LOUGHBOROUGH
UNIVERSITY OF TECHNOLOGY
LIBRARY

AUTHOR/FILING TITLE
DONG, L

ACCESSION/COPY NO.
040101544

VOL. NO. CLASS MARK

LOAN COPY

17 MAR 1995
21 MAR 1997
13 JUN 1997

0401015440

Gate Oxide Failure in MOS Devices

by

Liqin Dong B.Sc

A Master Thesis

Submitted in partial fulfilment of the requirements for the award of the Degree of Master of Philosophy of Loughborough University of Technology

August 1994

© by L. Dong, 1994.

Loughborough University
of Technology Library
Date Jul 95
Class
Acc
No 040101544
V8913473

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to my supervisor, Dr. Vincent Dwyer for providing me the opportunity of this work, for his encouragement and support throughout the work, and for his help and time he put into the final preparation and production of this thesis.

All members in the International Electronics Reliability Institute would have helped me in some ways, their friendship is appreciated. In particular, I would like to thank Dr. Martin Tunnicliffe for the many rewarding discussions I have had with him, for the help on experiments and reading this thesis. Thanks also go to Mr. Paul Donaldson for all his advice and help in experiments and in English. Thanks to Jonathan and Susan for their helps in many ways.

I am greatly indebted my husband, Zaiqing, for his love, encourage and tremendous support throughout many years.

To Zaiqing, I dedicate this work.

ABSTRACT

The thesis presents an experimental and theoretical investigation of gate oxide breakdown in MOS networks, with a particular emphasis on constant voltage overstress failure. It begins with a literature search on gate oxide failure mechanisms, particularly time-dependent dielectric breakdown, in MOS devices.

The experimental procedure is then reported for the study of gate oxide breakdown under constant voltage stress. The experiments were carried out on MOSFETs and MOS capacitor structures, recording the characteristics of the devices before and after the stress. The effects of gate oxide breakdown in one of the transistors in an nMOS inverter were investigated and several parameters were found to have changed.

A mathematical model for oxide breakdown, based on physical mechanisms, is proposed. Both electron and hole trapping occurred during the constant voltage stress. Breakdown appears to take place when the trapped hole density reach a critical value.

PSPICE simulations were performed for the MOSFETs, nMOS inverter and CMOS logic circuits. Two models of MOSFET with gate oxide short were validated. A good agreement between experiments and simulations was achieved.

SYMBOLS AND NOMENCLATURE

A	Ampere.
Å	Angstrom (10^{-8}cm).
a.c.	Alternating current.
AF	Accelerate factor.
Al	Aluminium.
B	Bulk of MOS device.
C	Capacitance, Coulomb.
cm	Centimetre.
CMOS	Complementary metal oxide semiconductor.
Cr	Chromium.
CV	Capacitance vs. voltage.
CT1	Tektronix current transformer probe.
°C	Centigrade.
CAD	Computer aided design.
C_{ox}	Oxide capacitance.
CR	Capacitance resistance.
D	Drain electrode of MOSFET.
DC, d.c.	Direct current.
D-mode	Depletion mode.
DUT	Device under test.
E	Energy.
E_a	Activation energy.
E_{bd}	Energy required to support oxide breakdown.
E_c	Energy of conduction band edge.
E_f	Fermi level.
E_G	Energy of band gap.
E_i	Intrinsic Fermi level.
E_I	Ionization energy.
E-mode	Enhancement mode.
EOS	Electrical overstress.
EPROM	Erasable programmable read only memory.
ESD	Electrostatic discharge.
E_T	Energy associated with defect.
E_v	Energy of valance band edge.
eV	Electron-volt.
F	Electric field, farad.
f	Frequency.
F_{an}, F_{cat}	Anode and cathode electric fields.
F_{bd}	Field threshold for dielectric breakdown.
FET	Field effect transistor.
F-N	Fowler-Nordheim.
G	Giga (10^9), gate of MOS device.
G, g	Trap generation rate.

g_m	Transconductance (or 'mutual conductance').
GOS	Gate oxide short.
H	Hydrogen.
h	Planck's constant.
$\hbar$	Reduced Planck's constant ($h/2\pi$).
HCL	Hydrochloric Acid.
HPIB	Hewlett-Packard interface bus.
Hz	Hertz.
I	Current.
$\Im$	Imaginary part of complex number.
I_0, I_1	Modified Bessel functions.
IC	Integrated circuit.
IGFET	Insulated gate field effect transistor.
IR	Impact recombination.
I/V	Current vs. voltage.
J	Current density.
k	Kilo (10^3).
L	MOSFET channel length, correlation length for fluctuation.
L-C-R	Inductance-capacitance-resistance.
LO	Longitudinal optic (phonon mode).
LSI	Large scale integration.
M	Mega (10^6).
m	Metre, Milli (10^{-3}).
MESFET	Metal semiconductor field effect transistor.
MOS	Metal oxide semiconductor.
MOS-C	Metal oxide semiconductor capacitance.
MOSFET	Metal oxide semiconductor field effect transistor.
MSI	Medium scale integration.
m_0	Electron rest mass.
m^*	Effective electron mass.
n	Electron density, Nano (10^{-9}).
Na	Sodium.
n-type	Semiconductor with majority electrons.
NMOS	Technology based on n-channel MOSFET.
NXOR	Not exclusive-OR logic.
O	Oxygen.
P	Phosphorus.
p	Pico (10^{-12}).
PMOS	Technology based on p-channel MOSFET.
PSG	Phosphosilicate glass.
p-type	Semiconductor with majority holes.
Q	Charge.
q	Electron charge.
Q_{bd}	Charge to breakdown.
Q_d	Total charge of depletion layer.
Q_n	Electron charge of inversion layer.
Q_{ox}	Effective oxide space-charge density.
Q_{ss}	Surface-state charge density.

R	Resistance.
$\Re$	Real part of complex number.
R_b	Resistance of bulk silicon.
RAM	Random access memory.
S	Source of MOSFET.
SEM	Scanning electron microscope.
Si	Silicon.
SiO_2	Silicon dioxide.
SMU	Source/measure unit.
SPICE	Simulation program with integrated circuits emphasis.
SSI	Small scale integration.
T	Temperature.
t	Time.
t_{bd}	Time to breakdown.
t_c	Charge time.
t_d	Discharge time.
TDDB	Time-dependent dielectric breakdown.
T_{ox}	Oxide thickness.
ULSI	Ultra large scale integration.
V	Voltage, Volt.
V_{dd}	Drain supply voltage.
VLSI	Very large scale integration.
V_{ss}	Source supply voltage.
V_T	Threshold voltage of MOS device.
W	MOSFET channel width.
W_d	Depletion layer width.
x	Position coordinate.
XOR	Exclusive-OR logic.
Z	MOSFET channel width.
Z_d	Impendance.
α	Hole generation efficiency, temperature coefficient of resistance.
β	Field acceleration factor.
Δ	R.m.s. height of the fluctuation in the interface.
ϵ_0	Permittivity of free space.
ϵ_{ox}	Dielectric constant of SiO_2.
ϵ_s	Dielectric constant of Si.
μ	Carrier mobility, micro (10^{-6}).
σ	Trap capture cross section, standard deviation.
ϕ	Electric flux.
Φ_{ms}, ϕ_{ms}	Effective work-function difference between gate and substrate.
ϕ_F	Fermi potential in the bulk Si.
Ξ	Deformation potential.
ρ	Crystal density.
τ	Function time, relaxation time.
χ_{ms}	Effective electron affinity.
Ω	Ohm.

CONTENTS

Page No.

ACKNOWLEDGEMENTS (i)

ABSTRACT (ii)

SYMBOLS AND NOMENCLATURE (iii)

CHAPTER 1: INTRODUCTION 1

1.1 Historical Background 1
1.2 Reliability and Quality Assurance 2
1.2.1 Bathtub Curve 3
1.2.2 Screening 4
1.2.3 Accelerated Testing 4
1.3 Instabilities and Failure Mechanisms in MOS Devices 5
1.4 Study Synopsis 6
1.5 References 7

CHAPTER 2: GATE OXIDE FAILURE MECHANISMS 8

2.1 Introduction 8
2.2 Physics of Silicon Dioxide 8
2.2.1 Properties of SiO_2 8
2.2.2 Atomic Structure 9
2.2.3 Band Structure 11
2.2.4 Defects in SiO_2 11
2.2.5 The Si-SiO_2 Interface 11
2.2.6 The Gate-SiO_2 Interface 13
2.3 Dielectric Breakdown 14
2.3.1 Phenomenology of Dielectric Breakdown 14
2.3.2 Time Dependent Dielectric Breakdown 16
2.3.2.1 The Acceleration of TDDB 17
2.4 Charge Trapping during Oxide Wearout 18
2.5 Effects of Oxide Breakdown and GOS upon Circuits 19
2.6 Summary 19
2.7 References 19

CHAPTER 3: EXPERIMENTAL APPARATUS AND PROCEDURE 24

3.1 Experimental Apparatus 24
3.1.1 Wafer Chuck and Microprober System 24
3.1.2 Electrical Instruments 25
3.1.2.1 HP4145B Parametric Analyzer 25
3.1.2.2 Programmable Voltage Source 26
3.1.2.3 L-C-R Meter 26
3.1.2.4 HP54111D Digital Oscilloscope 26
3.1.2.5 TG102 Signal Generator 27
3.2 Test Samples 27
3.2.1 Devices on the Wafers 27
3.2.2 Characterization of Samples 29
3.2.2.1 C-V Curve Measurement 29
3.2.2.2 Extraction of Oxide Thickness 29
3.2.2.3 The Measurement of Bulk Resistance 31
3.3 Experiments on MOS-C Structures and MOSFET Networks 34
3.3.1 Introduction 34
3.3.2 Field Accelerated TDDB Life Test 34
3.3.3 Temperature Accelerated TDDB Life Test 35
3.3.4 Experimental Results 35
3.3.4.1 Time-dependent Failure Distributions 35
3.3.4.2 Injection Current 37
3.4 Experiments on nMOS Inverter 39
3.4.1 *Characterisation of the Pull-down Transistor* 39
3.4.2 Characterisation of the nMOS Inverter 40
3.4.3 Constant Voltage Stress on Gate Oxide of the Pull-down Transistors 41
3.4.4 Re-characterisation of the Pull-down Transistor 41
3.4.5 Re-measurement of the Characteristics of the Inverter 43
3.5 Experimental Conclusions 43
3.6 References 43

