is more likely than tunnelling directly into the conduction band.

4. Once in these trap sites, electrons are able to move further into the oxide by means of the Poole-Frenkel effect. Emission over the barriers presented by the Coulombic potential wells at these sites.

5. n-type silicon can inject more charge into $\text{SiO}_2$ than p-type silicon, because of the positions of the Fermi levels and the greater probability of tunnel emission at the surface.

6. Since the $\text{SiO}_2$ conduction process under d.c. voltage stress, is thermionic, it is temperature dependent. In addition, combined temperature and field effects can generate further trap sites by affecting any weak bonds in the $\text{SiO}_2$.

Pulsed voltage conditions
7. Pulsed voltage breakdown is initiated by direct tunnelling of electrons from the cathode into the SiO₂ conduction band. There they undergo impact ionization through collision with the lattice, causing an avalanche multiplication of electrons in the conduction band.

8. When the injecting surface is in accumulation, the very fast response times of majority carriers enables them to contribute to the impact ionization process within the risetime of the applied pulse.

9. When the injecting surface is in inversion, the response time is very much slower than the risetime of the pulse, and the effective electric field in strong inversion is lower than the actual applied voltage pulse. Also, the tunnelling probability from inverted surfaces is less than that from accumulated surfaces for the same applied voltage.

10. Tunnelling into the SiO₂ conduction band can be enhanced by resonance tunnelling in oxide trap sites, suggesting that defects will influence breakdown under pulsed voltage conditions.

11. The actual breakdown mechanism is a result of the collision of electrons, travelling in the SiO₂ conduction band or between defect sites, with the SiO₂ lattice. Energy is transferred from the electrons to the lattice by means of phonons.

12. The breakdown process under pulsed voltage conditions is not influenced by temperature. This is supported by the experimental data on MOSFETs presented in Chapter 4.
6.8 REFERENCES


CHAPTER 7

DISCUSSION AND CONCLUSIONS

7.1 INTRODUCTION

Chapters 2 to 6 have identified and examined failure mechanisms in semiconductor devices with detailed work being particularly related to the effects of electrical overstress in MOS devices. The purpose of this chapter is to show the unified nature of the approach used in tackling this project. Major points made in each chapter are emphasised here and discussed in relation to this project as a whole.

In Chapter 2, the principal failure mechanisms in semiconductor devices were identified by means of a combined literature and industrial survey. Chapters 3, 4 and 5 detail experiments conducted into the most common failure mechanism, namely Electrical Overstress/Electrostatic Discharge Damage, with a view to gaining a deeper understanding of the physical processes involved in the damage. A model of the breakdown mechanisms is described in Chapter 6 enabling the analysis to be extended to predict the influence of similar stress factors on other types of MOS devices.

7.2. SURVEY ON FAILURE MECHANISMS IN SEMICONDUCTOR DEVICES

As a result of this survey it became evident that although many failure mechanisms exist, one mechanism dominated device reliability. Over 50% of all failures both in the field and in laboratory lifetests were related to EOS/ESD. Although listed as an event-dependent failure mechanism, EOS/ESD has become of increasing concern to
both semiconductor and equipment manufacturers because of the complex, state-of-the-art applications required of semiconductor devices in modern industry. Solutions to the problem were centred on protection circuitry capable of dissipating any excess voltages (within the design limits of the protection circuit) before they reached the device stress points. Therefore an area in need of further research was revealed.

A study on failure mechanisms could not be totally divorced from the concepts of reliability modelling and reliability assurance. Analysis of these topics led to the conclusion that reliability modelling must be considered purely as an indicator of device reliability. The uncertainty factors involved in any mathematically-based model made a mockery of the very precise reliability figures provided by these models. Similarly, conditions under which laboratory lifetests were conducted and the methods used in extrapolating the data to real time-scales and operating environments, meant the figures obtained should only be used as a guideline to device reliability. Furthermore, particular applications favoured particular models, and data obtained from an environmentally benign model such as those developed for telecommunication or computer applications, may bear no relationship to actual failure rates of devices used in the automobile or aircraft industry. It becomes apparent that a universal model, such as that presented in MIL-HDBK-217D, must necessarily be pessimistic, and even heavily pessimistic, when applied to less demanding environments, compared to actual failure data.

Increasing complexity in integrated circuits and the large number of functions they are required to incorporate, make it very difficult to "test-in" quality. In fact it has been increasingly acknowledged that "test-in" quality was both uneconomical and impractical. The solution is "built-in" quality, or the principle of "getting-it-right-first-time". This requires that the entire workforce from the management levels downwards must be committed to producing good devices first time. Furthermore, close
cooperation between the consumer and the device manufacturer, can ensure that more complex devices are functionally tested for the specific purpose for which they are to be used. Such a policy would reduce test costs and also provide a higher guaranteed reliability to the consumer. Finally, there is the trend towards "design-for-testability" whereby chips are designed with functional test programmes in mind. On the whole it appears that the problem of testing complex i.c.'s may prove to be one of limitation by the actual size of an integrated circuit. The advent of automatic test equipment should phase out the human interface on the semiconductor device production line, and therefore increase the quality of manufactured devices.

In order to raise the inherent reliability of a semiconductor device it is necessary to understand and eliminate the failure mechanisms which affect that device. This requires a knowledge of the physics of the failure process if the whole thing is not to become a black art. An understanding of failure physics also enables predictions to be made of the possible effects of any changes in the device structure or fabrication techniques on device performance and reliability. Hence, it becomes possible to evaluate devices which can be considered to still be in their infancy, as far as the manufacturing technology is concerned, with regard to their behaviour under various stress conditions.

7.3. EXPERIMENTS ON CONTINUOUS AND PULSED VOLTAGE STRESSING OF MOS STRUCTURES

7.3.1. Pulsed voltage stressing of enhancement mode n-channel MOSFETs

Breakdown threshold voltages under pulsed (or ESD) stress conditions are not well defined quantities. Because of the number of factors influencing breakdown,
particularly with regard to the mechanisms involved in the breakdown process, it is possible for identical devices to respond differently to a given ESD pulse. However, taken over a large sample size, distinct trends can be observed which enable the overall response of a given device type to be established.

The main conclusions drawn from this experiment were that ESD damage in discrete enhancement-mode n-channel MOSFETs was

(i) Voltage dependent
(ii) Temperature independent
(iii) Dimension dependent

It was also observed that sequential pulsing increased the degradation in damaged devices and eventually resulted in catastrophic failure, although the extent of the degradation after each pulse could not be predetermined.

The implications of this work are significant, and these experiments can be looked upon as a springboard from which an in-depth investigation into the physics of EOS/ESD damage can be launched. Firstly, the mechanism involved in pulsed voltage damage, which in most cases is due to high voltages (up to 20kV, but usually around 100 V) with risetimes between $10^{-8}$ secs. and $10^{-6}$ secs. is that of impact ionization. Therefore, when considering ESD sensitivity of devices, it becomes necessary to consider the influences of the stress factors whatever they may be, on the breakdown process, which includes the tunnel injection of carriers into the SiO$_2$ from both the Si and existing oxide defects. To this end, minimizing the number of traps during the fabrication of thermally grown oxides, could mean lower ESD sensitivity. Secondly, the degradation effect on the devices is important. It was observed that degraded devices had electrical characteristics which would have passed screening tests based
on a pass/fail system. However, these devices were now able to contribute to the scores of latent failures which usually occur during their operational life. Sequential pulsing of damaged devices confirmed that this could happen. Many field failures attributed to EOS could well have originated with latent ESD damage [1]. A third factor was highlighted by the temperature independence of the damage. Experimental work on packaged i.c.'s have found that the input protection circuit introduces a factor affecting reliability which is not present in the device itself, i.e. the temperature dependence [2]. Therefore, if the breakdown mechanism can be inhibited either by improved fabrication techniques [6] or by changing the MOSFET structure to reduce high field charge injection into the oxide, then the improved i.c.'s will be more economical (no large input protection circuit) and have comparable or even better reliability than present i.c.'s.

7.3.2. Pulsed voltage stressing of other MOS structures

The aim of these experiments was to expand the pulsed voltage breakdown analysis to cover all standard MOSFET structures. Although p-channel depletion mode MOSFETs were not available, results from the other three structures are extrapolated to predict the behaviour of these devices under similar stress.

Essentially, both the D-mode n-channel MOSFETs, and the E-mode p-channel MOSFETs showed degradation behaviour of similar nature to that observed in E-mode n-channel MOSFETs. The trend towards degradation is very significant, and since the proposed breakdown models for E-mode NMOS transistors do apply to these devices as well, such a large sample size was not required in order to evaluate their behaviour under the various stress factors. E-mode p-channel MOSFETs that were subjected to ESD pulses showed degradation characteristics very similar to the E-mode n-channel MOSFETs. A lower $I_{DS}$ was observed for a given $V_{DS}$ at the same gate voltage.
Degradation characteristics in D-mode n-channel MOSFETs varied slightly because of the implanted inversion-layer already present in the channel as discussed in Chapter 5.

The polarity dependence of pulsed breakdown was observed in experiments on MOS gate-oxide capacitor structures on the same wafers as the MOSFETs. Positive voltage breakdown thresholds were twice the negative voltage breakdown thresholds for p-Si capacitors. n-Si capacitors showed about 50% larger positive voltage breakdown thresholds compared to the negative voltage breakdown thresholds. For n+–Si capacitors the difference in breakdown thresholds was almost negligible. In all cases, the breakdown thresholds with negative voltage were about the same (-100V) while the positive voltage breakdown varied according to the type of semiconductor. This signified a dependence on the dopant concentration in the silicon. Results with continuous voltage (d.c.) stressing confirmed that this was indeed the case.

Capacitors were also subjected to positive ESD pulses below their known (experimentally determined) breakdown thresholds, and C-V curves made. Comparisons with C-V curves made before application of the pulse showed that:

(i) in the case of the p-Si and n-Si capacitors, no noticeable change had occurred, within the bounds of experimental accuracy.

(ii) n+–Si capacitors had a very much higher minimum capacitance than before the application of the pulse.

The latter result indicates that charge trapping has taken place at the Si-SiO₂ interface, which, with larger pulses, will enhance the prospect of oxide breakdown. The fact that no change was observed in the NMOS devices, together with their higher positive breakdown thresholds, implied that the injection of carriers into the oxide was important to the breakdown process.
7.3.3. Continuous voltage stressing of MOS Structures

The very short durations of the pulses (≈10^{-7} secs) meant that the breakdown mechanisms were time-independent. Under time-dependent continuous (d.c.) voltage conditions different mechanisms become important, and different breakdown thresholds may be expected. Experimentally, d.c. stress on MOS capacitors showed that this was indeed the case. The breakdown thresholds with both positive and negative voltage stresses were in the region of 35 V, almost five times less than the pulsed thresholds. A polarity dependence was still observed, however, as a function of the semiconductor dopant concentration. The p-Si capacitors showed symmetrical $I_{\text{leak}}$ vs. $V_{\text{appl}}$ curves, whereas n-Si capacitor curves were less symmetrical, and the p-Si capacitors were very much asymmetrical. Carrier injection was still, therefore, important to the breakdown process. The significant temperature dependence was in agreement with the time-dependent nature of the stress voltage, and indicated that the breakdown process was related to thermionic conduction processes within the SiO$_2$. Thus the breakdown mechanism was concluded to be due to hopping conduction via impurity sites and electron traps in the oxide. Defects of this kind could be generated by the high fields (≈10^{7} V cm^{-1}) and enhanced by high temperature (≈200°C), by which weak Si-Si or Si-O bonds could be dissociated.

"Freeze frame" pictures of the breakdown processes were obtained by making C-V plots after applying voltage stresses below the breakdown thresholds. Shifts in the capacitance minima and distortion of the curve confirmed that traps did have a significant role in the breakdown mechanism.

7.4. THE OXIDE BREAKDOWN MODELS
7.4.1. Theoretical analysis

The theoretical analysis presented in Chapter 4, is based on two aspects of the high-field behaviour of MOS structures. Charge injection at the oxide interfaces and electronic conduction process in the oxide itself. While the actual theoretical backgrounds of the subjects have been well developed, their effects on the oxide have always been considered in isolation of each other. The qualitative models presented earlier in this work consider the two phenomena as interacting to cause high field oxide breakdown. In essence, electrons are injected into the oxides, either at the oxide interfaces or from donor impurity sites within the oxides, by tunnelling or thermionic field emission (i.e. Poole-Frenkel effect). Conduction through the oxide can either be directly in the conduction band, or by hopping between impurity sites in the SiO$_2$ forbidden gap.

The models are based on the experimental data in Chapter 4. The time-dependent feature of the pulsed voltage stress was also considered together with the time-dependence of d.c. breakdown. Hence, the model also includes minority/majority carrier response times in the silicon, the thermal generation of defects and thermally-activated emission from defects in the SiO$_2$.

Quantum-mechanical tunnelling theory shows that the tunnelling probability is dependent on the transmission coefficient at the barrier, and also the supply function, i.e. the number of electrons available for tunnelling. High electric fields cause severe band-bending at the surface, leading to energy quantisation effects at the semiconductor surface which influences the transmission coefficient. The extent of the band-bending varies for different doping concentrations in the silicon. This means that the supply function for an n-type silicon surface will be higher than that of a p-
type surface, thus influencing the amount of injected charge in p-Si and n+-Si structures under identical stress factors.

The influence of the Fermi level on the injection of charge to interface states was considered. Since the energy levels of the interface states extend uniformly across the forbidden gap of silicon, the higher the Fermi level, with respect to the valence band, the greater probability there is of these states being occupied. High electric fields under d.c. conditions cause these carriers to be emitted to traps further in the oxide bulk, leaving empty states, thereby enhancing the probability of charge injection at the interface.

Oxide traps are also known to increase the probability of tunnelling. Resonance tunnelling, whereby, traps affect the electronic wave function, means that oxide defects contribute to the charge injection and conduction process in SiO₂ and lower the oxide breakdown thresholds.

7.4.2. Model of oxide breakdown under pulsed voltage stress

In modelling pulsed breakdown, the effect of the applied voltage on the silicon surface becomes an important consideration. Two states are possible under high field conditions;

(i) the surface is in accumulation
(ii) the surface is in inversion

Since the n+ polysilicon gate has a very large concentration of electrons (≈10²¹ cm⁻³) charge injection with negative applied voltages can be considered identical to that of an n-type surface in heavy accumulation. Positive voltage pulses will see different
behaviour from surfaces in heavy accumulation (n-type) and surfaces in heavy inversion (p-type).

Upon application of a large positive voltage pulse to an MOS device, for an accumulated surface (n-type silicon) the majority carrier response time is important. Since this is very fast (≈10^{-12} secs.) the surface charge will respond almost instantaneously to pulses in the region of 10^{-7} sec. risetimes. Charge injection into the oxide then takes place, and the field at breakdown (≈10^9 V cm^{-1}) causes electrons to be injected directly into SiO$_2$ conduction band. Here the electrons gain sufficient energy from the electric field to cause impact ionization which leads to lattice heating and eventually breakdown. The time scales involved in this process are very short. Impact ionization to breakdown proportions is possible in <10^{-9} secs.. This is well within the risetimes of any standard ESD pulses. It is also possible that the high electric fields can cause trapped electrons to tunnel through the coulombic potential barriers into the SiO$_2$ conduction band and contribute to the breakdown process.

Inversion in p-type devices requires minority carriers to respond to the applied voltage pulse. The response times of minority carriers is slow compared to majority carriers, typically about 10^{-3} secs. Therefore, when the surface is in strong inversion, the effective voltage across the oxide would have decreased (≈0.9 V$_{appl}$). Tunnelling probabilities and field emission will be less than in an accumulated device with the same surface conditions and the breakdown threshold is increased. Impurities and trapped electrons in the bulk oxide will however still be able to tunnel through to the oxide conduction band, as in the accumulated case, and therefore can influence breakdown if available in sufficient quantities.

7.4.3. Model of oxide breakdown under continuous voltage stress
Under d.c. voltage stress, electrons quantum-mechanically tunnel from the silicon electrodes into the oxide. The tunnel current is dependent upon the surface charge concentration, higher injection being obtained with more n-type surfaces. Conduction through the oxide is by hopping via impurity sites and electron traps in the bulk oxide. Hopping conduction is a thermionic field-enhanced mechanism indicating that the breakdown threshold is lowered under high temperature and high voltage conditions. The thermal energy gained by the lattice under high fields can result in dissociation of Si-Si, Si-O and O-O bonds at weak points in the lattice, generating more defect sites [5][9][10]. Such a process would have a finite lifetime (> 1 sec), but under d.c. stress conditions the time scales would be long enough to encourage defect generation [7][8].

Breakdown in the d.c. case, therefore, takes place due to a combination of impact ionization and hopping conduction. Electrons may tunnel directly from the silicon conduction band to the SiO₂ conduction band, but are most likely to tunnel to trap sites close to the Si-SiO₂ interface and to travel through the oxide by hopping conduction. The number of defects in the oxide play a considerable role in the breakdown process, and a lower number of defects would mean a lower oxide conduction current, and a higher breakdown threshold.