CHAPTER 4: ANALYSIS OF THE EXPERIMENTAL RESULTS 44

4.1 Introduction 44
4.2 Modelling of the Oxide Breakdown 44
4.2.1 Introduction 44
4.2.2 The Mechanisms of Oxide Wearout 45
4.2.2.1 The Entry of Charges into Oxide 45
4.2.2.2 Charge Trapping in Oxides 45
4.2.2.3 Electric Field Change due to Trapped Charges in Oxides 48

4.2.2.4 The Breakdown of Oxide 49
4.2.3 Mathematical Model 49
4.2.3.1 Physical Mechanisms 49
4.2.3.2 Quantitative Model 51
4.2.3.3 Calculated Results 53
4.3 The Assessment of the MOS-C as An Oxide-integrity Test Vehicle 54
4.4 The Distributions of the Oxide Failure 57
4.5 Degradation of Transconductivity of MOSFET 59
4.5.1 The Important Scattering Equations 60
4.5.2 Scattering Rates and Mobility 61
4.6 References 62

CHAPTER 5: PSPICE SIMULATION 64

5.1 Introduction 64
5.2 MOSFET Theory 64
5.2.1 Introduction 64
5.2.2 Basic Operation of MOSFET 64
5.2.3 A Mathematical Model of MOSFET 66
5.2.3.1 Threshold Voltage V_T 66
5.2.3.2 Drain Current 67
5.2.3.3 Transconductance and Conductance in MOSFET 69
5.3 MOSFET Models in PSPICE 71
5.3.1 Introduction to PSPICE Simulator 71
5.3.2 MOSFET Models 72
5.4 Simulation Results for MOSFETs with GOS 75
5.4.1 Models for MOSFETs with GOS 75
5.4.2 Simulation Results for Transistor with GOS 78
5.5 Simulation Results of the nMOS Inverter with GOS 81
5.5.1 Introduction to Inverter 81
5.5.2 Delay and Power 81
5.5.3 Simulation Results of nMOS Inverter 84
5.5.4 Conclusion 85
5.6 Test Considerations of MOS circuits with GOS 86
5.6.1 CMOS Inverter 86
5.6.2 NMOS Inverter 87
5.6.3 Other CMOS Logic Circuits 88
5.6.4 Conlusion 89
5.7 References 89

CHAPTER 6: CONCLUSIONS 91

APPENDIX: REPRODUCTIONS OF PUBLISHED PAPERS 93

Paper I (Proc. 5th ESREF, 1994) 94
Paper II (Proc. 5th ESREF, 1994) 99
Paper III (Journal of Electroceramics, 1993) 105

CHAPTER 1

Introduction

1.1. Historical Background

The evolution of electronic technology has been so rapid that it is sometimes described as a revolution. It is not an exaggeration to say that most of the technological achievements of the past two decades have depended on microelectronics. Small and reliable sensing and control devices are essential elements in complex systems from spacecraft, communications satellites to handheld calculators and digital watches. Somewhat subtler, but perhaps eventually more significant, is the effect of microelectronics development upon the computer.

The earliest transistor design was patented in the United States and Canada by J.E.Lillenfeld in 1925 [1]. According to Lillenfeld's proposal, the current flowing in a copper sulphide channel between gold 'drain' and 'source' electrodes was controlled by a voltage applied to an aluminium 'gate' electrode. In 1928, Lillenfeld patented the earliest insulated gate field effect transistor (IGFET) design, in which the conductivity of a copper sulphide channel was controlled by the voltage on an isolated gate, separated from the channel by a layer of aluminium oxide [1,2]. The IGFET was also patented by Heil in Great Britain in 1935 [3]. These designs were the forerunners of the modern MESFET and MOSFET transistors. It is extremely doubtful, however, that any workable transistors were constructed at this time.

The first known operational transistor, a germanium bipolar device, was developed in 1947 by Shockley, Bardeen and Brattain [4,5]. By the late 1950's, small-scale bipolar integrated circuits (SSI) were manufactured, but these were limited by high power dissipation and complicated fabrication processes.

To overcome these problems, the manufacture of the conceptually simpler metal-oxide-semiconductor (MOS) device was perfected in the mid 1960's, and it is this technology that has permitted the production of today's highly complex microprocessor

and large capacity memory circuits. From the evolution of circuit complexity shown in Fig.1.1, it can be seen that to a first approximation the number of components per integrated circuit has been increasing exponentially - a phenomenon usually referred to as Moore's Law. Small-scale integration (SSI) was the state of art in the early 1960's when the circuits were built using bipolar technology. The introduction of MOS technology allowed circuits of medium-scale integration (MSI) to be produced in the mid to late 1960's and the large-scale integration (LSI) of circuits in the 1970's.

Very large scale integration (VLSI) followed in the 1980's, with the fabrication of 32 bit microprocessors and 256K RAMs. Such scales of integration can approach one million devices per chip [6].

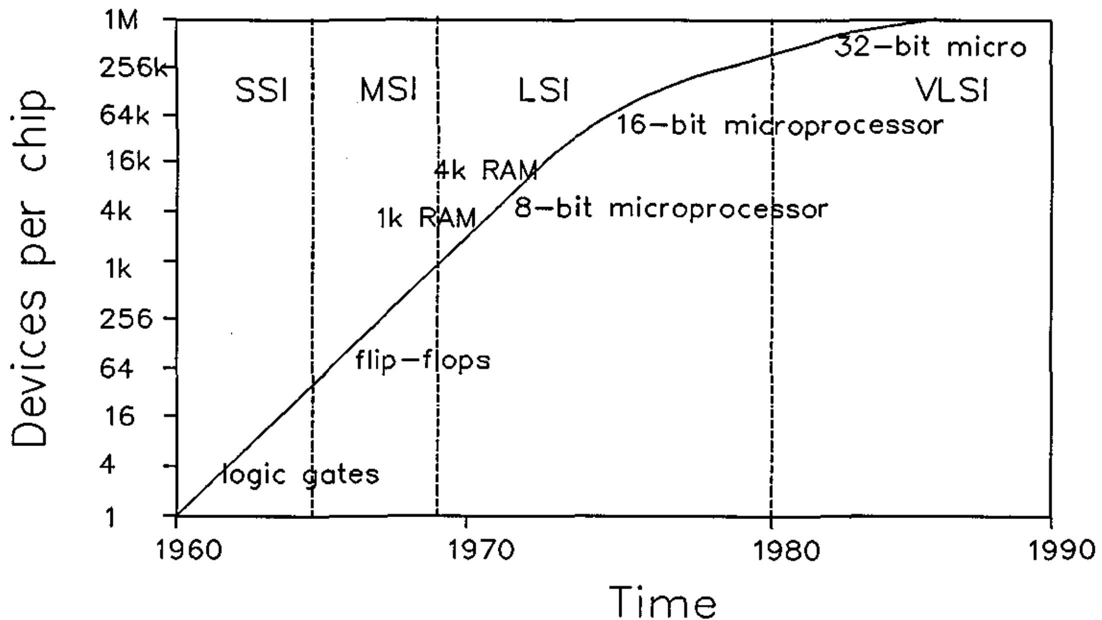


Figure 1.1: Integrated circuit complexity as a function of time (Moore's law).

1.2 Reliability and Quality Assurance

This thesis is concerned with reliability and failure mechanisms in small-dimension

MOS devices. This section presents a general discussion of semiconductor reliability theory. The specific failure mechanisms associated with MOS devices are considered in Section 1.3.

1.2.1 Bathtub Curve

A well-known graphical reliability model, which has been universally accepted since the beginning of the reliability discipline is the *'bathtub' curve* [8] shown in Fig.1.2. This curve has three distinct regions, each with its own unique characteristics. The first region with a rapidly decreasing failure rate from a high initial value represents 'infant mortality'. The second region, with almost a constant failure rate, extends well beyond the useful life of most components. The third region, marked by a steep rise in failure rate, is the final 'wearout' period.

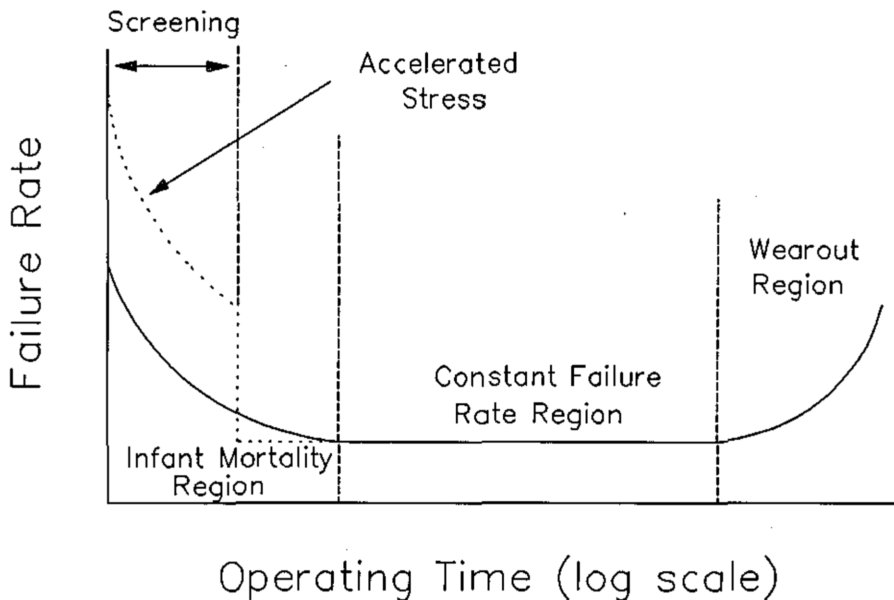


Figure 1.2: The bathtub curve.

The initial high failure rates in the infant mortality region arise from the fact that a small fraction of the population is inherently defective (weak) and fails in a very short

period of time. Most of the weak devices fail during this period. The region of constant (or nearly constant) failure-rate corresponds to normal operation of electronic equipment. Eventually, the ICs begin to fail at a rapid rate as a result of intrinsic failure mechanisms such as dielectric breakdown or electromigration.

1.2.2 Screening

In principle, *screening* is a process of eliminating defective components from the production batch. Reliability screening requires acceleration of the mechanisms that give rise to infant mortality. The aim of the screening process is to accelerate the failure rate for a limited period of time, such that the surviving population would begin its effective operational life within the middle steady region of the bathtub curve as shown in Fig.1.2.

Temperature, bias or a combination of the two are often used to accelerate failures within the screening period. The most commonly used screening procedure is burn-in, which combines the effects of voltage and temperature. The ICs with weak oxides, defective metallization, and contamination are typically eliminated by this screen.

1.2.3 Accelerated Testing

Failure rates in today's VLSI circuits tend to be so low in normal operation that it takes many years for the steep rise in failure rates due to wearout to begin. However, systems manufacturers require an assurance that the component failure rates remain low and steady over the intended operating life of the equipment. Therefore, it is an industry-wide practice to test ICs under conditions that accelerate wearout in order to provide statistically sufficient data and meaningful estimates of failure rates in an economically reasonable period of time. From the test results, acceleration factors (AF) and activation energies (E_a) are derived which permit the estimation of failure rates under any set of conditions of operation in the field.

Temperature, voltage, humidity, mechanical shocks and vibrations and radiation are amongst the types of stresses used to accelerate failures. Since the accelerated testing of IC's tends to be destructive, tests have to be performed on a small representative sample size drawn from the total population.

Interpretation of the test results requires an understanding of the accelerating effects of the applied stresses. Since the main purpose of these tests is to quantify all the possible known failure mechanisms under accelerated-stress conditions, an accurate model is necessary in order to extrapolate these mechanisms to normal operating conditions.

1.3 Instabilities and Failure Mechanisms in MOS Devices

Although in 1960, MOS devices with good characteristics were demonstrated, they did not become commercially reliable until the end of the decade. This was because the early devices were plagued with instability problems arising from **charges in the oxide or at the semiconductor interface**, e.g. *ionic contamination, dipole polarization* and *carrier injection into the oxides.*

Oxides can become contaminated by processing and handling. The environment, materials, machines, and operators are the main sources of contamination. In MOSFETs, the effect of ionic contamination is manifested as a drift in the strong-inversion threshold voltage, although this is usually negligible in the state-of-the-art technology.

The phosphosilicate glass (PSG) that immobilizes sodium ions can have a destabilizing effect on active devices because of its polarizability. These dipoles create image charges that can invert the silicon surface and produce effects on devices similar to those by mobile ions, however, the effects of dipole polarization are much less pronounced than those of mobile ions [8].

In order to meet the demands of increased packing density, the feature sizes are continuously being scaled down. With decreasing MOSFET dimensions, the ICs are becoming increasingly susceptible to drift as a result of the injection of carriers. Electrons and holes can enter the oxide by either tunnelling through or by surmounting the barrier. These injected carriers, when trapped in the oxide, are capable of changing the I-V characteristics of MOSFET. Meanwhile, latchup is a key concern to CMOS circuits. It stems from parasitic bipolar transistors, which are structurally inherent to bulk CMOS [9]. As CMOS technologies are scaled down, the protection against latchup will become more difficult.

Although metal electromigration, packaging and interconnections are among the

common failure mechanisms in MOS circuits, oxide breakdown is a great threat to MOS VLSI circuits and has been a subject of research since the mid 1920's [10]. Numerous theories have been proposed, however, no consensus has ever been reached concerning the physical mechanisms involved [11].

The effects of oxide breakdown upon circuits are also concerned by many researchers. Latent failure is a great threaten to reliability, moreover, 'gate oxide shorts (GOS)', i.e. short circuit paths across the insulator [12] can cause distortion of the transistor I/V characteristics and/or loss of transistor action [13].

1.4 Study Synopsis

The aims of this study are twofold: (i) develop a physical model to explain the experimental data and understand the oxide breakdown mechanism; (ii) examine the effects of Gate Oxide Shorts upon devices and circuits and try to find a way to detect them.

The remainder of thesis is divided into the following five chapters.

Chapter 2 outlines the properties of SiO_2, introduces oxide breakdown mechanism with emphasis on time-dependent dielectric breakdown.

Chapter 3 describes the apparatus and test samples used for the experimental work, together with the experimental procedures and results.

Chapter 4 develops a mathematical model based on the charge trapping mechanism in order to understand the physical mechanisms behind the experimental results, analyzes the main results from Chapter 3.

Chapter 5 focuses on the changes in the device characteristics and circuits after gate oxide breakdown, compares two different GOS equivalent circuits using PSPICE simulator and validates techniques for the detection of GOS in MOS digital circuits.

Chapter 6 draws conclusions from the results of the thesis.

1.5 References

[1] Gosling,W, Townsend,W.G., Watson,J, "Field-Effect Electronics", Butterworth, London, 1971.

[2] Lillenfeld,J.E., U.S.Patent No.1745175, 1930.

[3] Heil,O, British Patent No.439137, 1935.

[4] Bardeen,J, Brattain,W.H., "The Transistor, A Semiconducting Triode", Physical Review, **74**, p.230, 1948.

[5] Shockley,W, "The Theory of p-n Junctions in Semiconductors and p-n Transistors", Physical Review, **80**, p.467, 1950.

[6] Russell,G, Kinniment,D.J., Chester,E.G., McLauchlan,M.R., "Chapter 2: Technology and Design", CAD for VLSI, Van Nostrand Reinhold (UK) Co. Ltd, 1985.

[7] Forester,T, "The Microelectronics Revolution", Basil Blackwell, Oxford, 1980.

[8] Sabnis, A.G., "VLSI Reliability", Academic Press, Inc., 1990.

[9] Troutman,R.R., "Latchup in CMOS Technology: The Problem and Its Cure", Kluwer Academic Publishers, 1986.

[10] Amerasekera,E.A., "Failure Mechanisms in MOS Devices", Ph.D Thesis, 1986.

[11] Tunnicliffe,M.J., Dwyer,V.M., "Monitoring the Integrity of MOS Gate Oxides: A Preliminary Study", International Journal of Electronics, 1993.

[12] Soden,J.M., Hawkins,C.F., "Test Considerations for Gate Oxide Shorts in CMOS ICs", IEEE Design and Test, August, pp.56-64, 1986.

[13] Tunnicliffe,M.J., Dwyer,V.M., Campbell,D.S., "Parametric Drift in Electrostatically Damaged MOS Transistors", Proc. 14th Electrical Overstress/Electrostatic Discharge Symposium, Dallas, Texas, pp.112-120, 1992.

CHAPTER 2

MOS Gate Oxide Failure Mechanisms

2.1 Introduction

Silicon dioxide plays vital roles in MOS integrated circuits: firstly as a 'passivating' layer and secondly as the MOSFET gate dielectric. The quality of the gate oxide is a crucial factor in determining the performance of an MOS transistor. MOS ICs depend upon the dielectric integrity of thin oxide layers for the high input impedance of MOSFETs and the charge storage ability of MOS capacitors.

Oxide breakdown is probably the most widely recognized failure mechanism in MOS VLSI circuits and has indeed become a serious reliability problem in electronic equipment. The breakdown strength of gate oxide layers usually varies between 7 and 14MV/cm [1,2] depending on stress conditions and method of preparation. An understanding of the silicon dioxide breakdown behaviour is extremely important for prediction and estimation of IC failure rates in field operation.

Charge generation and trapping in SiO_2 has long been suspected to be responsible for the time-dependent dielectric breakdown of MOS structures. Degradation of MOSFETs by trapped electrons and/or holes in the SiO_2 films is the most serious technical problem in assurance of reliability in Si VLSI circuits. It is therefore extremely important to investigate the charge trapping and trap generation mechanisms.

2.2 Physics of Silicon Dioxide

2.2.1 Properties of Silicon Dioxide

SiO_2 or 'silica' is the native oxide of crystalline silicon. It grows rapidly on clean

silicon surfaces to a depth of about 2nm on exposure to room temperature air. Since such thin layers have little practical use, thicker oxides are grown at about 1000°C in oxidation furnaces over periods of time between 15 minutes to 2 hours [3]. The oxidising ambient can be pure O_2 for 'dry' oxides or steam for 'wet' oxides, the wet oxide having a considerably faster growth rate at the expense of electrical stability. Since electrical stability of the passivation is not normally of paramount importance, and since it is usually thick (of the order of 1μm), 'wet' oxides are generally used for this application [3].