7.5. FUTURE WORK

Future work can be described in three areas;

(i) investigations into pulsed voltage breakdown effects,
(ii) quantitative modelling of the oxide breakdown mechanisms,
(iii) experimental verification of the predicted breakdown effects in silicon submicron MOSFETs.

(i) Investigations into pulsed voltage breakdown effects

Analysis of the pulsed breakdown mechanisms showed that MOS damage was related to the silicon dopant concentration as well as the majority/minority carrier response times at the surface under pulse conditions. The influence of oxide defects on breakdown thresholds is also of importance. Future work would focus on these effects.

Experimentally, the majority/minority carrier response times could prove to be a limiting factor in pulsed breakdown. For instance, a positive voltage across a p-type MOS capacitor achieves full inversion after ~0.01 secs. Therefore the actual rise time of the pulse (< 10^{-6} secs.) is not important, while the decay time does become important. Extending this theory to p-n junctions, it will be possible to evaluate the necessity in obtaining very fast risetimes in ESD test equipment and the influence of faster risetimes on breakdown mechanisms. Since, pulsed breakdown is attributed to impact ionization, the time required to generate sufficient secondary electrons to initiate breakdown could also be significant. Therefore it may be that a pulse with a very fast risetime (< 10^{-12} secs.) does not result in breakdown until well into the decay period.

The influence of oxide defects on pulsed breakdown can be investigated further. Examining the pulsed sensitivity of devices with oxides of different defect concentrations ranging from high to very low (e.g. radiation hardened) would enable a more quantitative understanding of the effects to be gained. Methods of reducing the pulsed sensitivity of MOSFETs can be analysed. Such methods would involve reducing the probability of charge injection at the Si-SiO_2 interface. To achieve this it
would be necessary to either increase the Si-SiO₂ work function perhaps by introducing an intermediate material at the Si-SiO₂ interface or by moving the conducting channel to the device bulk (a two-dimensional epitaxial interface, still influenced by the gate potential is a highly speculative possibility). Improving the integrity of the oxide by reducing the number of defects, as is currently achieved in radiation-hardened structures could increase device reliability to EOS/ESD. Possible defects caused by ion implantation, for example, could be minimised by using a pre-implanted oxide layer. The effectiveness of such techniques would need to be experimentally proven on test samples.

(ii) Quantitative modelling of the oxide breakdown mechanisms.

The models presented in this thesis are qualitative in that they do not lay a mathematical base for the breakdown mechanisms. A quantitative model would include mathematical treatment of the tunnelling equations to the Si-SiO₂ interface as well as between trap-sites within the oxide. Coupled with analysis of the Poole-Frenkel effect under these conditions it will be possible to obtain mathematical expressions capable of predicting breakdown voltages for given MOS structures, under both pulsed voltage and continuous voltage stress conditions. Silicon dopant concentrations, oxide defect densities and device dimensions should all be incorporated in the final expression for the breakdown threshold.

Experimental verification of these calculations should coincide with the results presented in this work for the type of devices used here.

(iii) Experimental verification of the predicted breakdown effects in silicon submicron MOSFETs.
The predicted breakdown effects in submicron MOSFETs, as outlined in Appendix 3, could not be verified because of the unavailability of suitable samples. Small gate dimension capacitors with varying silicon dopant concentrations and submicron size MOSFETs would be required to obtain confirmation of the predictions. This work could obviously be considered as an extension of (i), the samples being used for both experiments. The quantitative model outlined in (ii) above would also be supported by experimental work on submicron MOS devices.

7.6. SUMMARY

1. A survey of the most common failure mechanisms in semiconductor devices was made. EOS/ESD damage was found to be the most frequent cause of failure accounting for over 50% of all observed damage.

2. The effects of EOS/ESD damage on MOSFETs were studied using continuous and pulsed voltages simulating the stress conditions. Hence it was possible to identify the physical processes involved in the breakdown mechanisms in these conditions.

3. A qualitative model of oxide breakdown in MOS structures was developed, based on the experiments. Two different mechanisms were proposed for continuous and pulsed voltage breakdown.

4. Continuous voltage breakdown is the result of hopping conduction of charge carriers through the oxide, via electron traps and impurity sites. The mechanism is both voltage and temperature dependent.

5. Pulsed voltage breakdown is the result of electron conduction in the SiO₂ conduction band, having been injected directly into the band via quantum-
mechanical tunnelling. Impact ionization and avalanche multiplication of carriers under the high electric fields then leads to breakdown. The mechanism is voltage dependent but independent of temperature.

6. Both mechanisms were found to be dependent on the surface charge concentration at the SiO₂ interface and therefore polarity dependent. The charge injection process into the SiO₂ is important to the breakdown mechanisms.

7.7. REFERENCES


APPENDIX 1

SPECIFICATIONS OF WAFERS No: 1 AND 2

1.1. WAFER No: 1 - PRINCIPAL PROPERTIES

1. Diameter of silicon wafer = 3 inches
2. Material = p-type
3. Orientation = < 100 >
4. Resistivity = 20 Ω cm
5. Field implant = Boron B⁺
6. Implant Boron = 4.0 x 10¹⁵ cm⁻³
7. Implant Phosphorous = 1.0 x 10¹⁵ cm⁻³
8. n⁺ doping = 1.0 x 10²¹ cm⁻³
9. Field Oxidation = 0.5 to 0.7 μm at 900°C
10. Gate Oxide (thermally grown) = 425 Å at 950 °C
11. Polysilicon gate (n⁺ implanted) = 4552 Å at 620 °C
12. Metallisation = Al sputtered, 1.5% Si/Al/Cu
13. Passivation = Silicon Glass (SILOX),
                  10485 Å at 400 °C
14. Anneal = 425 °C
15. Strip resist using fuming HNO₃
16. Correction factor to channel
    length due to dopant diffusion (ΔL) = 1.067 μm
17. Correction factor to gate width
    due to field oxide inset (ΔW) = 1.512 μm
18. Threshold voltage varies between 0.75 volts and 1.38 volts depending on dimension.

19. Junction depth = 0.4 μm

20. Capacitors; polysilicon (n⁺) -gate oxide - silicon;

   Area = 49.0 x 10³ μm²

   Periphery = 910 μm

21. Transistor dimensions are shown in TABLE A1.1.
<table>
<thead>
<tr>
<th>1</th>
<th>W 1.4</th>
<th>L 1.4</th>
<th>W 1.4</th>
<th>L 2.1</th>
<th>W 1.4</th>
<th>L 3.2</th>
<th>W 1.4</th>
<th>L 3.5</th>
<th>W 2.1</th>
<th>L 2.1</th>
<th>W 1.4</th>
<th>L 102.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>W 2.1</td>
<td>L 2.5</td>
<td>W 2.1</td>
<td>L 3.2</td>
<td>W 2.1</td>
<td>L 3.5</td>
<td>W 2.1</td>
<td>L 1.4</td>
<td>W 2.5</td>
<td>L 2.1</td>
<td>W 2.5</td>
<td>L 3.2</td>
</tr>
<tr>
<td>3</td>
<td>W 2.5</td>
<td>L 4.2</td>
<td>W 2.5</td>
<td>L 4.6</td>
<td>W 2.5</td>
<td>L 5.3</td>
<td>W 2.5</td>
<td>L 10.2</td>
<td>W 2.5</td>
<td>L 20.7</td>
<td>W 3.2</td>
<td>L 1.4</td>
</tr>
<tr>
<td>4</td>
<td>W 3.2</td>
<td>L 2.5</td>
<td>W 3.2</td>
<td>L 3.2</td>
<td>W 3.2</td>
<td>L 3.5</td>
<td>W 3.2</td>
<td>L 4.2</td>
<td>W 3.2</td>
<td>L 4.6</td>
<td>W 3.5</td>
<td>L 1.4</td>
</tr>
<tr>
<td>5</td>
<td>W 3.5</td>
<td>L 3.2</td>
<td>W 3.5</td>
<td>L 3.5</td>
<td>W 3.5</td>
<td>L 4.2</td>
<td>W 3.5</td>
<td>L 4.6</td>
<td>W 3.5</td>
<td>L 6.3</td>
<td>W 3.5</td>
<td>L 20.7</td>
</tr>
<tr>
<td>6</td>
<td>W 4.2</td>
<td>L 2.5</td>
<td>W 4.2</td>
<td>L 3.2</td>
<td>W 4.2</td>
<td>L 3.5</td>
<td>W 4.2</td>
<td>L 4.2</td>
<td>W 4.2</td>
<td>L 4.6</td>
<td>W 4.6</td>
<td>L 4.6</td>
</tr>
<tr>
<td>7</td>
<td>W 4.6</td>
<td>L 2.5</td>
<td>W 4.6</td>
<td>L 3.2</td>
<td>W 4.6</td>
<td>L 3.5</td>
<td>W 4.6</td>
<td>L 5.3</td>
<td>W 6.3</td>
<td>L 2.5</td>
<td>W 10.2</td>
<td>L 10.2</td>
</tr>
<tr>
<td>8</td>
<td>W 10.2</td>
<td>L 2.5</td>
<td>W 20.7</td>
<td>L 3.2</td>
<td>W 20.7</td>
<td>L 2.5</td>
<td>W 102.9</td>
<td>L 1.4</td>
<td>W 102.9</td>
<td>L 2.1</td>
<td>W 102.9</td>
<td>L 5.3</td>
</tr>
<tr>
<td>9</td>
<td>W 102.9</td>
<td>L 2.5</td>
<td>W 102.9</td>
<td>L 3.2</td>
<td>W 102.9</td>
<td>L 3.5</td>
<td>W 102.9</td>
<td>L 4.2</td>
<td>W 102.9</td>
<td>L 4.6</td>
<td>W 102.9</td>
<td>L 6.3</td>
</tr>
<tr>
<td>10</td>
<td>W 2.5</td>
<td>L 3.5</td>
<td>W 3.5</td>
<td>L 3.5</td>
<td>W 4.2</td>
<td>L 3.5</td>
<td>W 4.6</td>
<td>L 3.5</td>
<td>W 6.3</td>
<td>L 3.5</td>
<td>W 10.2</td>
<td>L 3.5</td>
</tr>
<tr>
<td>11</td>
<td>W 2.5</td>
<td>L 10.2</td>
<td>W 3.5</td>
<td>L 10.2</td>
<td>W 4.6</td>
<td>L 10.2</td>
<td>W 6.3</td>
<td>L 10.2</td>
<td>W 10.2</td>
<td>L 10.2</td>
<td>W 102.9</td>
<td>L 10.2</td>
</tr>
</tbody>
</table>

**TABLE A1.1.**

Gate width (W) and gate length (L) dimensions of a block of transistors on a chip in Wafer No: 1.
1.2. WAFER No: 2 - PRINCIPAL PROPERTIES

1. Diameter of silicon wafer = 3 inches
2. Material = n-type
3. Orientation = < 100 >
4. Resistivity = 25 Ω cm
5. Field Implant
   - n-channel, implant Boron
     = 1 x 10^{15} cm^{-3}
   - p-channel, implant Arsenic
     = 1 x 10^{15} cm^{-3}
6. Source/Drain doping, Boron (p⁺)
   = 5 x 10^{19} cm^{-3}
   Arsenic (n⁺)
   = 2 x 10^{20} cm^{-3}
7. Field Oxidation = 7200 Å
8. Gate Oxide = 320 Å
9. Threshold voltage for PMOS is approximately = - 0.8 volts
10. Junction Depth, n-channel = 0.35 μm
    p-channel = 0.7 μm
11. Capacitors; polysilicon (n⁺) - gate oxide-silicon,
    Area = 4.9 x 10^{3} μm^2
    Periphery = 280 μm
12. MOSFET Dimensions are as shown in TABLE A1.2.
<table>
<thead>
<tr>
<th>P1</th>
<th>W 20.0</th>
<th>W 20.0</th>
<th>W 20.0</th>
<th>W 20.0</th>
<th>W 20.0</th>
<th>W 20.0</th>
<th>W 20.0</th>
<th>W 20.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1</td>
<td>L 1.0</td>
<td>L 1.25</td>
<td>L 1.5</td>
<td>L 1.75</td>
<td>L 2.0</td>
<td>L 2.25</td>
<td>L 2.5</td>
<td>L 3.0</td>
</tr>
<tr>
<td>P2</td>
<td>W 2.5</td>
<td>W 2.5</td>
<td>W 2.5</td>
<td>W 2.5</td>
<td>W 20.0</td>
<td>W 20.0</td>
<td>W 20.0</td>
<td>W 20.0</td>
</tr>
<tr>
<td>N2</td>
<td>L 1.25</td>
<td>L 1.5</td>
<td>L 2.0</td>
<td>L 2.5</td>
<td>L 20.0</td>
<td>L 7.5</td>
<td>L 5.0</td>
<td>L 3.5</td>
</tr>
<tr>
<td>P3</td>
<td>W 1.0</td>
<td>W 1.25</td>
<td>W 1.5</td>
<td>W 1.75</td>
<td>W 2.0</td>
<td>W 2.25</td>
<td>W 2.5</td>
<td>W 3.0</td>
</tr>
<tr>
<td>N3</td>
<td>L 10.0</td>
<td>L 10.0</td>
<td>L 10.0</td>
<td>L 10.0</td>
<td>L 10.0</td>
<td>L 10.0</td>
<td>L 10.0</td>
<td>L 10.0</td>
</tr>
<tr>
<td>P4</td>
<td>W 1.5</td>
<td>W 1.5</td>
<td>W 1.5</td>
<td>W 1.5</td>
<td>W 20.0</td>
<td>W 10.0</td>
<td>W 5.0</td>
<td>W 3.5</td>
</tr>
<tr>
<td>N4</td>
<td>L 1.25</td>
<td>L 1.5</td>
<td>L 2.0</td>
<td>L 2.5</td>
<td>L 10.0</td>
<td>L 10.0</td>
<td>L 10.0</td>
<td>L 10.0</td>
</tr>
</tbody>
</table>

**TABLE A1.2.**

Gate width (W) and Gate Length (L) dimensions in μm of a block of MOSFETs on chip on Wafer No: 2. P1, P2, P3 and P4 are p-channel devices. N1, N2, N3 and N4 are n-channel devices.
APPENDIX 2

SOLUTION OF THE SCHROEDINGER EQUATION FOR TUNNELLING AT A POTENTIAL BARRIER

The 1-D time-independent Schrödinger Equation for the metal-dielectric interface shown in Figure A2.1 is given by:
\[-\frac{\hbar^2}{2m} \cdot \frac{d^2\psi}{dx^2} + \left\{ V(x) - E_x \right\} \psi = 0 \quad \text{Eq (A2.1)}\]

where

\[V(x) = -F_x - \Delta\phi \quad \text{Eq (A2.2)}\]

\(F_x\) is the applied electric field,
\(\Delta\phi\) is the barrier lowering due to the image force at the interface.

\[\Delta\phi = \left(\frac{e^3 F_x}{\varepsilon}\right)^{1/2} \quad \text{Eq (A2.3)}\]

where \(\varepsilon\) is the permittivity of the dielectric

Solving for \(\psi(x)\) from Eq (A2.1)

\[\psi(x) = a \left\{ e^{ik_x x} + Re^{-ik_x x} \right\}, \quad x < 0\]

\[= T \left\{ Bi(-y) + Ai(-y) \right\}, \quad x > 0 \quad \text{Eq (A2.4)}\]

where

\[k_x = \left(\frac{2m E_x}{\hbar^2}\right)^{1/2} \quad \text{Eq (A2.5)}\]
and
\[ y = a^{1/2} \left( \frac{(E - q \phi)}{qF} + x \right) \]  
Eq (A2.6)

\[ a = \frac{2m q F}{\hbar^2} \]  
Eq (A2.7)

The functions $\text{Ai}(-y)$ and $\text{Bi}(-y)$ are two real linearly independent solutions to Airy's equation.

The probability current normal to the interface is given by
\[ p = \frac{j \hbar}{2m} \left\{ \gamma^* \frac{\delta \gamma^*}{\delta x} - \gamma \frac{\delta \gamma}{\delta x} \right\} \]  
Eq (A2.8)

\[ = \frac{\hbar k_x}{m} \left| T \right|^2 \]

The transmission probability, $D$, of the barrier is defined as the ratio of this current to the probability current $P_0 = \hbar k/m$ in the incident plane wave.
For a triangular barrier,

\[ D(E_X) = \frac{P}{P_0} \quad \text{Eq (A2.9)} \]

The current per unit area can be written as

\[ J = q \int_0^\infty D(E_X) N(E_X) dE_X \quad \text{Eq (A2.10)} \]

where,

\[ N(E_X) dE_X \] is the supply function representing the number of electrons in the energy range \( (E_X, E_X + dE_X) \) which are available for tunnelling through the barrier. Assuming that all electrons escaping from region 1 are swept away by the field in region 2.