Although thermally grown SiO_2 is the best known material for a gate insulator, its properties are far from perfect. 'Dry' oxidation is usually used, as the gate oxide integrity and stability is of importance to the circuit performance and reliability [3]. Strategies have been developed to improve the electrical properties of gate oxides, including the addition of HCL in the oxidising ambient [4] and the deposition of an additional yttrium oxide or silicon nitride dielectric layer [5,6]. The techniques used to grow (or deposit) such layers on silicon substrates have been extensively reviewed by S. Rigo in ref.[7].

2.2.2 Atomic Structure

Although thermally grown SiO_2 is chemically identical to crystalline quartz (see Fig.2.1(a)), its atomic structure is vitreous, containing a high degree of short-range order and chemical homogeneity but without any long-range organization (Fig.2.1(b)).

In the classical model of vitreous silica, the SiO_4 tetrahedra are linked via bridging O atoms to form a continuous network of Si-O-Si bonds. Bond angles and interatomic distances are shown in Fig.2.2.

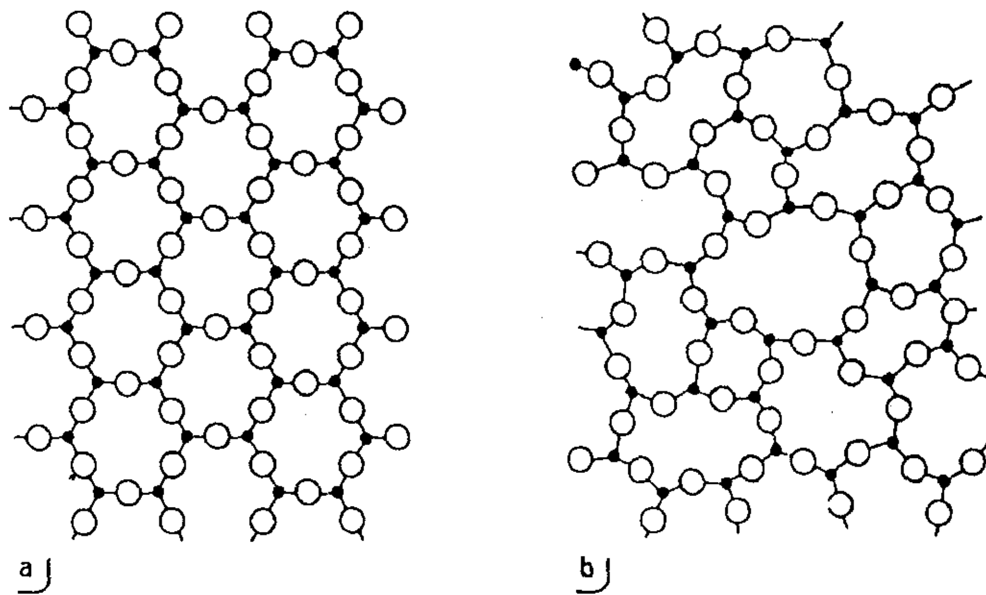


Figure 2.1: Two-dimensional representation of:
a) the lattice of 'perfect' quartz, b) the network of 'perfect' vitreous silica (black dots represent Si atoms, circles represent O atoms) [7].

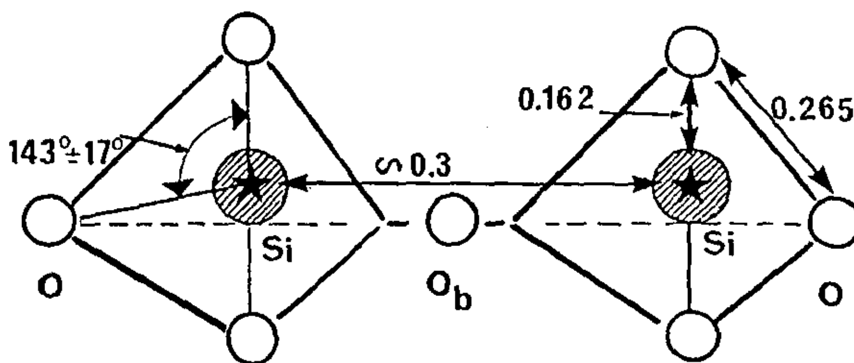


Figure 2.2: Typical dimension of the SiO_4 tetrahedra of vitreous silica. Dimensions are in nm, angles in degree [7].

2.2.3 The Band Structure of Vitreous Silica

Despite the lack of long-range order, vitreous silica has sufficient short-range regularity to diffract the electron wave functions into a valance and a conduction band. Optical absorption measurements performed on vitreous silica show that a band gap exists, which depends on the degree of perfection of the oxide network: the band gap is found to decrease when disorder increases. The most commonly measured E_G value for vitreous silica is 8.1eV, which agrees well with computed values based on one-dimensional theoretical models [7].

2.2.4 The Defects in Silica

Defects in silica can be classified into three categories: *point defects* which exist at single atomic sites, *complex defects* which are clusters of point defects, and *microheterogenities* which are devitrified regions in an otherwise non-crystalline oxide [7].

The presence of defects leads to a loss of the translational symmetry of the network and thus introduces supplementary energy levels. These levels may be located in the conduction band, and/or in the valence band, and/or in the band gap. Only those levels introduced in the band gap are of interest for instability phenomena [7].

A fundamental level E_T associated with a defect can be computed theoretically [7], and the ionization energy of the defect E_I (defined as $E_I = E_C - E_T$) can be measured by IR absorption or by thermally-stimulated conductivity measurements for shallow levels [7].

2.2.5 The Si-SiO_2 Defects

A cleaved Si crystal has a high density of surface trap states introduced by incomplete or 'dangling' orbitals. These behave as interband surface states or traps and have the same areal density as the surface atoms (i.e. $\sim 10^{15}$ traps/cm^2). Although a thermally grown SiO_2 layer passivates the surface, completing most of the surface orbitals, the periodic mismatch between the Si and SiO_2 networks implies that some Si bonds remain incomplete as shown in Fig.2.3.

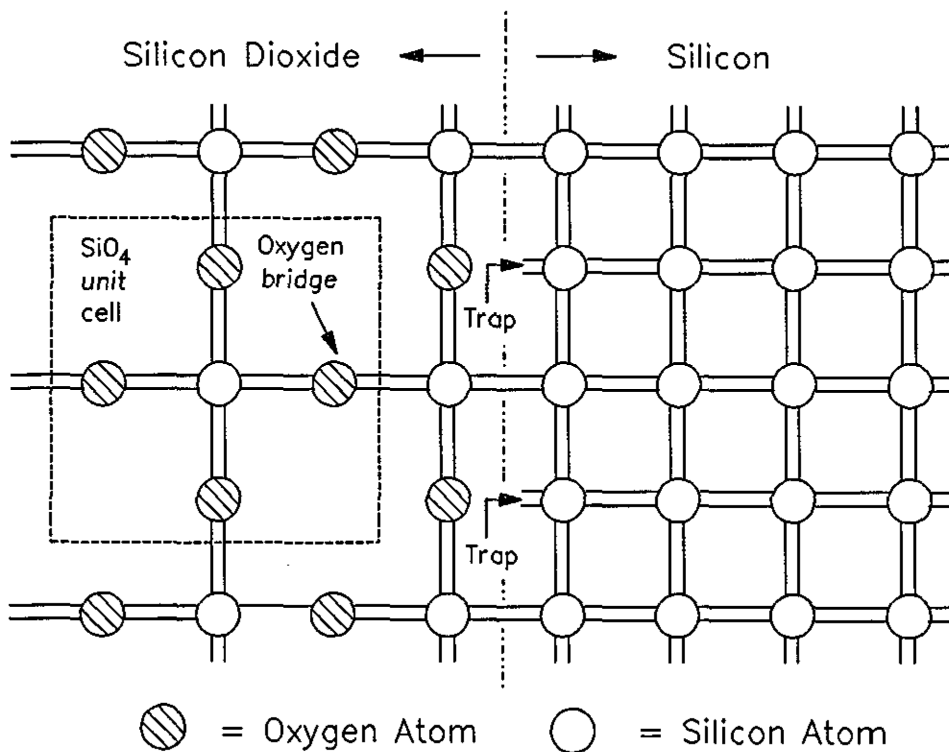


Figure 2.3: Schematic two-dimensional representation of Si-SiO_2 interface.

Stretched, twisted and dangling bonds do indeed appear on passivated Si-SiO_2 interfaces, introducing an 'interface' trap density of the order of 10^{12} traps/cm^2 [8]. Extrinsic defects associated with ionic contaminants can also exist [9] and although both donor and acceptor species exist at the interface, the donors are more numerous and the net surface charge is positive [10]. Hence the extrinsic surface state charge can be modelled by a positive charge density Q_{ss} (C/cm^2).

Interface state density is affected by the type and density of Si dopants, boron giving a higher trap density than phosphorus or arsenic due to an introduced oxygen deficiency [10]. Interface states can sometimes be annealed away by post-oxidation low-temperature thermal processing [10], although this can sometimes increase rather than decrease the defect density [11,12]. This is in some cases due to void formation under the deoxidation reaction ($Si + SiO_2 \rightarrow 2SiO\uparrow$) across the Si-$SiO_2$ interface. It can be prevented by

increasing the O_2 content of the annealing ambient, thereby causing compensatory reoxidisation [13]. Interface traps are also generated under electric field stress, although these may have slightly different properties from those of the pre-existing defects [14].

The charged interface states, together with unoxidised Si^+ ions, present a charge storage capacity at the interface, which can be detected using charge-pumping analysis [15], a.c. conductance measurements [16] or deep-level transient spectroscopy [17]. Carriers in MOSFET channel could enter the trap sites in oxide through a tunnelling mechanism, causing 1/f fluctuations in the channel conductance, and therefore the density of interface traps can be measured using a 1/f noise model [18].