This leads to the tunnelling current, calculated following the conventional Fowler-Nordheim derivation and using bulk statistics, of

\[ J = \frac{e^3 F^2}{8\pi \hbar \phi} \exp \left\{ \frac{-4\sqrt{2m}}{3\hbar eF} \phi^{3/2} \right\} Am^{-2} \quad \text{Eq (A2.11)} \]
References

1. FOWLER R., NORDHEIM L., Electron Emission in Intensive Electric Fields

2. DUKE C.B., ALFERIEFF M.E., Field Emission through Atoms Adsorbed on a Metal Surface


4. WEINBERG Z.A., Tunneling of electrons from Si into thermally grown SiO₂.
APPENDIX 3

DEGRADATION MECHANISMS IN SILICON SUBMICRON MOSFETS

A3.1. INTRODUCTION

As device geometries get smaller, the properties of the devices are affected by the changes in the structures and fabrication techniques required to achieve optimum performance at smaller dimensions. Very small dimension (VSD) MOS device can be classified as devices with gate widths and length $< 5 \, \mu m$. It is at these sizes that the electrical characteristics of the device begin to differ from those expected of a standard device. Dimensions of less than $1 \, \mu m$ constitute what are known as submicron devices. Silicon MOS submicron devices are expected to form the backbone of the future generation of semiconductor technology. Hence, an understanding of possible degradation mechanisms, which may be introduced by the manufacturing and design techniques employed, is important to their development. However, due to the difficulty at present in obtaining devices manufactured to submicron dimensions, much of the experimental work is carried out on VSD devices and extrapolated to gate lengths and widths of less than $1 \, \mu m$. Accurate computer simulation, necessarily involving complex 3-D numerical analysis, is another means of investigating the behaviour of submicron structures. The data gathered by these methods can be coupled with theoretical studies into the physics of the devices and their properties, to enable the behaviour to be effectively predicted under various operating conditions.
It is the intention in this chapter, to outline the principal mechanisms of degradation observed in VSD devices, and to apply them to submicron silicon MOSFETs. Particular emphasis will be placed on the effects of the oxide breakdown mechanisms proposed in Chapter 6, on the behaviour of submicron MOSFETs under stress conditions. Important features of submicron silicon MOSFETs are discussed in this chapter. A brief description is given of three dimensional computer simulation of VSD MOSFETs, and due to the difficulties in obtaining or developing a suitable 3-D program for this project, published results are presented. It then goes on to discuss known degradation mechanisms caused by the small dimensions involved, and finally extends the oxide breakdown analysis to submicron MOSFETs.
A3.2. IMPORTANT STRUCTURAL FEATURES OF VERY SMALL DIMENSION (VSD) MOS DEVICES

A3.2.1. Dimensions

The gate dimensions of a VSD device are of critical importance to the electrical properties of the device\(^1\). Investigations into short channel MOSFETs have shown, that with effective gate lengths of less than 2 \(\mu\)m, the electric fields in the conducting channel can be large enough under normal operating conditions, to cause degradation of device behaviour\(^2\). Degradation occurs as a result of electrons gaining sufficient energy (~ 3.2 eV) to surmount the Si-SiO\(_2\) interface barrier (hot electrons) or, by means of punchthrough conduction whereby the depletion regions of the source and drain are able to make contact providing carriers with a low resistance path through the device\(^3,4,16\).

The gate width has always been known to have an influence on the electrical properties of MOS devices. However, the development of three dimensional transistor modelling programmes has shown that at widths less than 2 \(\mu\)m, the surface potential, electric field and electron density in the conduction channel are affected by the gate width\(^5,29\). This is because the fringing effect of the electric field at the edges of the gate oxide can no longer be considered negligible, as the gate width approaches the depletion layer thickness in the transistor.

The gate oxide thickness is the third parameter of importance in the transistor operation since it influences the electric field in the channel. Defect mechanisms associated with oxides ranging from 200 \(\AA\)
to 600 Å have been discussed extensively in Chapter 6 of this work. Experimental data on 400 Å oxides has been presented in Chapter 6. In this chapter, the models presented earlier will be extended to VSD and submicron devices. High electric fields across the thin gate oxides, can lead to quantization of the energy bands in the inversion layer at normal operating voltages. Hence the electrons can be considered to be confined to a two dimensional layer at the surface, affecting electronic conduction properties and the electrical behaviour of small dimensions devices.

The shapes of the source and drain junctions are also of importance in VSD devices. Diffusion of the source/drain implants under the gate can be quite considerable (~ 0.7 μm) and will influence the device behaviour.

A3.2.2. Doping concentrations

In order that very small dimension transistors have the same electrical performance as long channel devices, it is necessary to increase the doping concentration. The exact increase is dependent on the threshold voltages and conduction properties required and may be two orders of magnitude greater than that of a conventional device. In addition the depths of the source and drain junctions affect the punch-through voltage and in order to control this, the energy of the source and drain implants must be reduced, requiring higher dose rates for the dopant concentrations necessary. Hot electron emission can be controlled by graded doping of the drain and source junctions which again adds further complications to the fabrication techniques.
The effect of this heavy doping on interface states may be significant, influencing the reliability of the device. High dopant concentrations will also contribute to surface quantization effects\(^1,6,8,23\).

It is also necessary to keep heat treatments and annealing steps at low temperatures and as short as possible to minimize excessive diffusivity of the implanted ions\(^8\).
A3.3. COMPUTER SIMULATION OF SUBMICRON SILICON MOS DEVICES

A3.3.1. Semiconductor device equations

The equations used in the simulation of silicon MOS devices are the Poisson equation (Eq. A3.1), the current equations for electrons (Eq. A3.2) and for holes (Eq. A3.3) and the continuity equations for electrons (Eq. A3.4) and for holes (Eq. A3.5) in three dimensions:9,10,26.

\[ \nabla^2 V(x, y, z) = \frac{-\rho(x, y, z)}{\varepsilon_s} \tag{A3.1} \]

\[ J_n = q \mu_n n \nabla F + q D_n \nabla n \tag{A3.2} \]

\[ J_p = q \mu_p p \nabla F - q D_p \nabla p \tag{A3.3} \]

\[ \frac{1}{q} \nabla \cdot J_n + \frac{\partial n}{\partial t} = G - R \tag{A3.4} \]

\[ \frac{1}{q} \nabla \cdot J_p + \frac{\partial p}{\partial t} = G - R \tag{A3.5} \]

where

\[ \rho(x, y, z) = -q(n - p + N_D - N_A) \tag{A3.6} \]
is the total space charge in the semiconductor, \( n \) and \( p \) are the
electron and hole concentrations respectively, and \( N_D \) and \( N_A \) are the
donor and acceptor concentrations respectively. \( V(x,y,z) \) is the
 electrostatic potential.

\( J_n \) and \( J_p \) are the electron and hole current densities
respectively, while \( \mu_n \) and \( \mu_p \) are the mobilities. \( G \) describes the
generation rate incorporating impact ionization or carrier generation
by external radiation, while \( R \) describes recombination processes. \( F \)
is the electric field given by \(-\nabla V(x,y,z)\). \( D_n \) and \( D_p \) are the
diffusion coefficients for electrons and holes respectively.

Poisson's equation is derived from Maxwell's Laws and describes
the charge distribution within the semiconductor as a function of the
electrostatic potential at any given point \( V(x,y,z) \). The current
equations define the absolute values, directions and orientation of
the electron and hole currents, while the continuity equations
characterise the balance of sinks and sources for electron and hole
currents.

The equations are solved using complex numerical methods based on
either Finite Element or Finite Difference techniques for partial
differential equations, to obtain a simulation of the charge
concentrations, potentials and electric fields in a MOS device
depending on the boundary conditions\(^{13}\). Figure A3.1. shows the area
over which the equations are to be solved. The boundary conditions
are chosen according to the type of device being considered and may
vary depending on the model\(^9-12,26\). Basically, charge neutrality is
assumed at the source and drain contacts which are considered ohmic,
as well as the substrate contact, giving:

\[
n - p + N_D - N_A = 0 \quad \text{Eq (A3.7)}
\]
Figure A3.1 CROSS SECTION OF AN NMOS TRANSISTOR SHOWING THE AREA OVER WHICH POISSON'S EQUATION IS TO BE SOLVED BOUNDED BY THE THICK RECTANGLE
In thermal equilibrium

\[ np = n_i^2 \]  \hspace{1cm} \text{Eq (3.8)}

The electric displacement \( \mathbf{D} \) is considered to be continuous at the Si-SiO\(_2\) interface, giving

\[ \mathbf{D}_{\text{ox}} = \frac{\varepsilon_s}{\varepsilon_{\text{ox}}} \mathbf{E}_s \]  \hspace{1cm} \text{Eq (3.9)}

where \( \mathbf{E}_{\text{ox}} \) is the electric field at the interface and \( \varepsilon_{\text{ox}} \) is the oxide permittivity. Using

\[ \varepsilon = -\nabla \cdot \mathbf{D} \]  \hspace{1cm} \text{Eq (4.10)}

the electrostatic potential can be determined.

Various modifications have to be made to these boundary conditions when applying the solutions to submicron devices, in order to accommodate high-field phenomena and heavy doping effects as well as the edge effects at the gate width\(^5,12,26\).
5.3.2. **Simulation of very small dimension MOS devices**

Very small dimension MOS devices have been successfully simulated using three dimensional techniques to observe the effects of the gate length and width on device performance. Due to the restrictions on the export of computer software of this nature from the USA, the author has had to rely on published results to evaluate the influences of small geometries on device behaviour. The high complexity of the numerical analysis and the extent of computer programming involved made it impractical to develop a programme for this project, here at Loughborough. The published results have therefore been used in the rest of this chapter to form the basis of the analysis into degradation in submicron devices presented here.

Three dimensional simulation of an n-channel small geometry (1 µm x 1 µm) MOSFET device has provided results for the potential distribution as a function of gate length and gate width as shown in Figure A3.25. The substrate voltage $V_{\text{sub}} = -3.0$ V, gate voltage $V_{\text{GS}} = 0.9$ V and the drain-source voltage $V_{\text{DS}} = 2.0$ V. It can be seen that the channel width does affect the potential distribution. The effective channel width is not uniform along the channel length, being wider near the source and then tapering off to a minimum at the pinch-off point, widening again towards the drain. The width widening effect at the source is attributed to the substrate bias, $V_{\text{sub}}$, although in the channel it is due to the width influencing the surface potential. The electric field dependence on channel length and channel width (Figure A3.3.) confirms the width influence. A decrease in effective channel width causes a decrease in surface potential. At
the pinch off point the channel width is at a minimum and hence the total electric field near the drain is a minimum as indicated by the trough in Figure A3.3. However, the field in the vicinity of the drain is extremely high under the biasing conditions (V_{sub} = 2.0 V, V_{GS} = 1.6 V, V_{DS} = 8.0 V).

Results for avalanche breakdown have shown the ionization integral to decrease as the channel width got smaller. Hence the avalanche breakdown threshold voltage was greater for smaller devices. Punchthrough conduction is found to decrease at channel widths below 4.0 μm. This is attributed to a decrease in potential at smaller widths with an increase in the potential barrier at the source to bulk junction. Current flow is therefore reduced\textsuperscript{5,29}. 


FIGURE A3.2. The potential distribution in an MOS transistor in 3-D (ref. 5).

FIGURE A3.3. The electric field distribution in an MOS transistor in 3-D (ref. 5).
A3.4. THE EFFECTS OF VOLTAGE STRESS ON SUBMICRON MOSFETS

A3.4.1. Introduction

In this section it is proposed to use the understanding of submicron MOSFETs gained by computer simulation (as described in section A3.3.2.), in conjunction with the understanding of oxide breakdown mechanisms described in chapter 6, to predict the effects of voltage stress on submicron MOSFETs. Degradation mechanisms caused by stress in the lateral direction, i.e. \( V_{DS} \), will be briefly discussed with regard to published results, for completeness.

A3.4.2. Lateral field effects

Figure A3.4 shows a typical distribution of equipotentials in an n-channel MOSFET under the influence of a positive \( V_{DS} \), with \( V_{GS} > 0 \). Electric field lines and current flow will be in a direction normal to these equipotentials. Therefore, it can be seen that the highest electric fields are obtained close to the drain and along the semiconductor surface. As device geometries get smaller and electric fields get larger, three main effects influence degradation mechanisms in MOSFETs.

(i) Avalanche breakdown through carrier multiplication at the large drain-substrate reverse-biased junction\(^{15}\).

(ii) Punchthrough conduction and drain induced barrier lowering due to interaction of the large depletion regions at the drain and source\(^{16,17}\).
FIGURE 3.4. The potential distribution in a MOSFET under the influence of a drain-source voltage ($V_{DS}$).
(iii) Hot carrier effects where electrons or holes may gain sufficient energy from the drain-source electric field to surmount the potential barrier at the Si-SiO₂ interface.

Degradation of device performance occurs as a result of interface state generation which affects the properties of the conduction channel. Extensive work has been carried out into hot electron and hot-hole effects and methods of improvement in small geometry devices²,⁷,¹⁸-²⁰,²⁷,²⁸.

Avalanche breakdown and punch-through conduction in very small dimension devices has been briefly discussed in the previous section (83.3.2.) Hot electron effects are believed to be reduced, as the channel length decreases below about 5 µm¹⁸. This is attributed to be proximity of the drain and source diffusions causing electrons generated in the substrate, as a result of impact ionization at high fields, to be swept directly to the diffusion region rather than to the surface. However, a decrease in channel width would increase the hot carrier effects, because of the edge effects¹⁸. The higher electric fields generated at the gate edge would enhance the probability of emission from silicon into the SiO₂ increasing the hot-electron emission current. Improvements are centred on lightly doped drain structures or graded drain doping,⁷,²⁰ although the influence on degradation of traps and interface states present in the oxide and therefore the quality of the oxide, must be considered²⁸.
A3.4.3. Vertical field effects

Vertical field effects are a consequence of stress voltages applied to the gate of the MOSFET. Chapter 6 has covered the effects of gate voltage stress on oxide breakdown mechanism, and highlighted the following features.

(i) Breakdown was caused by the conduction of electrons through the SiO₂ either by hopping conduction (d.c. stress) or by impact ionization in the conduction band (pulsed voltage stress).

(ii) Electrons were injected into the SiO₂ by means of tunnel emission at the Si-SiO₂ interface, a mechanism which is dependent on the electron concentration at the injecting electrode as well as the interface traps.

(iv) Breakdown was enhanced by the presence of traps and donor sites in the oxide bulk.

Section A3.4.2. has also shown that under the influence of high electric fields in the channel, a further injection mechanism at the Si-SiO₂ interface, in the form of hot carrier emission is introduced. These injected carriers will also contribute to electronic conduction process in the SiO₂ and therefore to the mechanisms of oxide breakdown.

Figure A3.5. shows a cross sectional view through a MOSFET indicating areas most sensitive to breakdown is small dimension devices. The underdiffusion of the drain and source wells will take up a greater proportion of the gate length with respect to the channel.
FIGURE A3.5. MOSFET showing areas of high breakdown probability in submicron devices (cross-section through drain and source).
length compared to a long channel device\textsuperscript{24,25}. The surface charge concentration dependence of oxide breakdown will therefore mean that n-channel MOSFETs will be more sensitive to gate voltage stress, especially pulsed (ESD) voltages, at submicron levels. Hot electron emission is highest at the drain junction and under dynamic operating conditions, these carriers can contribute to the surface charge concentration effect to result in lower breakdown thresholds (d.c. gate voltages) for n-channel submicron MOSFETs. In p-channel MOSFETs the high channel dopant concentrations required to maintain performance characteristics will mean more electrons at the semiconductor surface. These devices will therefore be more sensitive in the channel region to positive voltage pulses, particularly of an ESD nature, than n-channel MOSFETs.

Figure A3.6. shows a width cross-section through a MOSFET device, indicating the edge effects of the electric field. Higher electric fields at the edges enhance hot carrier emission under dynamic operating conditions, as well as contributing to the oxide breakdown mechanisms.

However, oxide thicknesses below 150 Å are expected to show an increase in pulsed breakdown thresholds because of their lower sensitivity to impact ionization. The thinner oxide providing less time for the ionization process to reach breakdown proportions\textsuperscript{21}.

\textbf{A3.4.4. Conclusions}

Oxide breakdown mechanisms proposed in previous chapters have been applied to submicron MOSFET device structures to identify areas which may be sensitive to voltage stresses. The drain-source voltage
in small MOSFETs causes hot carrier emission which can contribute to these breakdown processes making devices even more sensitive under bias conditions. n-channel and p-channel MOSFETs are sensitive at different regions, i.e. the drain and source underdiffusion and the channel region respectively. Submicron CMOS devices can therefore be considered to be more sensitive to breakdown because of these effects than standard geometry devices. Narrow gate widths (< 2 μm) universally enhance the oxide breakdown process, both under gate voltage stress (high electric fields due to edge effects) and dynamic bias conditions (increased hot carrier emission due to edge effects).

Experimental observations of pulsed voltage effects on small dimension devices indicate a trend supporting the increasing probability of gate to drain/source breakdown with smaller channel lengths.
FIGURE A3.5. MOSFET showing areas of high breakdown probability in submicron devices (cross-section through channel indicating electric field effects due to width).
A3.5. Summary

1. Important features of very small dimension MOSFETs are:
   a) the high electric fields in the conduction channel region
   b) the increasing influence of edge effects as width dimensions approach the depletion layer dimensions.
   c) higher dopant concentrations required to obtain device performance comparable to standard devices.
   d) drain and source underdiffusion and dopant concentrations and their influence on oxide breakdown mechanisms and interface states.

2. Computer simulation of MOSFETs in three dimensions is essential with devices approaching submicron sizes, since width effects can no longer be ignored. Programmes of this nature already exist, and published results, showing width and length effects on the surface potential and electric field of VSD devices, have been presented in this chapter.

3. Published results from the computer simulation of VSD devices show that as the width of the device decreases, the avalanche breakdown threshold voltage at the drain increase, and the punch through current decreases for a given voltage.

4. Hot electron effects are found to increase with smaller channel widths, although at channel lengths less than 5 μm, the emission current decreases. Degradation in VSD MOSFETs is the result of interface state generation in the presence of hot carrier injection at the Si-SiO₂ interface.

5. The smaller geometry of VSD devices compared to standard MOSFETs make them more sensitive to some of the oxide breakdown
mechanisms described in Chapter 3 and 4. Particularly sensitive is the gate oxide at the drain and source underdiffusion regions in n-channel MOSFETs and at the channel regions in p-channel MOSFETs because of the higher surface charge concentrations and the corresponding effect on oxide breakdown. Higher electric fields at the gate edges (along the channel width) and the prospect of higher hot carrier emission can make the gate oxide in this region also very sensitive to breakdown.

6. Experiments on small dimension devices described in Chapter 4 have indicated a trend towards increased gate to drain/source breakdown as the channel length decreases. This is considered to be justification for the validity of the proposed breakdown sensitivity of silicon submicron MOSFETs presented here.
A3.6. REFERENCES


2. HOFMANN K.R., WERNER C., WEBER W., DORDA G., Hot-electron and Hole-emission Effects in Short n-Channel MOSFET's.


4. BLICHER A., Field-Effect and Bipolar Power Transistor Physics;

5. CHAMBERLAIN S.G., HUSAIN A., Three-Dimensional Simulation of VLSI MOSFET's: The Three-Dimensional Simulation Program WATMOS.


7. TAKEDA E., KUME H., TOYABE T., ASAI S., Submicrometer MOSFET structure for minimising hot carrier generation.

8. FICHTNER W., LEVIN R.M, TAYLOR G.W., Experimental Results on Submicron-Size p-channel MOSFET's.

10. BUTURLA E.M., COTTRELL P.E., GROSSMAN B.M., SALSBURG K.A.,
    Finite-Element Analysis of Semiconductor Devices:
The FIELDAY Program.
11. MAYERGOYZ I.D., Solution of the nonlinear Poisson equation of
    semiconductor device theory.
12. FICHTNER W., ROSE D.J., BANK R.E., Semiconductor Device
    Simulation.
13. BANK R.E, ROSE D.J., FICHTNER W., Numerical Methods for
    Semiconductor Device Simulation.
14. SCHMITZ A.E., CHEN J.Y., Design, modelling and fabrication of
    subhalf-micrometer CMOS transistors.
15. TOYABE T., YAMAGUCHI K., ASAI S., MOCK M.S., A Numerical Model of
    Avalanche Breakdown in MOSFET's.
16. BARNES J.E., SHIMOHIGASHI K., DUTTON R.W., Short-Channel MOSFET's
    in the Punchthrough Current Mode.
17. TROUTMAN R.R., VLSI Limitations from Drain-Induced Barrier
    Lowering.
18. COTTRELL P.E., TROUTMAN R.R., NING T.H., Hot-electron Emission
    in N-channel IGFET's.
19. TSUCHIYA T., FREY J., Relationship Between Hot-Electron/Holes and Degradation of p- and n-channel MOSFET's.


23. BARKER J.E., FERRY D.K., On the physics and modelling of small semiconductor devices-II.

24. ONG T-C, TAM S., KO P-K., HU C., Width Dependence of Substrate and Gate Currents in MOSFET's.

25. WANG P.P., Device Characteristics of Short-Channel and Narrow-width MOSFET's.

26. PFEISTER J.R., SHOTT J.D., MEINDEL J.D., Performance Limits of CMOS ULSI.

27. TAKEDA E., JONES G.A.C., AHMED H., Constraints on the Application of 0.5 μm MOSFET's to ULSI systems.
28. MATSUMOTO H., SAWADA K., ASAI S., HIRAYAM M., NAGASAWA K.,
Effect of Long-Term stress on IGFET Degradations due to
Hot Electron Trapping.

29. LAI P.T., CHENG Y.C., An analytical model for the narrow-width
effect in ion-implanted MOSFETs.

30. PISCES II - developed by M.R. Pinto, C.S. Rafferty, R.W. Dutton;
Center for Integrated Systems, Stanford University,
California.
APPENDIX 4

ABSTRACTS OF REPORTS ASSOCIATED WITH THIS WORK.

Report No: 1 MOD Contract No: A5a/1265
LUT Contract No: RXB 211W
(August 1983 - March 1984)

Report No: 2 MOD Contract No: A5a/1265
LUT Contract No: ELB 211W
(March 1984 - September 1984)

Report No: 3 MOD Contract No: A5a/1412
LUT Contract No: ELB 247L
(September 1984 - April 1985)
ABSTRACT OF REPORT No: 1

An investigation has been conducted into the failure mechanisms of semiconductor devices based on a study of the available literature on the subject.

This report provides an extensive coverage of the existing theoretical and experimental aspects with respect to each area of semiconductor reliability. The survey has included analysis of reliability engineering and failure physics pertinent to a thorough understanding of semiconductor failure mechanisms. Hence, manufacturing procedures, screening processes, quality assurance and reliability modelling have been dealt with extensively.

The principal object has been to identify the possible sources of failure mechanisms, and to investigate methods for detection and elimination which could be used to improve the reliability of semiconductor devices.

Over 300 references have been studied and provide much of the information contained in this report. Discussions with semiconductor device manufacturers, electronic equipment manufacturers and users have helped to evaluate the literature in a practical context.
The Electrostatic Discharge (ESD) sensitivity of small dimension NMOS Field Effect Transistors has been investigated. Discrete NMOS FETs of varying dimensions and a constant gate oxide thickness of 400 A were each subjected to a single ESD voltage pulse of between 50 volts and 250 volts, at temperatures between 25°C and 200°C. Over 4000 devices were used, all resident on a single 3 inch silicon wafer. The object of the experiments conducted, was to determine the dependence of device ESD sensitivity on temperature, applied voltage and the gate dimensions. The voltage pulse was generated using the "human-body model" circuit as described in Method 3015.2 of the MIL-STD 883 C.

Device I-V characteristics obtained after application of the pulse were classified into five categories.

1) No change
2) Reduced characteristic - a degradation in device performance
3) A reduced characteristic with anomalous high gate-field effects.
4) Complete failure with no current being obtained.
5) Linear I-V characteristics

At 50 volts, the % number of devices showing "no change" was 72%, with only 5% showing complete breakdown. At 250 volts, the % number showing "no change" had dropped to 7%, with about 10% showing complete breakdown. Degradation accounted for over 80% of the devices.

No temperature dependence of device ESD sensitivity was observed within the range of the experiment.
A cumulative ESD effect was observed, whereby the degradation of device performance was found to increase with the number of applied pulses. All devices showed complete breakdown with the application of 25 pulses at 1 kV.

Complete breakdown was found to be gate dimension dependent increasing with increased channel length at a constant width.

Analysis of the breakdown characteristics enabled a model to be developed describing the mechanisms of ESD damage.
ABSTRACT OF REPORT No: 3

This report describes the work done on the contract "Failure Mechanisms in Semiconductor Devices" between November 1984 and April 1985.

Further analysis of the ESD results reported in Report No. 2 have been made and this has been followed by an investigation of the application of continuous ramped voltages to MOS capacitors of the same oxide thickness as that of the gate oxides of the MOSFETs. The following conclusions have been reached:

(a) The lengths of the polysilicon gate contacts do not have an attenuating effect on the amplitude of the applied ESD voltage pulse. Both experimental and analytical investigations show that all devices with a common gate contact would see the same gate voltage.

(b) The geographical distribution of the damage categories 1 to 3 do not show a distinct preference for any particular location on the chip. Category 4 devices are found to show a preference for devices with small gate width (1.4 μm < w < 2.1 μm). While category 5 devices are clustered around the gate contact pads.

(c) Experiments on MOS Capacitors showed a significant differences between the breakdown voltages of devices upon application of an ESD pulse and with a continuous ramped voltage. Damage with an applied ESD pulse occurs at approximately 200 volts compared with approximately 40 volts when the stress is continuous.
Work has been extended into failure mechanisms in submicron MOSFETs and it has been decided that the effect of the fringing field becomes of considerable importance in small dimension devices.
Failure Mechanisms in Semiconductor Devices by E.A. Amerasekera and D.S. Campbell
(to be published by John Wiley and Sons (U.K.) 1986.)

ABSTRACT

This monograph provides an extensive coverage of the existing theoretical and experimental aspects with respect to semiconductor reliability. It includes analysis of reliability engineering and failure physics pertinent to a thorough understanding of semiconductor failure mechanisms. Hence, as well as reliability as such, manufacturing procedures, screening processes, quality assurance and reliability modelling have been considered in detail.

The principal object has been to identify the possible sources of failure mechanisms, and to investigate methods for detection and elimination which could be used to improve the reliability of semiconductor devices.

Over 300 references are quoted and provide much of the information contained in this monograph. Discussions with semiconductor device manufacturers, electronic equipment manufacturers and users have also helped to formulate this survey and strengthen the practical context.
APPENDIX 6

Electrostatic Pulse Breakdown in NMOS Devices
ELECTROSTATIC PULSE BREAKDOWN IN NMOS DEVICES

1 A AMERASLIRA AND D S CAMPBELL

Electronics Component Technology Group Electronics and Electrical Engineering Department, Loughborough University of Technology, Loughborough, Leicestershire LE11 3TU, U.K.

SUMMARY

The electrostatic discharge (ESD) sensitivity of small dimension n-channel metal oxide semiconductor (NMOS) field effect transistors (FETs) has been investigated. NMOS FETs of varying dimensions and a constant gate oxide thickness of 400 Å were each subjected to a single ESD voltage pulse of between 50 and 250 V at temperatures between 25 and 200°C. Over 4000 devices were used, all residing on a single 3 inch silicon wafer. The object of the experiment was to determine the dependence of ESD sensitivity on temperature, voltage, and device dimensions as well as to investigate the mechanisms that cause oxide breakdown as a result of ESD damage.

The temperature dependence of device ESD sensitivity was observed within the range of the experiment. A significant voltage dependence was observed with degradation accounting for over 80 percent of cumulative ESD effects observed, whereby the degradation of device performance was found to increase with the number of applied pulses. Analysis of the breakdown characteristics revealed the cause of damage was oxide breakdown. Application of the ESD pulse appears to lead to oxide breakdown through impact ionization within the oxide. The very short duration of the pulse not being favourable to processes involving electron trapping unless these traps are already present in the oxide.

KEY WORDS: Semiconductors, MOS technology, Electrostatic discharge damage (ESD)

1 INTRODUCTION

The problems associated with the presence of static potentials in working environments have been well documented. Static potentials of 20 kV are easily generated. The dielectric strength of SiO₂ is approximately between 9 and 14 MV/cm (i.e. between 90 and 140 kV/cm for a 1000 Å oxide thickness). Hence MOS devices are very sensitive to electrostatic discharge (ESD) damage.

With the advent of devices of very small dimensions, the problems associated with ESD are becoming more critical. Oxide thickness down to 150 Å and MOSFET dimensions approaching 1 μm are even more sensitive to ESD than earlier devices.

Investigations into ESD sensitivity and failures are wide-ranging, covering a vast multitude of devices and operating conditions. There is, however, a grey area regarding the mechanisms behind the breakdown of ESD damaged devices.

The objectives of the experiments presented here were twofold:

(a) to examine the ESD sensitivity of NMOS devices as a function of temperature

(b) to investigate the mechanisms of ESD damage and its subsequent effect on device behaviour.

Item (a) is based on a paper by Hart et al. which presents a case for the temperature dependence of ESD sensitivity. However, the devices investigated by Hart et al. were packaged MOSLSI devices containing input protection circuits. Failure was traced to damage in the protection circuit, as a combined result of the ESD pulse and the applied temperature.

A need was therefore seen to investigate NMOS FETs without protection circuits at the gate contact. For this purpose wafers consisting of discrete small dimension NMOSFETs were obtained by courtesy of Plusey Research (Caswell) Ltd. These wafers consisted of transistors of a variety of dimensions, making them ideal for the experiments at hand.

The paper consists of four sections detailing the work. In Section 2 a brief outline is given of existing dielectric breakdown theory. Sections 3 and 4 present the experimental procedure and the results respectively, and Section 5 provides the analysis and discussion of the data obtained. Over 4000 devices were used in the experiment, providing a very large data bank for analysis.

2 DIELECTRIC BREAKDOWN

2.1 Introduction

There are two mechanisms which are dominant in the electrical breakdown of dielectric materials. They are:

(i) impact ionization

(ii) electron transport through the oxide by means of electron traps in the oxide.

The two mechanisms are briefly outlined below.
2.2 Ionization

In the presence of high electric fields, in excess of 1MV/cm, the conduction band edge is distorted. Hence electrons can gain access to the conduction band from the valence band. Electrons are injected by Schottky barrier emission (potential barrier lowering), Poole-Frenkel emission (field-enhanced thermal excitation of electrons into the conduction band) or Fowler-Nordheim tunnelling, from both the metallic contacts and donor impurity sites within the oxide.

Once in the conduction band electrons may be accelerated by the applied electric field to an energy large enough to ionize the lattice and produce a destructive electronic cascade. Energy transfer between electrons and the lattice takes place as a result of electron-phonon interactions within the oxide. Consequently localized heating occurs which could lead to thermal burnout.

2.3 Electron trapping

Distortion of the band edges in the presence of high electric fields could also enable electron transport through the oxide by hopping between donor impurity sites and trap impurity sites. Hence a kind of "conduction band is formed within the forbidden energy gap". Harari has demonstrated a strong relationship between electron trap generation under high stress conditions and oxide breakdown. Trapped electrons could also result in localized electric fields in excess of the dielectric field strength.

Experimental evidence presented by Hughes and Harari has shown that defects can be generated in oxides which were originally defect free, under time-dependent high stress conditions. Such a mechanism would be significantly temperature dependent.

2.4 Temperature and voltage dependence of oxide breakdown

It is important to bear in mind that the electrostatic discharge sensitivity of MOS devices would not involve a time-dependent factor: the duration of the pulse being very small (0.1-15ms).

Osburn and Ormond have shown a very small temperature dependence of the breakdown strength of 400Å SiO2 between -30°C and 300°C. They suggested that impact ionization rather than the thermally derived electron trap mechanism was the dominant mechanism.

The electron trap mechanism as presented by Harari is also less suited to the very narrow pulse widths associated with ESD. The long time-dependence required to generate sufficient traps to cause oxide breakdown would not be relevant to ESD. Harari suggested a temperature dependence based on this mechanism which would therefore not be applicable to ESD breakdown.

Osburn and Wentzmann found an increase in leakage current through the dielectric as the applied field is increased, in accordance with the impact ionization model. They also found that the leakage current and maximum dielectric field strength were not significantly influenced by the substrate type or dopant concentration.

3 EXPERIMENT

3.1 The silicon wafer

As noted earlier the tests and results to be described were obtained on process assessment wafers kindly provided by Plessey Research (Caswell) Ltd. The NMOS transistors are manufactured on 3 inch slice p-type wafers. On these wafers n⁺ wells are diffused to form the source and drain using arsenic ions, with a concentration of 6x10¹⁰ cm⁻² at 100 keV. A gate oxide of 400Å thickness is thermally grown at 950°C. The polysilicon n⁺ gate is grown at 620°C and is 4500Å thick. In order to obtain positive threshold voltages for the enhancement type device, the channel is preferentially doped with boron ions with a concentration of 4x10¹⁰ cm⁻² at 30 keV. The substrate resistivity is 20Ω cm.

The water consists of about 100 chips. Each chip consists of three blocks of NMOS transistors, two of which consist of enhancement mode devices and the third of depletion mode devices. The transistors are set in rows, each with a common source and a common gate contact. In some cases individual MOS transistors are constructed in parallel and connected as a single transistor in order to enable better current measurements to be made. Results were examined to ensure that this parallelism did not affect the results in any way. Eight drain contacts with effectively eight transistors available in each row. The transistors are of varying channel lengths and gate widths ranging from 1um and 102nm.

There are approximately 176 devices available for experimentation per chip and a sum total of approximately 17,600 devices on a wafer.

3.2 The apparatus

The ESD pulse is generated using the "human body" model circuit as described in Method 3015.2 of the MIL-STD-883C and shown in Figure 1(a). A 10nF capacitor is charged up to the required pulse voltage and then discharged through a 1kΩ resistor directly onto the gate contact of the device under test. Both the device contact and the source contact were left floating. The substrate was grounded.