2.2.6 The Gate-SiO_2 Interface

The most common materials used for gate fabrication are aluminium and polysilicon (polycrystalline silicon). Both materials are applied to the SiO_2 by vapour deposition, and exhibit a polycrystalline structure. The polysilicon/SiO_2 junction can be considered to display similar properties to the thermal Si/SiO_2 interface. The energy barrier between the polysilicon gate and the SiO_2 can be characterised in terms of the effective work function ϕ_{ms}, i.e. the potential difference between the silicon Fermi-level and the oxide conduction band. Alternatively the effective electron affinity χ_{ms}, i.e. the potential difference between the oxide and silicon conduction band edges can be used. Both parameters, shown schematically in Fig.2.4, can be measured using the internal photoemission technique [10,19]. The Si-SiO_2 interface appears to have a ϕ_{ms} of approximately 3.15eV, a value which varies slightly with the Si crystal orientation [10]. If the gate material is metallic, the effective electron affinity is meaningless since the conduction band edge bears little relation to the energy of cathode electrons. The effective work function is therefore the only meaningful parameter, equalling 4.1eV for an aluminium gate material.

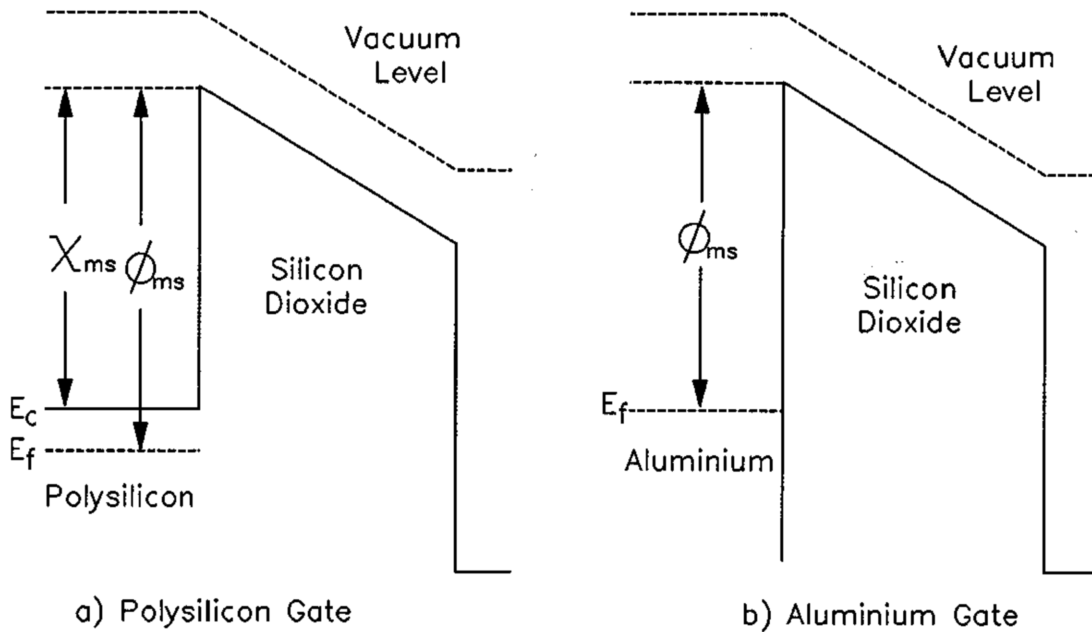


Figure 2.4: Energy band diagrams for poly-Si/SiO_2 and Al/SiO_2 MOS structures.

2.3 Dielectric Breakdown

2.3.1 Phenomenology of Dielectric Breakdown

The breakdown of thin silica films is usually studied through electrical tests performed on capacitor-like structures. At breakdown, a large localised current flows in the dielectric, melting a region of gate material and producing a visible crater (see Fig.2.5). The breakdown site can be located by electron microscopy [20,21] or by electroluminescence under current injection [22]. If the gate electrode is thin then the heat dissipated by breakdown may evaporate the gate material in the vicinity of the crater, isolating the defect and *self healing* the oxide [20]. However, either the molten gate material or an air discharge may form a conduction path between the gate and the underlying silicon, sustaining the breakdown current and causing damage propagation across the gate area [20]. Alternatively the current path across the oxide may fuse under

further stress, providing another mechanism for self healing [23].

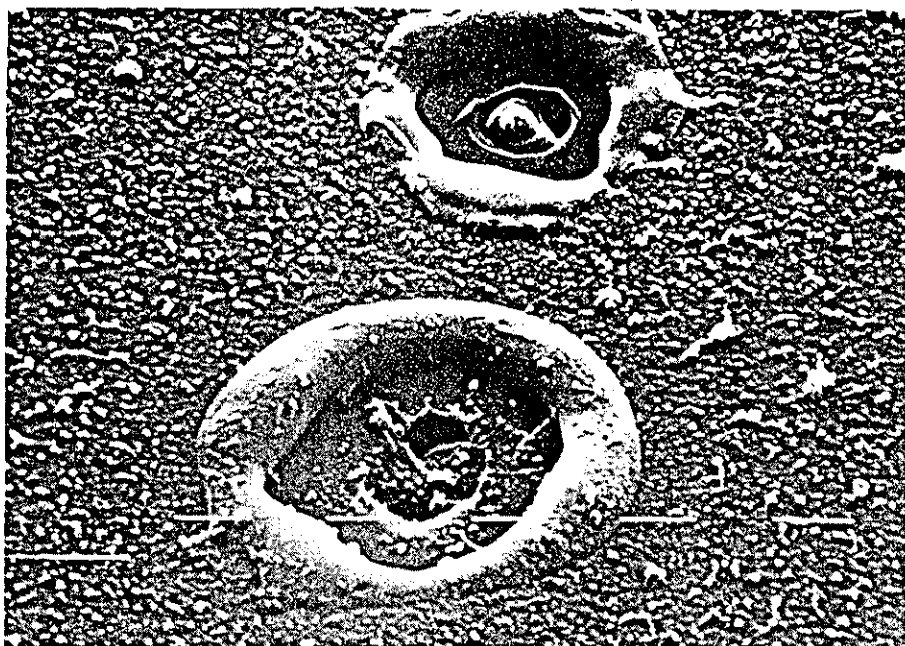


Figure 2.5: A SEM photograph of two breakdown spots. The polysilicon electrode has melted away over $\sim 20\mu m^2$. The polysilicon is spread over a large area in the form of dust and little globules. The size of a dash is $1\mu m$.

Although the definitions of what constitutes a damaged oxide varies in the literature, a typical example is a leakage current of $10\mu A$ at 5V bias [24]. However, the rapid resistivity change at breakdown (typically from $1000G\Omega$ to $1K\Omega$ in less than $1\mu s$) appears to render the exact definition somewhat unimportant. Some recent experiments on ultra-thin SiO_2 (about 55Å) show that nondestructive multiple breakdown events can occur prior to final breakdown as the oxide switches between these high and low conductance levels [25]. Later studies have uncovered the existence of multiple conduction levels in the oxide, between which the dielectric may switch prior to breakdown [26,27].

2.3.2 Time Dependent Dielectric Breakdown (TDDB)

The present consensus maintains that SiO_2 wearout is driven by Fowler-Nordheim tunnelling injection of electrons from the cathode into the dielectric. Such wearout tends to be characterised either by the time-to-breakdown t_{bd} or by the charge Q_{bd} injected prior to breakdown. Fig.2.6 illustrates the causal nature of wearout. It shows the Fowler-Nordheim current J plotted against the injected charge Q for two oxides, one stressed at a continuous 12MV/cm, the other stressed first at 12MV/cm and then at 11MV/cm after an indefinite interruption. Clearly only a small additional charge is required to push the oxide into breakdown at 11MV/cm after the injection it has already sustained. This is called time-dependent dielectric breakdown (TDDB).

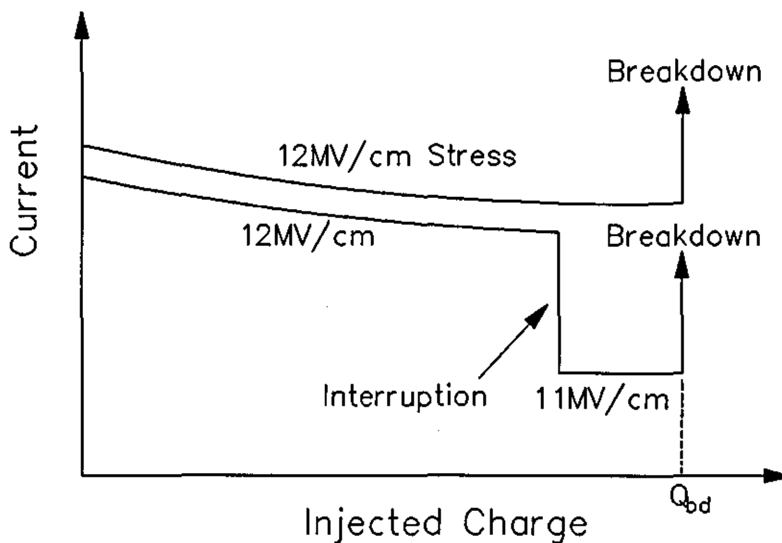


Figure 2.6: Illustration of time-dependent dielectric wearout (after Wolters et al. [28]).

Time-dependent dielectric breakdown (TDDB) is judged to be a potential failure mode for MOS integrated circuits and has received considerable attention [29-37]. Early investigations centred around the effects of mobile ions on enhanced conduction via ionic drift to the Si-SiO_2 interface [31,32], or on the effects of radiation damage on dielectric breakdown [33,34]. More recently, with improved processing, oxides have become

relatively free from both Na^+ contamination and radiation damage; thus TDDB attention has turned to intrinsic breakdown [30, 35]. The occurrences of TDDB, experimentally observed on test capacitors and actual memory devices, appear to be lognormally distributed (i.e. with the logarithm of time) [30] or to follow a Weibull distribution [36,37].

2.3.2.1 The Acceleration of Time-dependent Dielectric Breakdown

The occurrences of TDDB can be accelerated thermally and electrically. The temperature dependence has been found to follow the Arrhenius relationship, i.e.

$$t_{bd}(T) \propto \exp\left(\frac{E_{tbd}}{kT}\right) \tag{2.1}$$

where t_{bd} is the time-to-breakdown, E_{tbd} is an activation energies, although a wide range of values has been reported [30,38-41]. The field dependence of TDDB follows the equation

$$t_{bd} \propto \exp\left(\frac{G}{F_{ox}}\right) \tag{2.2}$$

where G is a constant and F_{ox} is the oxide electric field. In order to characterize or project the lifetime of oxides, the concept of 'Electric-field acceleration factor' β is widely used. β is defined as:

$$\beta \equiv \frac{d\log(t_{bd})}{dF_{ox}} \tag{2.3}$$

Using the acceleration factor β, oxide life time under normal (low-field) operation conditions would be extrapolated from the high-field acceleration test data. However,

again, many β values have been reported in the literature, varying from 7 to 1.5 decade/MV/cm [30,42,43].

2.4 Charge Trapping during the Oxide Wearout

Any defect in the SiO_2 which introduces an energy level in the band gap can act as a carrier trap. The current decay evident in Fig.2.6 may indicate electron trapping, making the oxide more negative and reducing the cathode field which supports tunnelling. Current increase has also been observed [31,44,45], possibly caused by mobile positive ions drifting towards the cathode or by hole trapping increasing the cathode field.

Charge trapping in thermally grown SiO_2 on silicon has been studied extensively [45-52]. Both electron and hole trapping have been observed experimentally [45,49-51]. Electron trapping depends on current density (therefore the field) [50], temperature [46,52], thickness of the oxide [46], and the conditions of oxidation process [46,49].

Traps in SiO_2 can be classified by the following three types with respect to their origin: (1) extrinsic trap states related to impurities such as sodium and heavy metals, (2) semi-intrinsic trap states generated by water- or hydrogen-related species, (3) intrinsic trap states induced by Si-Si stretched bonds or oxygen vacancies in SiO_2. The first type of trap states can be neglected in modern oxidation processing, although the second and third types remain important.

Both electron and hole trapping have been observed experimentally [45,49-51]. It has been suggested that the generation of electron traps during high-field stress may occur at either the gate-SiO_2 or the SiO_2-bulk Si interfaces, depending on the polarity of the stress [51], Other workers suggest that the negative charges in oxide are distributed more or less uniformly and that the positive charges are located near the Si-SiO_2 interface [45,49]. Although ultra-dry oxide films have a very small density of electron traps (less than $10^{11}/cm^2$), they have twice as many hole traps as wet oxide films [49]. Electron trapping and hole trapping are strongly dependent on the stressing conditions and a wide range of trap cross sections have been reported.

2.5 Effects of Oxide Breakdown and GOS upon Circuits

Amerasekera et al. [53] experimentally studied the characteristic changes of n-channel MOSFET after ESD pulse. Both parametric degradation and catastrophic failure were observed. Greason et al. [54] showed experimental evidence of latent failure on devices after the gate oxide was subjected to ESD stress, thermal shock or exposed to ultraviolet light, and concluded that latent failure are mainly related to the instability of the silicon-silicon dioxide system in MOS devices and charge trapping appears to be a key factor in the phenomenon. Tunnicliffe et al. [55] found that the threshold voltage V_T and transconductance g_m of the MOSFET drifted before the catastrophic failure, showed the effects of a 'walking-wounded' MOSFET upon the circuitry characteristics of CMOS inverter.

Gate oxide failure and their resulting shorts greatly affect MOS IC reliability. Gate oxide defects usually cannot be detected by stuck-at fault model test sets. In CMOS circuits, this problem can be solved by accompanying the each test vector with an I_{DD} measurement [56].

2.6 Summary

Although numerous theories have been proposed to explain oxide breakdown, no consensus has ever been reached concerning the physical mechanisms involved. New experiments are needed to further study the physical mechanisms of oxide breakdown.

Test for the gate oxide failure in CMOS ICs has been proposed, however, it is not effective universally. Alternative ways are needed.

2.7 References

[1] Scatterfield,H.W., Werber,R.F., "Statistical and Graphical Analysis of Oxide Thickness and ESD Failure Modes", Proc. 9th. EOS/ESD Symp., pp.258-264, 1987.

[2] Osburn,C.M., Ormond,D.W., "Dielectric Breakdown in Silicon Dioxide Films, Part II: Influence

of Processing and Materials", J. Electrochem. Soc., **119(5)**, pp.597-603, 1972.

[3] Davis,J.R., "Instabilities in MOS Devices", Electrocomponent Science Monographs Volume 1, Gordon and Breach, New York, 1981.

[4] Hamasaki,M, "Effect of Adding HCL to the Oxidizing Ambient on Characteristics of Metal-Oxide-Semiconductor Capacitors", J. Appl. Phys., **52(5)**, pp.3484-3490, 1981.

[5] Manchanda,L, Gurvitch,M, "Yttrium Oxide/Silicon Dioxide: A New Dielectric Structure for VLSI/ULSI Circuits", IEEE Electron Dev. Lett., **9(4)**, pp.180-182, 1988.

[6] Moslehi,M.M., Saraswat,K.C., "Thermal Nitridation of Si and SiO_2 for VLSI", IEEE Trans. Electron Dev., **ED32(2)**, pp.106-123, 1985.

[7] Barbottin,G, Vapaille,A, "Instabilities in Silicon Devices", Vol.I, Elsevier Science Publishers B.V., 1986.

[8] Goetzberger,A, Klausmann,E, Schulz,M.J., "Interface States on Semiconductor/Insulator Surfaces", CRC Critical Reviews in Solid State Science, **6**, pp.1-43, 1976.

[9] Rosencher,E, Coppard,R, "Transient Capacitance Spectroscopy of Na^+-Induced Surface States at the Si/SiO_2 Interface", J. Appl. Phys., **55(4)**, pp.971-979, 1984.

[10] Verwey,J.F., Amerasekera,E.A., Bisschop,J, "The Physics of SiO_2 Layers", Rep. Prog. Phys., **53**, pp.1297-1331, 1990.

[11] Reed,M,L., Plummer,J.D., "Si-SiO_2 Interface Trap Production by Low-Temperature Thermal Processing"' Appl. Phys. Lett., **51(7)**, pp.514-516, 1987.

[12] Burte,E.P., Matthies,P, "Time-Resolved Thermal Annealing of Interface Traps in Aluminium Gate - Silicon Oxide - Silicon Devices", IEEE Trans. Nuc. Sci., **NS35(5)**, pp.1113-1120, 1988.

[13] Hofmann,K, Rubloff,G.W., Liehr,M, Young,D.R., "High Temperature Reaction and Defect Chemistry at the Si/SiO_2 Interface", Appl. Surf. Sci., **30**, pp.25-31, 1987.

[14] Vuillaume,D, Goguenheim,D, Bourgoin,J.C., "Nature of the Defects Generated by Electric Field Stress at the Si-SiO_2 Interface", Appl. Phys. Lett., **55(5)**, pp.490-492, 1991.

[15] Elliot,A.B., "The Use of Charge Pumping Currents to Measure Surface State Densities in MOS Transistors", Sol. State Electron, **19**, pp.241-247, 1976.

[16] Noras,J.M., "Parameter Estimation in MOS Conductance Studies", Sol. State Electron, **31(5)**, pp.981-987, 1988.

[17] Bauza,D, Morfouli,P, Pananakakis,G, "Detection of Interface and Volume Traps in Very Thin Oxide MOS Structures using DLTS, Quazi-Static and Conductance Measurements", Sol. State Electron, **43(9)**, pp.933-936, 1991.

[18] Celik-Butler,Z, Hsiang,T.Y., "Determination of Si-SiO_2 Interface Trap Density by 1/f Noise Measurements", IEEE Trans. Electron Dev., **ED35(10)**, pp.1651-1655, 1988.

[19] Williams,R, "Photoemission of Electrons from Silicon into Silicon Dioxide". Phys. Rev., **140(2A)**, pp.A569-575, 1965.

[20] Klein,N, Gafni,H, "The Maximum Dielectric Strength of Thin Silicon Oxide Films", IEEE Trans.

Electron Dev., **ED12(12)**, pp.281-289, 1966.

[21] Klein,N, "Switching and Breakdown in Films", Thin Solid Films, **7**, pp.149-177, 1971.

[22] Hansel,G, Villareal,E, "Cost Effective Failure Analysis Method for Detection Failure Site Associated with Extremely Small Leakage", Proc. **9th**, EOS/ESD Symp., pp.71-77, 1987.

[23] Taylor,R.G., Woodhouse,J, "Junction Degradation and Dielectric Shorting: Two Mechanisms for ESD Recovery", Proc. **8th**. EOS/ESD Symp., pp.92-99, 1986.

[24] Amerasekera,E.A., Campbell,D.S., "Charge Limited Breakdown in MOS Capacitors", Proc. **17th**. Annual Solid State Device Research Conference, pp.733-741, 1987.