A heated chuck with thermostat control enables the temperature of the water to be set at the desired level (±2°C) when investigating the temperature dependence of ESD sensitivity. Figure 1(b) shows a block diagram of the total experimental arrangement.
3.3 Experimental procedure

3.3.1 Measurements The current-voltage (I-V) characteristics obtained using a curve tracer (CT) contain information from which most of the electrical properties of the transistor can be determined. Therefore, the CT was used to monitor changes in transistor characteristics after application of the ESD pulse.

All measurements were made at room temperature. When a wafer was heated and pulsed, its electrical properties were examined at room temperature (298K).

3.3.2 Voltage range The ideal voltage range was considered to be one in which a steady rise in catastrophic failure would be observed with increasing temperature, but 100 per cent failure of all devices in a group was not desired. The voltage range found to be most suitable was between 50 and 300 V. Therefore, the devices were subjected to pulses of 50, 100, 150, 200, and 250 V.

3.3.3 Temperature range The temperature range was required to encompass the standard maximum operating temperatures for commercial devices (70°C) and military devices (125°C). In addition, measurements were also made at room temperature, which were used as a control group. Extreme temperature dependence was examined with a group at 200°C. The five selected temperatures were therefore 25, 70, 110, 150, and 200°C.

3.3.4 Number of devices The total number of devices used in this experiment was approximately 5 x 176 = 800 at a given temperature 5 x 800 = 4000 in total. Thus 25 chips were used.

3.3.5 Precautions It was important that maximum precautionary measures were taken to ensure that devices which were damaged accidentally either by the measuring technique or by spurious ESD pulses were not included in the analysis. Thus, any devices which show $f_{on}$ vs $f_{off}$ characteristics in which $f_{on}$ differs by more than ±10 percent from the
standard provided by unpulsed devices, or which do
not work in the pre-experiment measurements, are
noted and left out of any further experimentation.
Before any experiments were begun, and before the
wafer was handled, the operator discharged himself
of any static potential that may have accumulated.

3.3.6 Method The experimental method con-
Sisted of the following steps:
1. The characteristics of all the devices on 25 chips
   were recorded.
2. The wafer was brought to room temperature,
   25°C being selected for uniformity.
3. All devices on one chip were subjected to a
   single ESD pulse of the required voltage at a
   given temperature.
4. The experiment was repeated for the selected
   voltage range on five different chips.
5. Steps 3 and 4 were repeated until the entire
   temperature range had been covered.
6. The wafer was then allowed to return to room
   temperature.
7. The device characteristics of all 25 chips were
   again recorded, emphasis being laid on those
   devices which showed changes in the character-
   istics recorded in step 1.

3.3.7 Definition of failure It was decided that
no particular specification would be laid down outside
of which a device was deemed to have failed. Instead,
any recorded variation in device characteristics from
measurements made before application of the ESD
pulse was categorized.

3.3.8 Cumulative ESD effect The above exper-
iment subjected the devices to only one ESD pulse. In
order to observe the effect of a series of consecutive
ESD pulses on the device characteristics a small
number of devices were subjected to a sequence of
pulses.

Only 24 devices were used for each set of pulses, 8
being pulsed at 50V, 8 at 150V and 8 at 250V. The
experiment was repeated at 10 pulses, 15 pulses, 20
pulses, 25 pulses, 50 pulses and 100 pulses, and the
characteristics were recorded.

Subsequent devices were subjected to a constant
number of pulses and various voltages between 350
and 1000V. Twenty-five pulses at 350V, 525V, 700V
and 1000V were applied, the characteristics being
recorded after application of the 25 pulses.

4 RESULTS

4.1 ESD pulse

4.1.1 Waveform The ESD pulse generated us-
ing the human body model is of the shape shown in
Figure 2, as observed on an oscilloscope and using a
low capacitance probe. The rise time of the pulse \( t_{\text{rise}} \)

\[ t_{\text{rise}} = 500\text{ns} \]  The decay time of the pulse, \( t_{\text{fall}} \) = 1ms
(this is dependent on the input impedance of the load
circuit and hence will vary for different gate imped-
ces).

4.1.2 Polarity The majority of experiments
were performed with positive pulses. However a few
experiments were performed using negative pulses
over the same voltage range. As the effects observed
were similar to those for positive pulses, no specific
analysis of these observations needs to be considered.

4.2 Characteristics

A comparison of the \( I_{DS} \) vs \( V_{DS} \) characteristics
observed before and after application of the ESD
pulse, shows that over 95 per cent of the devices could
be classified into one of the following four categories.

4.2.1 Category 1—no change All parameters
remain unaltered, the typical transistor character-
istics, as shown in Figure 3, being obtained.

4.2.2 Category 2—degradation A typical I-V
characteristic for devices in this category is shown in
Figure 4. The device has the same dimensions as the
device associated with Figure 3. However, a compar-
ison of the two characteristics shows significant de-
gradation to give Figure 4. The mutual transconduc-

dt
inance \( (g_m) \) has decreased along with the drain-source conductance \( (g_{ds}) \) in the linear region. \( g_m \) in the saturation region has increased.

Measurements of the gate-drain resistance, \( R_{GD} \), and the gate-source resistance, \( R_{GS} \), given in Table I, show that only a very small leakage current is present.

Table I: Gate-source resistance \( (R_{GS}) \), gate-drain resistance \( (R_{GD}) \) and gate-substrate resistance \( (R_{GS}) \) together with associated current values for devices exhibiting the five categories of characteristics (N.B. No two of the devices measured in the table below have the same dimensions.)

<table>
<thead>
<tr>
<th>Category</th>
<th>( R_{GS} )</th>
<th>( R_{GD} )</th>
<th>( V_{GS} )</th>
<th>( I_{DS} )</th>
<th>( I_{GD} )</th>
<th>( I_{CS} )</th>
<th>( R_{DS} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( \times )</td>
<td>( \times )</td>
<td>( \times )</td>
<td>( \times )</td>
<td>( \times )</td>
<td>( \times )</td>
<td>( \times )</td>
</tr>
<tr>
<td>2</td>
<td>25</td>
<td>75</td>
<td>40</td>
<td>0.46</td>
<td>0.34</td>
<td>1.0</td>
<td>1000</td>
</tr>
<tr>
<td>3</td>
<td>25</td>
<td>75</td>
<td>40</td>
<td>2.5</td>
<td>0.94</td>
<td>0.04</td>
<td>2.3</td>
</tr>
<tr>
<td>4</td>
<td>109</td>
<td>1840</td>
<td>40</td>
<td>0.94</td>
<td>0.952</td>
<td>1400</td>
<td>0.0014</td>
</tr>
<tr>
<td>5</td>
<td>165</td>
<td>0</td>
<td>40</td>
<td>10</td>
<td>12.5</td>
<td>1470</td>
<td>0</td>
</tr>
</tbody>
</table>

4.2.3 **Category 3**—degradation with negative \( g_m \) at high \( V_{GS} \). In addition to the degradation described above, these devices also exhibit a drain-source current which begins to decrease as \( V_{GS} \) increases beyond a specific value (see Figure 5). Measurements show that oxide resistance values (Table I) are lower than those associated with category 2 devices.

4.2.4 **Category 4**—catastrophic failure. This category is so termed because \( I_{DS} = 0 \) for all values of \( V_{GS} \), and for all values of \( V_{GS} \) up to avalanche breakdown as shown in Figure 6. Typical \( R_{GD} \) and \( R_{OD} \) values in Table I indicate that the oxide at the source shows very high leakage.

4.2.5 **Category 5**—resistance characteristics. These characteristics are shown in Figure 7. The resistance values in Table I indicate an oxide breakdown at the drain which, owing to the electrical circuitry of the CT, results in these characteristics.

4.3 **The cumulative ESD effect**

The results of this experiment are presented in Figures 4 and 5.
However, published opinion is that the ESD sensitivity of MOS devices is influenced by temperature. Work presented by Hart et al. shows that this is indeed true. But the devices investigated by Hart were packaged integrated circuits incorporating input protection circuitry. The cause of failure was in fact traced to ESD damage of a diode in the protection circuitry. The implication, therefore, is that it is possible for the protection circuitry to create a temperature dependence which does not normally exist when dealing with only the NMOS FET.

5.1.2 The voltage dependence of ESD sensitivity.

From Figure 9, it will be observed that the number of devices in which no change in device characteristics is observed after ESD pulsing decreases as the applied voltage increases. A significant voltage dependence therefore exists for damage characteristics.

5.1.3 The dimensionless dependence of ESD sensitivity.

A plot was made relating the dimensions of the devices to the number of devices failing in the most catastrophic form. In this category (category 4), \( I_{DS} \equiv 0 \) and hence the device can be regarded as completely destroyed. From Figure 10 a definite trend can be observed between increasing failures and increasing device dimension for a constant gate width. This suggests that the area of the oxide affects the probability of oxide breakdown. Oxide breakdown must therefore be related directly to the presence of defects and detect mechanisms in the oxide.

The above conclusion is in accordance with the oxide breakdown model presented by Harari where electron traps due to oxide defects are claimed to be the cause of breakdown. However, this still does not contradict the impact ionization model because only the existing electron traps would affect breakdown. Electrons thus trapped would generate high localized electric fields, resulting in an increase in impact ionization in the area and eventual oxide breakdown.

The number of defects in the form of electron traps in thermally grown oxide films is therefore considered to be relevant to the ESD sensitivity of MOS devices.

5.2 Analysis of device characteristics after pulsing

5.2.1 Category 1—no change. Theoretical breakdown voltages for an oxide of thickness 400Å vary between 50 and 100 V. Therefore at 50 V one would expect the majority of devices not to be affected, and this is indeed so as seen in Figure 9. The percentage of devices unaffected by the ESD pulse decreases from 72 at 50 V to 7 at 250 V, emphasizing the voltage dependence of the damage.

5.2.2 Category 2—degradation. The degradation in drain-source current indicates that the surface of the semiconductor is not being inverted by the gate-source voltage, to the original extent. Conduction of charge carriers through the damaged oxide prevents satisfactory inversion from taking place.
From Figure 4 it can be seen that the characteristics are of a perfectly standard nature for an NMOS transistor. Without the benefit of the control characteristics of Figure 3, one would be justified in assuming that the transistor is in working condition, albeit to a different specification. Even after an ESD test as specified by MIL-STD-883C the device may still pass as fit for operation if used in a digital circuit. Hence, serious implications are placed on the long-term reliability of devices subjected to such damage.

5.2.3 Category 3—Degradation with negative gm
at high Vds. Devices in this category are more
obviously damaged than those in the previous category. Figure 11 shows a plot of \( I_{DS} \) against \( V_{GS} \). For a device showing these characteristics, Figure 12 shows \( g_m \) as a function of the gate-source voltage. A negative \( g_m \) is obtained at \( V_{GS} > 9 \) V.

Once again the oxide has been damaged, becoming a weak conductor, as indicated by the degradation. However, it appears that as the gate-source voltage is increased, it reaches a stage at which the \( E \) field generated is sufficient to conduct charge carriers away from the inversion layer via the gate oxide. The conduction mechanism could be either Fowler–Nordheim tunnelling or Poole–Frenkel emission due to distortion of the SiO\(_2\) conduction band edge brought on by the application of the ESD pulse.

5.2.4 Category 4—catastrophic failure. A high leakage through the oxide would prevent a depletion region from forming at the semiconductor surface with obvious repercussions on the magnitude of the drain-source current. Although the drain-source current \( I_{DS} \) does not always reduce to zero, the breakdown is large enough to reduce \( I_{DS} \) by factors of at least 10\(^7\) from the order of mA to that of \( \mu \)A.

5.2.5 Category 5—resistive characteristics. The electrical circuitry of the CT is such that the current measurement is made between the drain contact and the biasing supply. Hence, a leakage path directly between the gate and the drain diffusion could result in the resistance type traces obtained. Measurements of \( R_{GS} \) and \( R_{GD} \) confirm that the conduction path is between the gate and the drain with \( R_{GD} \ll R_{GS} \) (see Table 1).

5.2.6 Comparison of the statistical distribution of devices in each category. From Figure 9 it is seen that there is a steady increase of devices showing degrada-

![Figure 11: \( I_{DS} \) vs \( V_{GS} \) for device showing degradation with negative \( \mu \) at high \( V \), (category 3) \( L = 2.5 \) \( \mu \)m \( W = 50 \) \( \mu \)m \( V_{GS} = -60 \).](image1)

![Figure 12: \( I_{DS} \) vs \( V_{GS} \) for device showing degradation with negative \( \mu \) at high \( V \), (category 3) \( L = 2.5 \) \( \mu \)m \( W = 50 \) \( \mu \)m \( V_{GS} = -60 \).](image2)
tion characteristics, whereas the catastrophic failure categories only account for about 20 per cent of the devices even at 250 V. This can be attributed to the oxide being damaged as the breakdown voltage is reached (between 50 and 100 V). However complete breakdown still does not affect the majority of devices even at voltages over twice the maximum breakdown voltage expected for an oxide of the given thickness. This indicates that a longer pulse or more pulses are required to create the required avalanche current in the oxide to cause breakdown on every occasion.

5.3 Cumulative ESD effect

Complete failure was not observed as soon as one might expect. Most devices still showed signs of life after 100 pulses in the 50 to 250 V range. Eventually complete failure was obtained with 25 pulses at 1000 V. These results are in accordance with the impact ionization theory of dielectric breakdown.

The ESD pulse was deliberately not of sufficient magnitude to cause complete breakdown, and the duration of the pulse was not long enough to enable enough charge conduction to take place to cause a complete short circuit through the oxide. A certain amount of leakage through the oxide takes place during the pulse duration. This must be dependent on the presence of electron traps which, when filled, provide favourable regions of high electric field causing localized impact ionization and conduction paths.

The interdependence of the electron trap density and high electric fields for damage to occur is an important observation. A device with a very good oxide (a low number of defects) would therefore be expected to be less susceptible to the very short duration high voltage ESD pulse. Pulses in excess of 1 kV must obviously generate a sufficient electron cascade in the oxide during the short duration to cause complete failure.

The cumulative ESD sensitivity of MOS devices is also of importance when considering the development of reliability tests.

5.4 ESD and oxide breakdown

As a result of the study discussed above it is proposed that the oxide damage caused by the application of ESD pulses is primarily due to impact ionization within the SiO₂. Electron traps do influence breakdown but cannot be generated to any significant degree during the period of a typical ESD pulse.

Further experiments in this area are intended to establish a firm model of oxide breakdown under applied stress, both ESD and continuous voltages will be investigated.

6 CONCLUSIONS

The following conclusions have been reached:

1. The ESD sensitivity of NMOS FETs is not temperature dependent between 25 and 200°C at voltages between 50 and 1000 V.
2. A strong voltage dependence is observed; the devices showing increased susceptibility to damage as the voltage is raised from 50 to 1000 V.
3. Five categories of device characteristics are observed after the application of a single ESD pulse to the gate of a device. These five categories have been analysed and hence a qualitative model of the breakdown proposed. The categories are:
   (a) No change—characteristics are unaltered.
   (b) Degradation—typical $I_D$-$V_D$ characteristics but degraded indicating that damage has taken place. Such devices would be a major reliability hazard.
   (c) Degradation with negative $g_m$ at large $V_{DS}$ indicative that the conduction band edge of the SiO₂ has been affected by the ESD pulse.
   (d) Catastrophic failure—complete breakdown of the type envisaged when looking for failures.
   (e) Gate to drain short-circuiting. There is some evidence that gate-drain breakdown is more prevalent in devices of small dimensions due to the effect of gate-drain overlap for small dimension devices.

These five categories have been analysed and hence a qualitative model of the breakdown proposed.

4. Breakdown is due to damage of the gate oxide.

5. Oxide damage due to ESD pulsing is caused by the electronic cascade as a result of impact ionization within the oxide. Existing electron traps in the form of oxide defects (broken Si–O bonds etc.) make the oxide more sensitive to breakdown or damage, owing to the greater impact ionization energy available at localized areas around the trapped electrons.

6. A cumulative breakdown effect is observed as the number of ESD pulses is increased. Damage results in device performance being continuously degraded until eventually complete breakdown occurs.

7. As a result of the oxide breakdown mechanism, it was expected that the number of defects (i.e. trap levels) in the oxide would influence breakdown. This was confirmed with an increasing failure rate being observed when the channel length of the device is increased while the width is kept constant.

Summary

(a) Five categories of damage were observed.
   Category 1: No change.
   Category 2: Reduced characteristics. Lower $g_m$. Damage to gate oxide due to impact ionization in oxide.
Category 3: Reduced characteristic Negative gm at large Vgs, Increased oxide damage

Category 4: Catastrophic breakdown Badly damaged oxide prevents formation of surface inversion layer

Category 5: Linear I/V characteristics Conducting path between gate and drain

(ii) The effects of the various parameters were as follows:

Dimensions:
- Increase area — Increase catastrophic breakdown
- Increase channel length — Increase breakdown
- Increase source through drain
- Decrease gate width — Increase breakdown
- Increase chance of field distortion breakdown

Temperature:
- No effect on behaviour up to 200°C

Voltage:
- Increase damage as voltage increased
  - At 50V no damage most important
  - At 50V category 1 most frequent
  - At 250V categories 2 and 3 most frequent
  - At 1000V category 4 almost always occurs

Pulse sequences:
- Increase damage
  - A larger percentage show degraded behaviour due to increase damage — but not as rapidly as expected.

ACKNOWLEDGEMENTS

We wish to thank Roger Tomlinson of the Electronic Component Technology Group in the Electronic and Electrical Engineering Department here at Loughborough University of Technology for his valuable advice in setting up the experimental apparatus and the design and manufacture of some of the specialized components required in the circuits.

The probing equipment and the waters were provided by Plessey Research (Caswell) Ltd and we are especially grateful to Bill Holt, Ray Oakley and Nick Armstrong for their assistance.

This work has been carried out with the support of the Procurement Executive Ministry of Defence.

REFERENCES

2. R V Young. Cation-electrostatic discharge, a work IEEE, Trans Components, Hybrids and Manufacturing Technologics CHMT-5 (4) 512-515 (1972)

Authors biographies

Apath Amerasekera. Born in Sri Lanka Received a B Sc (Hons) degree in Electronic Engineering and Physics from Loughborough University in 1983 Is at present researching into the physics of failures of small dimension semiconductor devices at Loughborough University UK leading to a Ph D in 1986

David S Campbell. After a career in industry over 17 years culminating in being the Technical Manager of a large capacitor factory within the Plessey Group of companies he has been Professor of Electronic Component Technology in the Department of Electronic and Electrical Engineering at Loughborough University of Technology since 1971. He is now in charge of an active group in this area which has special interests in Interconnection Technology particularly thick film systems and also Reliability Analysis. The reliability work involves a major study on Field Failure of Electronic Components in Equipment a study which involves the University with major electronic companies in the U K and also with companies in Denmark through the involvement of the Danish Engineering Academy. Longby. Two studies are in hand on the reliability of components which are only now coming into use and these include Surface Mounting, Electrostatic Breakdown in MOS Silicon the effect on Failure Mechanisms of using submicron Silicon Devices and newly GaAs structures
APPENDIX 7

Oxide Breakdown in MOS Structures Under ESD and Continuous Voltage Stress Conditions
RELIABILITY TECHNOLOGY
Theory & Applications

European Reliability Conference - REE-Con'86
Copenhagen, Denmark, June 16-20, 1986
EUREL - DEF.CF.5

Edited by
J. MØLTOFT and F. JENSEN
The Engineering Academy of Denmark,
Lyngby, Denmark

Participants Edition - Part 2
ADVANCED PROFESSIONAL SEMINARS

NORTH-HOLLAND
AMSTERDAM • NEW YORK • OXFORD • TOKYO
STOR-A-FILE IMAGING LTD

DOCUMENTS OF POOR ORIGINAL HARD COPY

Date 12/02/08

Authorized by Simon Cockbill

Issue 2

Page 1 of 1

If in your copy, this page is UNCONTROLLED and only valid on date of issue 16-May-08
OXIDE BREAKDOWN IN MOS STRUCTURES UNDER ESD AND CONTINUOUS VOLTAGE STRESS CONDITIONS

E A AMERASEKERA and J S CAMPBELL
Electronic and Electrical Engineering Department
Loughborough University of Technology, Loughborough, Leicestershire, UK

Investigations have been conducted into the oxide breakdown mechanisms caused by the application of large voltage stresses in the form of both Electrostatic Discharge (ESD) pulses and continuous d.c. voltages, on metal-oxide-semiconductor (MOS) structures.

The MOS structures had been fabricated on p-type silicon wafers and had an oxide thickness of 2 400 Å. Both enhancement-type and depletion-type n-channel MOSFETs and MOS capacitors were constructed on the same wafer.

Earlier experiments on the ESD susceptibility of n-channel enhancement mode MOSFETs on the same wafer have shown there to be no temperature dependence of the breakdown although a significant voltage dependence was observed.

The experiments presented in this paper have investigated the correlation between the ESD sensitivity of MOS oxides and the continuous d.c. voltage breakdown strength, with the intention of aiding the identification of the principle mechanisms associated with each type of applied stress.

MOS capacitors were subjected to both ESD pulses, in accordance with the "human-body" model as presented in MIL-STD-883C, and continuous d.c. voltages. The breakdown of MOS capacitors upon application of the ESD pulse occurred at 200 volts (Breakdown being taken as the condition where the parallel resistance of the capacitor becomes finite, i.e. reduced by a factor of approximately 100.)

Identical capacitors were then subjected to a continuous voltage for approximately 300 sec. Breakdown then occurred at 16 volts. This is a factor of approximately 5 less than the sensitivity to ESD pulsing, indicating that two different oxide breakdown mechanisms are involved.

Such observations are in keeping with the time-independent nature of the ESD stress voltage, and the time-dependent nature of the continuous stress voltage. The time-dependent breakdown implies that the mechanism involved is that of impact-ionization of charge carriers within the oxide under high-field conditions which leads to thermal runaway and then burnout. This mechanism is temperature-independent. Time-dependent breakdown suggests that the generation of electron-traps leads to ultimate failure. As this mechanism is temperature-dependent, tests were carried out to investigate the temperature-dependence of continuous voltage breakdown. The results showed that continuous voltage breakdown is significantly temperature dependent between 25°C and 200°C, the breakdown voltage for a 40A oxide decreasing from 36V to 28V over this range.

1 INTRODUCTION

The sensitivity of MOS devices to Electrical Overstress (EOS)/Electrostatic Discharge (ESD) damage has long been established. However, the extent of device susceptibility is not well defined because of the difficulties in specifically locating and identifying a particular breakdown mechanism.

It is intended in this paper to aid the identification of the principle physical processes by which EOS and ESD manifest themselves in metal-oxide-semiconductor (MOS) devices. Once the particular mechanism by which breakdown takes place is understood, it becomes easier to predict the effect of the process on the different characteristics of the device.

In an earlier paper, experiments on the sensitivity of NMOS devices to ESD were described. As a result of these experiments, it was shown that the ESD sensitivity of enhancement mode NMOS devices was not temperature dependent. This is a result which shall be referred to later in this paper in context with the results presented here.

The experiments described here were conducted on both Enhancement-mode and Depletion-mode MOS structures. Devices have been subjected to both ESD and continuous voltage stress and the temperature dependence of the oxide breakdown strength has been determined. It is therefore possible to compare the principle physical processes of the respective breakdown mechanisms.
The experiments

2.1 The MOS Devices

MOS Capacitors were used in these experiments. The capacitors had been fabricated in chips on complete p-type silicon wafers and have an oxide thickness of 2400Å. Both enhancement (E) type and depletion (D) type capacitors were available on the same chip on the wafer. The wafers themselves were provided by Plessey Research (Caswell). The capacitors consist of

(a) polysilicon on thermal SiO₂ on enhancement (p-type) silicon (C1).

(b) polysilicon on thermal SiO₂ on depletion (n-type) silicon (C2).

Figures 1(a) and 1(b) show cross sections of the two capacitors. In particular, the different charge concentrations in the E-type and D-type capacitors are emphasised.

- Fig. 1(a) E-Type Capacitor. The Boron implant ensures a positive threshold voltage Charge density at the surface is \(4 \times 10^{11} \text{cm}^{-2}\) holes (p).

- Fig. 1(b) D-Type Capacitor. The Phosphorous is implanted over the Boron. Hence the residual charge density at the surface is \(8.5 \times 10^{11} \text{cm}^{-2}\) electrons (n).

The dimensions of the devices were as follows:

- The area of the capacitors = \(49 \times 10^3 \text{(µm)^2}\)
- Peripherial length of the capacitors = \(910 \text{µm}\)
- Oxide thickness = 400Å

2.2 The Procedure

2.2.1 ESD

A set of capacitors was subjected to a single ESD pulse generated using the "human-body" model as described in MIL-STD 883 C, Method 3015.2. Voltages of magnitudes ranging from 75V to 250V were applied to the devices. The parallel resistance across the capacitor was measured before applying the pulse and after application of the pulse.

A typical resistance across the gate oxide of an undamaged MOS capacitor is of the order of \(10^4\)Ω. The resistance bridge used in this experiment however had a maximum measuring capacitance of 10pF. Breakdown was deemed to have occurred when a finite parallel resistance was measured across the capacitor. A typical value of such a resistance is of the order of \(10^2\) which is very much less than that of an undamaged capacitor.

In order to evaluate the effect of ionic impurities in the oxide on the capacitor breakdown strength, another set of capacitors was subjected to a High Temperature Reverse Bias (HTRB) screen. The HTRB screen consisted of applying a voltage of -12V at a temperature of 150°C across the capacitors for approximately 5 minutes. This has the effect of sweeping impurities out of the oxide. These capacitors were then subjected to the ESD pulses described above.

2.2.2 Continuous Voltage Stress

MOS capacitors of both E-type and D-type structures were subjected to a voltage of increasing magnitude of the form shown in Fig. 2. Each voltage step lasted approximately 5 minutes. The leakage current was measured at each voltage step, and a plot made of leakage current as a function of the applied voltage.

The temperature of the wafer was increased in steps from room temperature (25°C) to 75°C, 125°C, 150°C and 200°C, the experiment being repeated at each temperature. It was therefore possible to plot the continuous voltage breakdown strength as a function of the ambient temperature.


**Oxide Breakdown in MOS Structures under ESD**

### RESULTS

#### Electrostatic Discharge

Table 1 gives the results of the ESD experiments on E-type and D-type capacitors that have not been subjected to HTRB screens. Typical values of capacitance and resistance are given for a single MOS capacitor. However, some capacitors did show resistance values as low as 1kΩ with a capacitance of 1 pF when damaged. The upper limit was about 300 kΩ. Breakdown was always seen to occur at 200 volts.

Table 2 gives the results of the ESD experiments on E-type capacitors that have been subjected to an HTRB screen. The lower limit of resistance is still around 1kΩ with C = 1 pF. The upper limit now is about 3 MΩ.

These results can be compared with those obtained in ref. 1. The slight degradation characteristics of the MOS transistors in that paper correspond to the upper limit value of resistance in the temperature, associated with a still functioning capacitor. The catastrophic breakdown characteristic corresponds to the lower limit of resistance which is 1 kΩ associated with zero capacitance.

#### Continuous Voltage Stress

These results are summarised in Figures 3, 4 and 5. 5 capacitors of each type were subjected to the stress. Figures 3 and 4 show the oxide leakage current as a function of applied voltage for a typical E-type capacitor and a typical D-type capacitor respectively. The oxide leakage current is seen to increase at around 32V in both graphs and rises to approximately 5 A at 36V. Resistance and capacitance measurements made at this stage for the E-type capacitor in Fig. 3 gave -

- parallel resistance = 197 kΩ
- parallel capacitance = 18 pF

and for the D-type capacitor in Fig. 4,

- parallel resistance = 4 MΩ
- parallel capacitance = 10 pF

These values are comparable with the breakdown figures given in Table 1.

Increase in voltage just beyond 36V resulted in a rapid increase in leakage current and burn-out occurring. 36V can therefore be considered to be the oxide breakdown voltage due to the applied continuous voltage stress.

Fig 5 shows the oxide breakdown voltage of an E-type capacitor as a function of temperature. As can be seen, the breakdown strength drops sharply from 36V at 25°C to 28V at 200°C.

---

### Table 1

<table>
<thead>
<tr>
<th>CAPACITOR</th>
<th>NO. OF DEVICES</th>
<th>APPLIED VOLTAGE</th>
<th>TYPICAL RESISTANCE</th>
<th>TYPICAL CAPACITANCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>10</td>
<td>150</td>
<td>120</td>
<td>100</td>
</tr>
<tr>
<td>E-type</td>
<td>10</td>
<td>200</td>
<td>165</td>
<td>150</td>
</tr>
<tr>
<td>D-type</td>
<td>10</td>
<td>150</td>
<td>130</td>
<td>120</td>
</tr>
<tr>
<td>Depletion</td>
<td>10</td>
<td>150</td>
<td>130</td>
<td>90</td>
</tr>
</tbody>
</table>

---

### Table 2

<p>| RESULTS OF APPLIING ESD PULSES TO MOS CAPACITORS WHICH HAVE BEEN HTRB SCREENED |
|-----------------------------|-----------------------------|-----------------------------|-----------------------------|</p>
<table>
<thead>
<tr>
<th>CAPACITOR TYPE</th>
<th>NO. OF DEVICES</th>
<th>APPLIED VOLTAGE</th>
<th>TYPICAL RESISTANCE</th>
<th>TYPICAL CAPACITANCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enhanced</td>
<td>10</td>
<td>150</td>
<td>200</td>
<td>100</td>
</tr>
<tr>
<td>Enhanced</td>
<td>10</td>
<td>200</td>
<td>165</td>
<td>150</td>
</tr>
<tr>
<td>Depletion</td>
<td>10</td>
<td>150</td>
<td>130</td>
<td>100</td>
</tr>
</tbody>
</table>

---

Fig. 2 Waveform of the voltage used in the continuous voltage stress experiments.
Fig. 3 Oxide leakage current ($I_{\text{leak}}$) vs Applied voltage ($V_{\text{applied}}$) for a typical E-type MOS capacitor. The parallel resistance across the gate of the capacitor measured after 36V had been applied was 197kΩ, indicating that breakdown had occurred.

Fig. 4 Oxide leakage current ($I_{\text{leak}}$) vs Applied voltage ($V_{\text{applied}}$) for a typical D-type MOS capacitor. The parallel resistance across the gate of the capacitor measured after 36V had been applied was 4MΩ, indicating that breakdown had occurred.

Fig. 5 Oxide breakdown voltage as a function of temperature under continuous voltage stress conditions.
4 DISCUSSION

4.1 Oxide Breakdown

The breakdown voltage of 200V for an ESD pulse is 15 times greater than the 36V observed with a continuous voltage stress. The strong temperature dependence of the continuous voltage stress breakdown is in direct contrast to earlier results establishing the temperature independence of ESD breakdown. Therefore it is possible to conclude that two different oxide breakdown mechanisms are involved. As discussed in ref. 1, these two mechanisms are:

a) impact ionization
b) electron-trap generation.

4.1.1 Impact Ionization

Impact ionization occurs when free electrons in the conduction band have gained sufficient energy to enable them to transfer enough energy to a valence electron upon collision such that the electron is elevated to the conduction band. The whole process occurs in about 10^-15 seconds, which for the purposes of this paper can be considered time independent.

High electric fields (10^6 V/cm) are necessary to enable the electrons to gain the energies required.

4.1.2 Electron-trap Generation

The fabrication process can result in dangling (unattached) Si or O bonds being present in the oxide which then become traps for charge carriers. However, in the presence of an electric field, it is possible that the lattice structure can be deformed sufficiently that thermal vibrations would create new electron traps. The generation of new electron traps has been experimentally shown by Harari as a function of both electric field and temperature. Oxide breakdown in these experiments is time-dependent (100 seconds).

Once traps are established within an oxide, the application of an electric field sufficiently distorts the SiO2 conduction band enabling charge to hop from one trap to another. This process is known as hopping conduction.

Traps can also result in high electric fields occurring in localized areas of the SiO2, thereby enhancing the prospect of impact ionization occurring.

4.2 Electrostatic Discharge

The ESD pulse is of very short duration (10^-6 seconds) and any change produced by such a pulse can therefore be considered time-independent. The ESD breakdown process is also temperature independent. It can, therefore, be concluded that impact ionization is responsible for oxide breakdown.

As Table 4 it was shown that devices which had less ionic contamination in the oxide had higher ESD breakdown voltage at 250V. It is therefore possible for electron traps already available in the oxide before the application of the pulse to enhance the probability of impact ionization and thereby reduce the breakdown voltage of the devices.

4.3 Continuous Voltage Stress

The breakdown observed with continuous voltage stress is in keeping with Harari's results. The temperature dependence and the low electric fields (cf. ESD) signify that electron-trap generation and the associated conduction processes is the mechanism concerned.

4.4 Implications

ESD is shown to be due to a high-energy impact ionization breakdown mechanism. It is significant that electron traps already available in the oxide can contribute to ESD breakdown. Hence the device sensitivity to ESD would be a function of the number of electron-traps in the oxide. This has far reaching implications when considering device reliability at the fabrication stage. It is important to reduce the number of traps in an oxide at this stage and processing techniques such as those used in the manufacture of radiation hardened devices along with thermal annealing techniques, may provide the solution.

However, processing techniques such as ion implantation may result in a high number of electron traps being created in the oxide. Energies in the range 20keV to 100keV are presently quite common. With the requirements for higher ion implant doses at higher energies being used in the manufacture of sub-micron devices, it is possible that the number of traps thus created could have a significant effect on device reliability.

5. CONCLUSIONS

1. A large disparity was observed in the breakdown voltages of MOS capacitors due to

(a) Electrostatic Discharge pulsing (200V)
(b) Continuous d-c voltage stress (36V)

2. A strong temperature dependence was observed for oxide breakdown under continuous voltage stress compared with earlier results showing that oxide breakdown under ESD pulsing was not temperature dependent.