[25] Neri,B, Olivo,P, Riccò, "Low-frequency Noise in Silicon-gate Metal-Oxide-Silicon Capacitors Before Oxide Breakdown", Appl. Phys. Lett., **51(25)**, pp.2167-2169, 1987.

[26] Suñè,J, Farrès,E, Placencia,I, Barniol,N, Martín,F, Aymerich,X, "Nondestructive Multiple Breakdown Events in Very Thin SiO_2 Films", Appl. Phys. Lett., **55(2)**, pp.128-130, 1989.

[27] Saletti,R, Neri,B, Olivo,P, Modelli,A, "Correlated Fluctuations and Noise Spectra of Tunnelling and Substrate Currents Before Breakdown in Thin-Oxide MOS Devices", IEEE Trans. Electron Dev., **ED37(11)**, pp.2411-2413, 1990.

[28] Hamano,K, "Breakdown Characteristics in Thin SiO_2 Film", Jap. J. Appl. Phys., **13(7)**, pp.1085-1092, 1974.

[29] Wood,M.H., Euzent,B.L., "Reliability in MOS Integrated Circuits", IEDM Tech. Dig., p.50, 1984.

[30] Crook,D,L, "Method of Determining Reliability Screens for Time Dependent dielectric Breakdown", Proc. Int. Reliability Phys. Symp., pp.1-7, 1979.

[31] Raider,S.I., "Time Dependent Breakdown of SiO_2 Films", Appl. Phys. Lett., **23(15)**, pp.34-36, 1973.

[32] Osburn,C.M., Ormond,D.W., "Na-induced Barrier Height Lowering and Dielectric Breakdown in SiO_2 Films on Silicon", J. Electrochem Soc., **121**, pp.1195-1198, 1974.

[33] Li,S.P., Maserjian,J, "Electron Radiation Effects on Time Dependent Dielectric Breakdown in SiO_2 Films", Solid State Electronics, **18**, pp.287-288, 1975.

[34] Li,S.P., "Ion Implantation Effect on Time Dependent Breakdown in SiO_2", J. Appl. Phys., **46**, pp.4833-4834, 1975.

[35] Anolick,E.S., Nelson,G.R., "Low Field Time Dependent Dielectric Integrity", Proc. **17th** Annual IEEE Rel. Phys. Symp., pp.8-12, 1979.

[36] Vollertsen,R.,P., " Statistical Modelling of Time Dependent Oxide Breakdown Distributions", Microelectronics Reliab., **33(11)**, pp.1665-1677, 1993.

[37] Tunnicliffe,M.J., Dong,L, Dwyer,V.M., "Monitoring the Integrity of MOS Gate Oxides", Int. Electroceramics, 1993.

[38] Baglee,D.A., "Characteristics and Reliability of 100Å Oxides", Proc. Int. Reliability Phys. Symp., pp.152-155, 1984.

[39] Yamabe,K, Taniguchi,K, "Time-dependent Dielectric Breakdown of Thin Thermal Grown SiO_2 Films", IEEE Trans. Electron Devices, **ED-32(2)**, pp.423-428, 1985.

[40] Chen,C.F., Wu,C.Y., Lee,M.K., Chen,C.N., "The Dielectric Reliability of Intrinsic Thin SiO_2 Films Thermally Grown on a Heavily Doped Si Substrate-Characterization and Modeling", IEEE Trans. Electron Devices, **ED-34(07)**, pp.1540-1552, 1987.

[41] Hokari,Y, Baba,T, Kawamura,N, "Reliability of 6-10nm Thermal SiO_2 Films Showing Intrinsic Dielectric Integrity",IEEE Trans. Electron Devices, **ED-32(11)**, pp.2485-2491, 1985.

[42] Hirayama,M, Matsukawaa,T, Tsubouchi,N, Nakata,H, "Time Dependent Dielectric Breakdown Measurement of High Pressure Low Temperature Oxidized Film", Proc. Int. Reliability Phys. Symp., pp.146-151, 1984.

[43] McPherson,J.W., Baglee,D.A., "Acceleration Factors for Thin Gate Oxide Stressing", Proc. Int. Reliability Phys. Symp., pp.1-5.

[44] Osburn,C.M., Weitzman,E.J., "Electrical Conduction and Dielectric Breakdown in Silicon Dioxide Films on Silicon", J. Electrochem Soc., **119**, pp.603-609, 1972.

[45] Jeno,C.S., Ranganath,T.R., Huang,C.H., Jones,H.S., Chang,T.L., "High-field Generation of Electron Traps and Charge Trapping in Ultra-thin SiO_2", IEDM Tech. Digest, pp.388-391, 1981.

[46] Young,D.R., Irene,E.A., DiMaria,D.J., De Keersmaecker,R.F., "Electron Trapping in SiO_2 at 295 and 77°K", J. Appl. Phys., **50(10)**, pp.6366-6372, 1979.

[47] Dutoit,M, Fazan,P, Benjelloun,A, Ilegems,M, "Charge Trapping and Breakdown in Thin SiO_2 Polysilicon-gate MOS Capacitors", Appl. Surface Science, **30**, pp.333-338, 1987.

[48] Liang,M-S, Hu,C, "Electron Trapping in Very Thin Thermal Silicon Dioxides", IEDM, pp.396-399, 1981.

[49] Miki,H, Noguchi,M, Yokogawa,K, Kim,B-W, Asada,K, Sugano,T, "Electron and Hole Traps in SiO_2 Films Thermally Grown on Si Substrates in Ultra-Dry Oxygen", IEEE Trans. Electron Devices, **35(12)**, pp.2245-2251.

[50] Haddad,S, Liang,M-S, "The Nature of Charge Trapping Responsible for Thin-Oxide Breakdown Under a Dynamic Field Stress", IEEE Trans. Electron Lett., **EDL-8(11)**, pp.524-527, 1987.

[51] Heyns,M.M., "Degradation and Wearout of Thin Dielectric Layers during Charge Injection", ESSDERC, pp.361-368, Sept., 1990.

[52] Avni,E, Loev,L, Shappir,J, "Temperature Effects on Electron Trap Generation and Occupation in SiO_2", J. Appl. Phys., **63(8)**, pp.2700-2703, 1988.

[53] Amerasekera,E.A., Campbell,D.S., "Electrostatic Pulse Breakdown in NMOS Devices", Quality and Reliability Engineering International, **2**, pp.107-116, 1986.

[54] Greason,W.D., Kucerovsky,Z, Chum,K.W.K., "Latent Effects Due to ESD in CMOS Integrated Circuits: Review and Experiments", IEEE Trans. Industry Applications, **29(1)**, pp.88-97, 1993.

[55] Tunnicliffe,M.J., Dwyer,V.M., Campbell,D.S., "Latent Damage and Parametric Drift in Electrostatically Damaged MOS Transistors", Journal of Electrostatics, **31**, pp.91-110, 1993.

[56] Hawkins,C.F., Soden,J.M., "Electrical Characteristics and Test Considerations for Gate Oxide Shorts in CMOS ICs", Proc. International Test *Conf., pp.544-555, 1985.*

CHAPTER 3

Experimental Apparatus and Procedure

3.1 Experimental Apparatus

3.1.1 Wafer Chuck and Microprober System

Two different wafer chucks were used to mount the silicon wafers. The first, intended for 4" wafers, was made of brass and held the wafer in place by means of a vacuum pump system. The second type, designed for 3" wafers, was made of aluminium and held the wafers in position using a system of mechanical clips arranged around the wafer periphery. The 3" chuck was heated by a series of wire-wound resistors, embedded in the body of the chuck, while its temperature was monitored by a Cr/Al thermocouple. This system accurately stabilised the chuck (and hence the wafer) at any user-defined temperature between ambient and 200°C. Temperature was independently monitored via a second thermocouple by a Comark Type 1602-2 electronic thermometer.

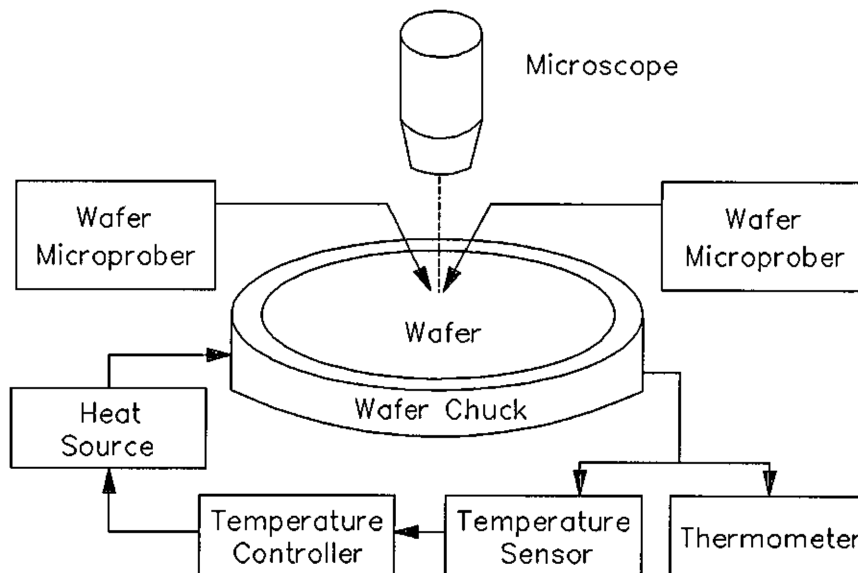


Figure 3.1: Wafer microprober station (together with temperature control system).

Electrical contact to the device structures was made using adjustable 20μm tip microprobes, by which the device bond pads were manually probed under microscopic observation. Substrate contact was made using the chuck itself, which was in contact with the metallised wafer base. The overall arrangement is shown in Fig.3.1.

3.1.2 Electrical Instruments

Various electronic instruments were used to study the semiconductor wafers, all of which are described below.