3. It is therefore concluded that two different mechanisms are responsible for oxide breakdown due to ESD pulsing and continuous voltage stress.

4. The time-independent property of the ESD pulse suggests that impact ionization is the mechanism through which breakdown takes place. High localized electric fields caused by electron traps already in existence within the oxide enhance the prospect of breakdown taking place. This has direct implications on device fabrication techniques involving high energy ion implantation.
5 It is postulated that oxide breakdown under continuous voltage stress is related to the generation of new electron traps within the oxide. The electric fields required for breakdown are, therefore, not as great as those required in the case of ESD.

6. ACKNOWLEDGEMENTS

Thanks are due to Plessey Research (Caswell) for the wafers used in the study. This work has been carried out with the support of the Procurement Executive, Ministry of Defence.

REFERENCES


BIBLIOGRAPHY

AJITH AMERASEKERA

Received a B Sc (Hons) degree in Electronic Engineering and Physics from Loughborough University in 1983. Is at present researching into the physics of failures of small dimension semiconductor devices at Loughborough University, UK, leading to a Ph.D in 1986.

DAVID S. CAMPBELL

After a career in industry over 17 years, culminating in being the Technical Manager of a large capacitor factory within the Plessey Group of companies, he has been Professor of Electronic Component Technology in the Department of Electronic and Electrical Engineering at Loughborough University of Technology since 1971. He is now in charge of an active group in this area which has special interests in Interconnection Technology, particularly thick film systems and also Reliability Analysis. The reliability work involves a major study on Field Failure of Electronic Components in Equipment, a study which involves the University with major electronic companies in the U.K. and also with companies in Denmark through the involvement of the Danish Engineering Academy at Lyngby. Also studies are in hand on the reliability of components which are only now coming into use and these include Surface Mounting, Electrostatic Breakdown in MOS Silicon, the effect on Failure Mechanisms of using Submicron Silicon Devices and finally GaAs structures.
DOCUMENTS OF POOR ORIGINAL HARD COPY
APPENDIX 8

A Comparison Between GaAs MESFET and Si NMOS ESD Behaviour
- Proc. ERA Seminar on ESD in Electronics
A COMPARISON BETWEEN GaAs MESFET AND Si NMOS ESD BEHAVIOUR

A.J. Franklin, E.A. Amerasekera and D.S. Campbell
Department of Electronic and Electrical Engineering,
University of Technology, Loughborough, Leicestershire, UK

ABSTRACT

Work is in progress at Loughborough to investigate ESD sensitivity of GaAs D-MESFETs and unprotected enhancement mode NMOS structures.

The work to date has revealed that NMOS structures can be severely degraded with ESD pulse above 200V as compared with 600V for GaAs MESFETs. It has also been shown that both NMOS and GaAs structures are polarity sensitive.

The breakdown of the oxide for NMOS devices can be explained by impact ionisation. The effect of the Schottky barrier to explain the polarity behaviour of GaAs is discussed.

1 INTRODUCTION

It has been recognised for some time that one of the major failure mechanisms in semiconductors was electrical overstress (EOS) which includes device failure due to Electrostatic discharge pulses (ESD).

Equipment manufacturers, users and semiconductor manufacturers now estimate that 60% of semiconductor failures can now be attributed to EOS, a large percentage of this being due to ESD.

In order to identify the failure mechanisms associated with unprotected MESFET and NMOS FETs, an investigation has been carried out and the results are presented in this paper.

2 BASIC STRUCTURE

The two basic structures that have been investigated are n-channel, depletion mode metal-semiconductor GaAs FETs and n-channel, enhancement mode metal-oxide-semiconductor Si FETs. Figure la and b, shows the structures and it can be seen that they are radically different.

The D-MESFET has a n-doped channel and the source and drain contacts are made using an ohmic metal. The gate contact is formed with a metal laid directly onto the surface of the GaAs forming a Schottky depletion region beneath the area of the gate contact.

The NMOS device has a lightly p-doped channel and has n wells diffused to form the source and drain contacts. An oxide of 400Å is grown on the wafer, and the gate is formed with n+polysilicon laid on the oxide.

Electrical modulation of the drain current is fundamentally different in each case.

Depletion MESFET operation uses the action of the Schottky depletion region, under negative voltage bias, to deplete the channel of carriers and hence decrease the current flow. An enhancement device can be achieved using a
shallow implant of the n-type channel or by a received gate structure.

The enhancement NMOS device is modulated by a positive voltage on the gate which induces an inversion layer beneath the gate oxide which electrically connects the source to the drain, allowing current to flow.

3 EXPERIMENTAL WORK

3.1 INTRODUCTION

The present study has been concerned with ESD pulses applied to the gate of the device under investigation, resulting in a breakdown which will have different causes in the two devices. The subject of ESD modelling has become of major importance, since various research centres have differing views on the model to be adopted.

In the present work the ESD pulses are derived using the MIL-STD-883C, "Human body model". The human body model comprises of a circuit in which a high voltage supply charges a 100pF capacitor through a 1MΩ resistor. The discharge is through a 1.5kΩ (body) resistance in series with the device under test. The charge/discharge circuits are selected by means of a mercury wetted 2 pole relay.

The pulse generated using the human body model is of the shape shown in Figure 2 as observed on an oscilloscope and using a low capacitance probe. The rise time of the pulse, t\[\text{rise}\] is less than 500ns, and the decay is approximately 1µs. In practice this value will vary depending on the load offered by the device to the discharge circuit.

3.2 RELATED DEVICE GEOMETRIES

The GaAs MESFET geometry is shown in Figure 3. It comprises of two AuGeNi ohmic contacts and a CrAu gate of dimensions 1 x 150µm. The gate/source-drain dimensions were 1µm and 2µm respectively. They are fabricated on a 2" Si-1on implanted GaAs LEC wafer which contains 1500 D-FET structures. In the majority of cases the whole active area was overlayed with polyimide.

The NMOS structures are fabricated as 3" p-type Si wafers which contained varying gate dimension transistors, MOS capacitors and integrated circuits. The FETs have polysilicon n+ gate and alloyed AlSiCu drain and source contacts. They are set in rows, each with a common source and a common gate.

3.3 PROCEDURE

The main part of the experimental programme consisted of pulsing the gate(s) with varying polarity voltages. The source and drain electrodes were either earthed (GaAs) or left floating (NMOS), with the substrate grounded.

The gate was pulsed with voltages between ±50 up to ±1kv, and in both cases this gave a spread of breakdown characteristics.

After the application of each pulse the devices were analysed electrically for any change in the \(I_D \text{ vs } V_{DS} (I/V)\) characteristics. The MESFETs were also examined with an optical microscope mounted on the wafer prober. Finally the MESFETs were examined for detailed breakdown changes with the scanning electron microscope.

4.3.2
4 RESULTS

4.1 CHARACTERISTICS

In both cases the degradation after the application of one ESD pulse to the gate could be classified into one of 4 categories, although with the NMOS devices the behaviour is more complex. These categories are based on the electrical I/V curves for the transistors before and after each pulse.

4.1.1 Category 1 - No change

All electrical parameters remain unaltered and the I/V curves, as seen in Figure 4a and b remain the same.

4.1.2 Category 2 - Degraded Characteristic

Typical I/V curves are shown in Figure 5a and b. The degradation manifests itself as a decrease in $q_m$ and an increase in $I_{GL}$, the gate leakage current to source or drain.

4.1.3 Category 3 - Total Burnout

This state is reached in the case of the MESFET when the I/V curves show constant resistive properties as in Figure 7. In the case of the NMOS structures total burnout is defined when $I_d = 0$ for all values of $V_{DS}$ and $V_{GS}$ or constant resistive properties for all drain/source configurations.

4.1.4 Category 4 - Partial Burnout

This category is part way between category 2 and 3. In both devices the same I/V curves are encountered, an example of which is given in Figure 6a. If the drain and source voltages are reversed the curves in Figure 6b are obtained.

4.2 VOLTAGE EFFECTS

The voltages applied to the gate have a significant effect on the degradation of the devices. It should be noted that it was not found possible to predict the exact behaviour of any one device and the figures given must be interpreted as statistical results found by measuring large numbers of devices.

At voltages between +200 and +800v the MESFETs fall into category 2, a degraded characteristic after the application of a single pulse. The NMOS FETs fall into category 2 between +50 and +200v. In both cases voltages lower than the stated minimum put the devices in category 1. (No change observed.)

Category 4, partial burnout, was encountered for the MESFETs between +800 and +1kv and for the NMOS FETs between +200 and +300v.

Finally, category 3, total burnout, was observed at voltages greater than +1kv (MESFETs) and +300v (NMOS FETs).
4.3 **TEMPERATURE EFFECTS**

No detailed studies of the effect of temperature on changes in the I/V characteristic caused by a single ESD pulse have been made for the MESFETs. However for the MOSFET devices it has been shown that temperature change has a negligible effect on the voltage degradation behaviour as described in paragraph 4.2.

4.4 **CUMULATIVE PULSE EFFECTS**

The cumulative effects of pulsing the gates of the devices resulted in degradation (category 2), eventually leading to partial burnout (category 4) and in some cases total burnout (category 3). The extent of degradation could not be predicted but there was an overall dependence on the voltage magnitude and polarity of the pulse. Figure 8a, b, and c shows the degradation of MESFETs subject to +600v pulses on a function of the number of pulses applied. It can be seen that the transconductance is decreasing and in many cases the transistors degraded to the single I/V curve normally associated with zero VGS, for all values of VGS. The NMOS FETs degraded to zero IDS for all VDS, VGS, or degraded until total burnout was obtained.

4.5 **POLARITY EFFECTS**

In both cases a polarity effect was evident. In section 4.2 the voltage magnitudes quoted for a given degradation category can be roughly halved for negative polarity pulses (i.e. partial burnout is encountered at voltages between -400 and -500 (MESFETs) and -100 and -150 (NMOS FETs). This polarity effect is shown in Figure 8 as a function of pulse number and polarity. It can be seen that the cumulative failure rate is greater for the negative pulses.

4.6 **PHYSICAL CHANGES AFTER THE APPLICATION OF AN ESD PULSE**

After the application of an ESD pulse to the gate of the MESFET any changes in appearance along the gate or in the channel was noted. Due to the vertical nature of the NMOS FETs this was not possible. After the application of a pulse of sufficient magnitude to cause severe degradation, or partial burnout, a discolouration of the gate/source region occurred. If total burnout was encountered the discoloration extended across the channel region from drain to source. In general, Partial burnout resulted in the discoloration between source and gate, but sometimes it occurred between gate and drain. This was traced to misalignment of the gate, moving the gate nearer the drain than the source.

SEM analysis of the channel region, after removal of the polyimide, revealed greater damage to the source electrode for a negative pulse, and the gate electrode for a positive pulse.

5 **DISCUSSION**

5.1 **DEGRADATION CHARACTERISTICS**

In both devices, degraded characteristics were manifested as a decrease in transconductance. This can be explained by a leakage path existing in:-

(a) The channel between gate and source (MESFET)

(b) The oxide either to the source, drain or channel (NMOS FET)
In the MESFET the leakage resistance acts like a potential divider and a proportion of the voltage which would act on the gate depletion region is dropped across this leakage resistance. Hence pinch-off of the device requires a greater voltage. The degraded transconductance is therefore a measure of the magnitude of the leakage path.

The NMOS FET breakdown is more complex, and the degradation characteristic (i.e. category 2, 3 or 4) will depend on the position of the leakage path in the oxide. If the breakdown was between oxide and the drain or source then partial burnout would be obtained. If the breakdown was directly over the channel then total burnout would be obtained. In the later case the presence of the leakage path in the oxide will not allow the inversion layer to form. In the former case, the leakage path will simply drop a proportion of the gate voltage as in the MESFET case.

In both cases, after a single low voltage pulse, the devices that show 'degraded characteristics' would still be capable of digital operation. They would, however, be potential failures later in life. Such devices are thus often referred to as "walking wounded". In such cases, the leakage resistance of both types of devices will still be relatively large and, therefore, device operation is still possible.

5.2 CUMULATIVE PULSE EFFECTS

In both devices cumulative degradation effects were present. It is postulated that the MESFETs degraded because the leakage path resistance decreased with each pulse. Hence the voltage dropped across this leakage resistance increased with each pulse, thus decreasing the transconductance.

The NMOS FETS degraded as explained in section 5.1 but the degradation category depended on the nature of the leakage path within the oxide.

5.3 POLARITY EFFECTS ON ESD SENSITIVITY

In both types of devices it was found that a negative polarity would tend to degrade the device to a greater extent than a positive pulse.

In the MESFET case, the action of the Schottky barrier under voltage stress is different for each polarity. This would account for the observed effect. In the case of the NMOS FETs the breakdown is through the oxide and is supposedly polarity independent. Studies on MOS capacitor structures has lead to the conclusion that surface charge effects under the gate oxide will affect the extent of breakdown for a given voltage and this will involve polarity behaviour.

5.4 PHYSICAL CHANGES DUE TO ESD BREAKDOWN

The GaAs MESFETs burnt out between gate and source and this can be explained by considering the geometrical arrangement of the channel/gate region. (The gate/source spacing is 1µm while the gate/drain is 2µm.) Since both of these electrodes are earthed then the highest field will be created between gate and source. It is postulated that the ohmic contacts have uneven edges which may also extend into the channel region creating localised high field regions. The breakdown will initiate at the opposite electrode to the current injection as observed in the SEM micrographs. This effect is also reported with D.C. drain/source breakdown.

The NMOS FETS showed no observable changes after pulsing.

4.3.5
5.5 SCHOTTKY BARRIER EFFECTS

A detailed understanding of the part played by the Schottky barrier in the degradation of MESFET devices subject to ESD pulses, is being evolved. However, data is needed on both temperature and continuous voltage effects before a consistent model can be presented. Detailed results in this area will be presented shortly.

5.6 OXIDE BREAKDOWN

Analysis of the results that have been obtained indicate that leakage paths created in the oxide can account for the changes in the I/V curves obtained. The position of the breakdown in the oxide will determine the changes in breakdown characteristics.

Previous work\(^4,5,6,8\) indicated that oxide breakdown under transient high voltage stress was due to an electron cascade caused by avalanche breakdown (impact ionisation) within the oxide. Oxide defects, present after growth, will influence the point of breakdown within the oxide and also surface charge effects\(^7\) will affect the breakdown point.

6 CONCLUSIONS

Degradation and burnout characteristics of GaAs MESFETs and Si NMOS FETs have been investigated.

It is concluded that degradation in the MESFETs is due to a leakage path forming at a high field point in the gate/source region. The probability of failure is dictated by the polarity and magnitude of the applied ESD pulse.

In the NMOS FETs, degradation is due to oxide breakdown by impact ionisation within the oxide. The position and probability of failure mode is also influenced by polarity and magnitude of the applied ESD pulse but for different causes.

7 ACKNOWLEDGEMENTS

We wish to thank Plessey Research (Caswell) Ltd, for supplying the NMOS test wafers and STC Components Ltd. for the GaAs MESFET wafers.

One of us (AJF) is grateful for support from STC Components Group and the other (EAA) is grateful for support from the MOD Procurement Executive (DCVD).

In detail we would like to thank Bill Holt, Roy Oakley and Nick Armstrong from Plessey and Richard Butlin, Ron Eggar and Steve Bland from STC Ltd. for their support and advice.

8 REFERENCES


Figure 1a  MESFET Cross-sectional Structure

Figure 1b  N-MOSFET Cross-sectional Structure
Volts

$t_{\text{RISE}} = 500\text{ns}$  $t_{\text{DECAY}} = 1\mu s$  (time)

Figure 2 Typical ESD "Human Body Pulse"

Figure 3 GaAs MESFET Plan Geometry
Figure 4 Typical I/V Plot for Both Devices

Figure 5 Degradation (Category 2) I/V Plot for Both Devices
Figure 6a, b  Partial Burnout Characteristics

Figure 7  Full Burnout Characteristics
Figure 8 Degradation of the transconductance for a GaAs MESFET as a function of the number of ESD pulses applied to the gate.
Figure 9  Transconductance degradation ($g_m$) as a function of voltage and polarity for a GaAs MESFET
APPENDIX 9

ESD Pulse and Continuous Voltage Breakdown in MOS Capacitor Structures
- Proc. EOS/ESD Symposium,
  Las Vegas, September 1986.
ABSTRACT

ESD PULSE AND CONTINUOUS VOLTAGE BREAKDOWN

IN MOS CAPACITOR STRUCTURES

E.A. Amerasekera and D.S. Campbell

Department of Electronic and Electrical Engineering

Loughborough University of Technology

Loughborough Leicestershre U K

Telephone in U K 0509 222811

Telephone outside U K 44 509 222811

Previous reported work by the authors on ESD pulse and D.C. breakdown in unprotected NMOS and PMOS structures has been extended by studying unprotected single n and p type MOS capacitor structures. The effect of polarity on the breakdown of both n and p structures has been examined and the results have been analysed and compared with ESD. A model has been developed to physically explain the results.
1. INTRODUCTION

As devices get smaller in size, and the device density per i.c. increases, the problem of EOS/ESD becomes significant to the device manufacturer. The existing solution of providing protection circuitry against overstress voltages uses up an increasingly larger proportion of the chip area and can become a limiting factor in Ultra Large Scale Integration (ULSI). It is therefore important to obtain an understanding of the physics of the breakdown process under high stress conditions. Identification of possible weak points can then be made and by altering the fabrication process or changing the device structure, less sensitive devices can be manufactured.

Previous work published by the authors [1,2] has shown that:-

(a) The ESD sensitivity of NMOS transistors is not temperature dependent, but that oxide breakdown is highly temperature dependent when subjected to continuous voltage stressing.

(b) The voltages required for breakdown with an ESD pulse is almost a factor of five greater than the breakdown voltage in the d.c. case.

Hence, it was suggested that ESD pulses resulted in breakdown through impact ionization within the SiO₂, a process which is aided by existing electron traps in SiO₂. Continuous voltage stressing generated electron traps within the oxide which encouraged hopping conduction through the oxide and eventual terminal breakdown.

In this paper, further experiments carried out on MOS capacitor structures are described. Both n and p-type capacitors were subjected to positive and negative polarity voltage stresses. Pulsed voltages (simulating ESD conditions) and continuous voltages (simulating EOS conditions) were applied to these devices. As a result of these experiments it has been possible to identify the principle features of the gate-oxide capacitor structures which influence the oxide breakdown process under EOS/ESD conditions.
2. THE EXPERIMENTS

2.1 The Capacitors

The MOS capacitors used in these experiments were fabricated on p-type and n-type silicon substrates, with an oxide thickness of 400Å. The p-type substrates had both enhancement (E)-type capacitors and Depletion (D)-type capacitors; i.e. the silicon surface in the E-type cases was preferentially doped with $p^+$ (Boron), while in the D-type cases the silicon surface was doped n (Phosphorous). The n-type substrates only supported E-type capacitors. Therefore, the three capacitor structures were:

(a) $n^+$ polysilicon on $SiO_2$ on $p^+$ (Figure 1(a))

(b) $n^+$ polysilicon on $SiO_2$ on n-Si (Figure 1(b))

(c) $n^+$ polysilicon on $SiO_2$ on n$^+$-Si (Figure 1(c))

The dopant concentrations as follows:

(1) $p^+$-Si = $4.0 \times 10^{11}$ cm$^{-2}$

(11) n-Si = $8.5 \times 10^{11}$ cm$^{-2}$

(111) n$^+$-Si = $10.0 \times 10^{11}$ cm$^{-2}$

(v1) $n^+$ poly Si = $2.0 \times 10^{16}$ cm$^{-2}$

Device dimensions were:

For p-substrate capacitors,

Area = $49 \times 10^3$ $\mu$m$^2$

Periphery = 910 $\mu$m

For n-substrate capacitors,

Area = $4.9 \times 10^3$ $\mu$m$^2$
A total of 10 capacitors of each type were subjected to each voltage stress.

2.2 The Procedure

2.2.1 ESD

The ESD pulse was generated using the "human-body" model circuit as described in MIL-STD-883C, Method 3015.2 [3]. Single voltage pulses of magnitudes ranging from 50 volts to 250 volts of both polarities were applied to the devices. The parallel resistance across the capacitor was measured before and after application of the pulse.

A typical gate-oxide resistance is of the order of $10^{14}$ $\Omega$.

The resistance bridge used in this experiment however had a maximum measuring capacitance of $10^{9}$ $\Omega$. Breakdown was deemed to have occurred when a finite parallel resistance was measured across the capacitor. A typical value of such a resistance is of the order of $5 \times 10^5$ $\Omega$ which is very much less than that of an undamaged capacitor.

2.2.2 EOS

The MOS capacitor structures were subjected to a voltage of increasing magnitude of the form shown in Figure 2. Each voltage step lasted approximately 5 minutes. Plots were made of the gate leakage-current as a function of the applied voltage.
3. **RESULTS**

3.2 **ESD**

Table 1 presents the results of the ESD experiments on E-type and D-type p-Si and E-type n-Si MOS capacitors. Values of capacitance and resistance are given for a typical capacitor in each category. However, some capacitors did show resistance values as low as 1kΩ with a capacitance of ~0 pF when damaged. The upper resistance limit was about 300 kΩ.

Breakdown of the E-type p-Si capacitors consistently occurred upon application of a single positive pulse of +200 volts, or a single negative pulse of -100 volts. With a D-type p-Si capacitor the disparity was less, negative pulse breakdown still occurred at around -100 volts, with positive pulse breakdown at 175 volts. E-type n-Si capacitors showed no polarity dependence. Because of the structural differences between the p-Si and the n-Si capacitors it is not possible to compare directly the absolute values of the respective breakdown voltages. However, the relative disparities in the positive and negative polarity breakdown voltages are comparable.

C-V curves were made before and after the devices were subjected to an ESD pulse. Figure 3 shows a C-V curve for an E-mode p-Si capacitor before and after application of a +150 volt ESD pulse. Figure 4 shows a C-V curve for an E-mode n-Si capacitor before and after application of a +50 volt pulse. The p-Si capacitor appears unaffected by the sub-breakdown threshold voltage while the n-Si capacitor is significantly affected.

3.2 **EOS**

The gate-oxide leakage current ($I_{g \text{ leak}}$), as a function of the applied voltage, is shown for p-Si capacitors in Figure 5 and for n-Si capacitors in Figure 6. It must again be emphasised here that the p-Si capacitors differ in size from the n-Si capacitors. The absolute values of the leakage currents and breakdown voltages of the two different structures cannot be directly compared. What is important are the rates at which $I_{g \text{ leak}}$ increases at the onset of breakdown.
In Figure 5, the curve for an E-type capacitor shows a large polarity effect. A sharp increase in $I_g$ leak is observed at around -30 volts, compared with a much more gradual increase in $I_g$ leak with positive voltage stress. This asymmetry is less obvious for D-type p-Si capacitors, where the rate of increase in $I_g$ leak at negative voltages is similar to that for E-type capacitors, but the positive voltages show a much sharper increase in $I_g$ leak. E-type n-Si capacitors show great symmetry between positive and negative $I_g$ leak values, indicating that the polarity of the applied voltage had an almost negligible effect on the breakdown threshold voltage of these structures in comparison to the p-Si capacitors.

Typical C-V curves as a result of positive and negative D.C. voltage stresses are shown in Figures 7 and 8 respectively for E-type p-Si capacitors. A positive voltage stress of +36 volts is seen to increase the capacitance minima and distort the well of the C-V curve but no lateral shift along the voltage axis is observed. By comparison a negative voltage stress of -30 volts resulted in a large lateral shift in the negative direction along the voltage axis, but otherwise no distortion of the C-V curve was observed.

4. DISCUSSION

4.1 Introduction

The major point in section 3 above, was the extent of the polarity effect on the breakdown thresholds of different MOS capacitor structures under stress conditions simulating both EOS and ESD.

Previous workers have observed that MOS structures with n-type surfaces had lower breakdown thresholds than p-type surfaces [4, 6]. There is also evidence to show that the breakdown threshold with an $n^+$ polysilicon gate as cathode is lower than when a less n-type surface is the cathode [7]. This disparity between the breakdown thresholds of the p and n-type structures has been attributed to intrinsically higher defect densities in SiO$_2$ film grown on n-type wafers, and the use of different oxidation furnaces for n and p-type materials [5]. However this explanation is not consistent with both the d.c. and pulsed voltage stress observations outlined in Section 3, which indicate that the injection of electrons into the SiO$_2$ must be considered [8].
4.2 Continuous Voltage Stress Breakdown

Electrons can be injected across the Si-SiO$_2$ potential barrier (3.1eV) either by Poole-Frankel (P-F) emission [9, 10] or by quantum mechanical tunnelling [11, 12].

In the case of P-F emission, trapped charge-carriers gain sufficient thermal energy from the external applied electric field to overcome the potential barrier at the interface. Since the mechanism is thermionic, it must be significantly affected by temperature. Experimental observations, described in an earlier paper, do show this to be true for MOS capacitor structures under continuous voltage stress conditions [2]. The effect of the stress voltages both positive and negative on the C-V curves indicate that interface states are being filled or emptied depending on the polarity of the applied voltage and the initial condition of the state [9, 13]. Charge movement across the interface may therefore, be considered to be taking place under D.C. stress conditions.

The tunnelling current at the interface is dependent on the electron density at the surface by means of the supply function $N(E_x)$, since

$$J = e \int_0^\infty D(E_x)N(E_x)dE_x \text{Am}^{-2} \quad \text{Eq. (1)}$$

where $E_x$ is the energy of the incident electron, $D(E_x)$ is the transmission probability and $N(E_x)$ is the number of electrons in the energy range $E_x$ to $E_x + dE_x$ available for tunnelling through the barrier. $N(E_x)$ is based on occupation statistics [14, 16]. An n-type surface under the influence of a high electric field (∼10 MV cm$^{-1}$) would have a higher supply function than a p-type surface in the same electric field. Also the higher number of occupied states in the n-type surface would imply a higher $D(E_x)$ compared to a p-type surface [8]. Hence, the tunnelling current for electrons at the Si-SiO$_2$ interface would be greater for an n-type accumulation surface compared to a p-type inversion surface under the same biasing conditions.
In summary therefore EOS breakdown is initiated by electrons being injected from the Si into the SiO₂. The higher the concentration of electrons in the silicon \( N_D \), the greater the leakage current, giving rise to the \( I_g \) leak vs \( V_{appl} \) curves obtained for the three MOS capacitor structures.

4.3 Pulsed Voltage Stress Conditions

The result of the ESD experiments showed that the breakdown thresholds for p-type MOS structures with positive voltages were almost twice as much as the negative voltage breakdown threshold. This difference became less as the silicon surface was made more n-type, again suggesting a charge injection effect. However, the effect of the voltage pulse on the surface must be considered in the analysis [18].

When a large positive pulse is applied across an n-type capacitor structure, the surface accumulates majority carriers (i.e. electrons). The majority carrier response time in silicon is about \( 10^{-12} \) secs [17], hence the surface has time to go into strong accumulation within the rise time of the applied pulse (<500 ns). Electrons, injected from the strongly accumulated silicon surface, together with electrons from traps and impurity sites already on the \( SiO_2 \) tunnel through to the \( SiO_2 \) conduction band. Here impact ionization takes place and results in oxide damage [2]. In the case of a negative polarity ESD pulse, the \( n^+ \) polysilicon gate becomes the injecting electrode. This surface has a large enough concentration of electrons to be able to match the process described for a positive ESD pulse. The breakdown thresholds are, therefore, very similar in the two cases for an n-type surface.

In a p-type surface, an inversion layer consisting of electrons is induced by a large positive voltage. Here the electrons are minority carriers and their response time is of the order \( 10^{-2} \) secs to 1 sec, which is outside the 500 ns rise time of the applied pulse. Initial injection of charge carriers into the \( SiO_2 \) conduction band, therefore, takes place from impurity and trap sites in the \( SiO_2 \) itself. However, during the decay time of the pulse (dependent on the resistance and capacitance of the oxide) the surface has time to go into strong inversion. By this time, though, the electric field across the oxide would have decreased thereby affecting the charge injection current.
The consequence is that p-Si capacitors would have a higher breakdown threshold than n-Si capacitors. A negative voltage pulse would have a breakdown threshold similar to that of an n-type structure because the injecting electrode would be the n⁺ poly Si gate. Therefore, the positive voltage breakdown threshold for p-Si capacitors is higher than the negative voltage breakdown thresholds. The more p-type the silicon surface, the higher the breakdown threshold is likely to be.
5. **CONCLUSIONS**

(1) A significant polarity effect was observed for the oxide breakdown thresholds of p-Si and n-Si MOS capacitors under both continuous and pulsed voltage stress conditions, simulating EOS and ESD respectively.

(2) Under EOS conditions it was found that the more p-type the silicon surface, the lower the gate leakage current with applied voltage.

(3) Under ESD conditions n-Si MOS capacitors were shown to be more sensitive to damage than p-Si capacitors. Polarity effects were negligible on n-Si capacitors, while the breakdown threshold of p-Si capacitors with negative ESD pulses was almost half the magnitude (100 volts) of the positive breakdown threshold (200 volts).

(4) A model of oxide breakdown based on charge injection processes at the Si-SiO₂ interfaces has been proposed.

(5) For EOS conditions, the injection current is proportional to the electron supply function which increases with more n-type doping.

(6) For ESD conditions, the majority and minority carrier in accumulated and invested surfaces, respectively, must be considered. When electrons are majority carriers (i.e. in n-type surfaces), the response time to the applied pulse is fast enough to respond immediately to the voltage. On the other hand, when electrons are minority carriers (i.e. p-type surfaces) the response time is much slower than the pulse rise time. As a consequence, strong inversion only occurs within the decay time of the pulse, and therefore the effective electric field is lower than that applied to the device. Hence, p-Si capacitors have a higher oxide breakdown threshold from n-Si capacitors.

(7) These conclusions are experimentally justified.
6. ACKNOWLEDGEMENTS

We thank Roger Tomlinson for his help in setting up some of the experimentation. Plessey Research (Caswell) Ltd., provided the wafers used in this work and we are grateful to them.

This work has been carried out with the support of the Procurement Executive, Ministry of Defence, U.K.
7. REFERENCES


# TABLE 1

RESULTS OF ESD EXPERIMENTS USING POSITIVE AND NEGATIVE VOLTAGE PULSES ON E-TYPE AND D-TYPE NMOS AND E-TYPE DMOS CAPACITORS

<table>
<thead>
<tr>
<th>CAPACITOR TYPE</th>
<th>NO. OF DEVICES</th>
<th>APPLIED VOLTAGE (V)</th>
<th>TYPICAL CAPACITANCE (pF)</th>
<th>TYPICAL RESISTANCE (kΩ)</th>
<th>COMMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>E-TYPE NMOS</td>
<td>10</td>
<td>+150</td>
<td>20.0</td>
<td>O/C</td>
<td>O/C</td>
</tr>
<tr>
<td>E-TYPE NMOS</td>
<td>10</td>
<td>+200</td>
<td>20.0</td>
<td>O/C</td>
<td>150</td>
</tr>
<tr>
<td>E-TYPE NMOS</td>
<td>10</td>
<td>-75V</td>
<td>20.0</td>
<td>O/C</td>
<td>O/C</td>
</tr>
<tr>
<td>E-TYPE NMOS</td>
<td>10</td>
<td>-100V</td>
<td>20.0</td>
<td>O/C</td>
<td>65</td>
</tr>
<tr>
<td>D-TYPE NMOS</td>
<td>10</td>
<td>+150</td>
<td>13.0</td>
<td>O/C</td>
<td>O/C</td>
</tr>
<tr>
<td>D-TYPE NMOS</td>
<td>10</td>
<td>+175</td>
<td>13.0</td>
<td>O/C</td>
<td>50</td>
</tr>
<tr>
<td>D-TYPE NMOS</td>
<td>10</td>
<td>-75V</td>
<td>13.0</td>
<td>O/C</td>
<td>O/C</td>
</tr>
<tr>
<td>D-TYPE NMOS</td>
<td>10</td>
<td>-100V</td>
<td>13.0</td>
<td>O/C</td>
<td>50</td>
</tr>
<tr>
<td>E-TYPE PMOS</td>
<td>10</td>
<td>+50V</td>
<td>11.5</td>
<td>O/C</td>
<td>O/C</td>
</tr>
<tr>
<td>E-TYPE PMOS</td>
<td>10</td>
<td>+75V</td>
<td>11.5</td>
<td>O/C</td>
<td>75</td>
</tr>
<tr>
<td>E-TYPE PMOS</td>
<td>10</td>
<td>-50V</td>
<td>11.0</td>
<td>O/C</td>
<td>O/C</td>
</tr>
<tr>
<td>E-TYPE PMOS</td>
<td>10</td>
<td>-60V</td>
<td>11.5</td>
<td>O/C</td>
<td>2.16M</td>
</tr>
</tbody>
</table>
Figure 1(c) An E-Mode n-Si MOS Capacitor, indicating the n⁺-type Arsenic implant at the Semiconductor Surface
Fig 2 - Form of the continuous voltage stress as applied to MOS structures.
Figure 3(a) C-V Curve of E-Mode p-Si MOS Capacitor, undamaged.
Figure 3(b) C-V Curve of E-Mode p-Si MOS Capacitor after application of a single ESD pulse of +150V
Figure 4(a) C-V Curve of an Undamaged E-Mode n-Si MOS Capacitor
Figure 4(b) C-V Curve of an E-Mode n-Si Capacitor after application of an ESD Pulse of +50V
Figure 5 Graph of oxide leakage current ($I_{\text{g leak}}$) as a function of the applied stress voltage for p-Si MOS Capacitors
Figure 6 Graph of Oxide Leakage Current ($I_{\text{leak}}$) as a Function of the Applied Stress Voltage for an E-Mode n-Si capacitor
Figure 7  C-V Curve of E-Mode p-Si MOS Capacitor after Continuous Voltage Stress at +36V
Figure 8  C-V Curve of E-Mode p-Si MOS Capacitor after Continuous Voltage Stress at -30V