3.1.2.1 HP4145B Parametric Analyzer

Parametric analysis of the devices was performed using a Hewlett Packard HP4145B semiconductor parametric analyzer, which is capable of performing a number of different functions. The HP4145B is basically a programmable d.c. measurement system consisting of two voltage sources, two voltage monitors and four source-monitor units (SMUs). The latter are two-mode devices which can source current and monitor voltage or source voltage and monitor current. The arrangements is shown schematically in Fig.3.2.

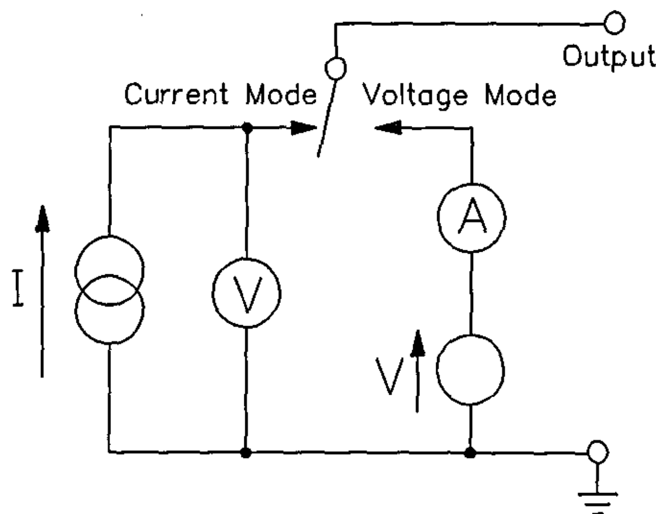


Figure 3.2: Schematic diagram of Source-Measure Unit (SMU).

The parametric analyzer can measure the static I/V characteristics of a device or alternatively measure current or voltage as a function of time. Accuracy is ultimately limited by system noise, and measurements to within 10mV/10nA are readily attainable. The instrument also has a range of analytical functions (e.g. measurement of the slope/intercept of a tangent to an I/V characteristic), enabling the user to extract device parameters from the measured characteristics.

Measured data were stored on 3½" floppy disks for future retrieval. Hard copies of characteristics were also obtained using a Hewlett Packard HP7475A plotter.

3.1.2.2 Programmable Voltage Source

The Keithley 230 programmable voltage source provides precision d.c. voltages between 0 and 100V in both polarities with a resolution below 10mV and an output current limit between 2mA and 100mA. The unit has an internal program memory, allowing it to generate a user-defined voltage waveform.

3.1.2.3 L-C-R Meter

The Wayne-Kerr 4210 L-C-R bridge was used to measure high frequency circuit elements. This instrument measures resistance, capacitance and inductance in parallel or series configuration using sampling frequencies of 100Hz, 1kHz and 10kHz. It also incorporates a "suppress" function, which allows the parasitic circuit elements of the test probes and leads to be eliminated from the measurements.

3.1.2.4 HP 54111D Digital Oscilloscope

The Hewlett Packard 54111D digital oscilloscope can observe single-shot events and transfer the captured waveform to a computer via an HPIB interface. The controlling software, developed by Dr.A.J.Franklin [1], runs on a Walters 286 AT compatible desktop computer and allows the waveform data to be stored on a disk or transferred to the Mathcad software for analysis. The maximum speed of the oscilloscope is 10^9 samples per second, allowing the measurement of frequency components up to 500MHz.

The signals monitored by the oscilloscope were either voltages, measured by the HP10431A (1MΩ/6.5pF) and 10440A (10MΩ/2.5pF) voltage probes, or electrical currents, measured by the Tektronix CT1 current-transformer probe.

The signals from the voltage and current probes needed very little in the way of mathematical processing. Although probe attenuations were corrected by the oscilloscope itself, a problem was experienced with the zero level, whose value tended to drift. For this reason, the computer was programmed to calculate its own zero-voltage level by averaging the leading tenth of the waveform (prior to the beginning of the pulse) and then subtracting this value from the signal, i.e.

$$V(t) = V_{sc}(t) - \frac{10}{T_{sc}} \int_0^{T_{sc}/10} V_{sc}(\tau)\, d\tau \qquad (3.1)$$

where $V_{sc}(t)$ is the voltage signal from the oscilloscope and T_{sc} is the time duration of the waveform.

3.1.2.5 TG102 Signal Generator

A THANSAR TG102 pulse generator was used to provide an a.c. input to the test circuits. This generator has a frequency range of 0.2Hz-2MHz and a maximum output level of 20V peak-peak from 50Ω source impedance. It can generate sine, square and triangle waveforms with variable DC offset.

3.2 Test Samples

3.2.1 Devices on the Wafers

4" wafers which were composed of <100> 40-55Ωcm, p-type bulk silicon were used throughout the experiments. The wafers contained a selection of E-mode and D-mode nMOS transistor network, capacitor structures, various resistors, EPROM device arrays

and nMOS inverter circuits. All devices had gate-oxides of nominal thickness 40nm, and n^+-doped polysilicon gate electrodes.

All the MOSFETs on the arrays within each die were interconnected with common source and gate contacts and individual drain terminals, such that gate-stress was applied to the entire network at once. The total gate-area of the network was approximately $3 \cdot 10^{-4}$cm^2. Channel lengths of both E-mode and D-mode transistors varied between approximately 1 and 50μm.

Although various capacitors were fabricated on the wafers, only the rectangular gate oxide capacitors with two different dimensions (215x268μm, and 416x553μm) were used in the experiments. Some of these capacitors were unimplanted, while others had received preferential p-type doping. The MOS-C gate inputs were isolated from all other devices on the die.

An nMOS inverter (see Fig.3.3) consisted of an E-mode nMOS transistor and a D-mode nMOS transistor was fabricated on each die of the wafers. The dimensions of the transistors are 15x3.5μm and 5x7μm respectively.

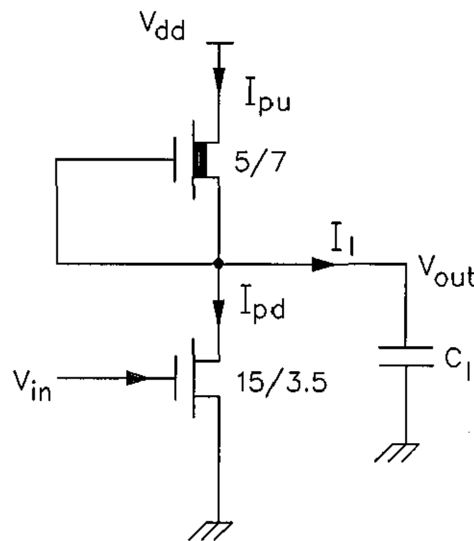


Figure 3.3: The nMOS inverter on wafers.

3.2.2 Characterization of Samples

3.2.2.1 C-V Curve Measurement

The Wayne-Kerr 4210 L-C-R meter was configured to monitor the capacitance of the MOS-C structures while a external voltage bias was linear increased from -4V to +4V. The measurement was carried out at two different frequencies 1kHz and 10kHz. The results are shown in Fig.3.4.

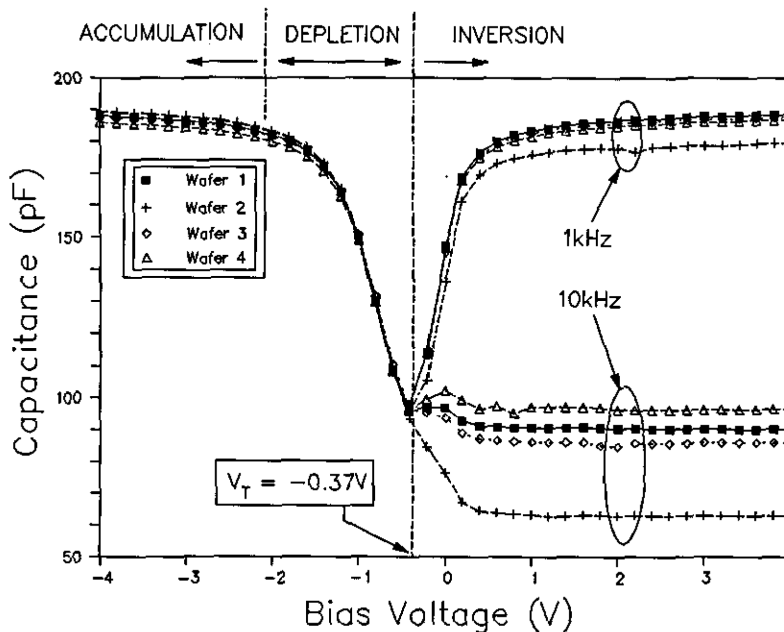


Figure 3.4: Capacitance versus voltage curves for the MOS-C structures.

3.2.2.2 Extraction of Oxide Thickness

Although all the devices have a nominal oxide thickness of 400Å, a more accurate measurement of local thickness was taken for each die using the wide-area 416x553μm MOS-C structure. It is assumed that all the other devices in the same die have the same oxide thickness.

The thickness (T_{ox}) for each die was gauged by measuring the capacitance (C_{ox}) of the wide-area capacitors. Measurement was performed at 10kHz using a Wayne-Kerr 420 L-C-R meter, with a -10V bias potential to ensure carrier accumulation at both oxide surfaces and this was guaranteed from the C-V curve shown in Fig.3.4. The oxide thickness for each die, estimated from the formula $T_{ox} = \epsilon_0\epsilon_{ox}\cdot\text{Area}/C_{ox}$ (where $\epsilon_0\epsilon_{ox}$ = oxide permittivity).

The results of thickness measurement for all the wafers are shown in Fig.3.5.

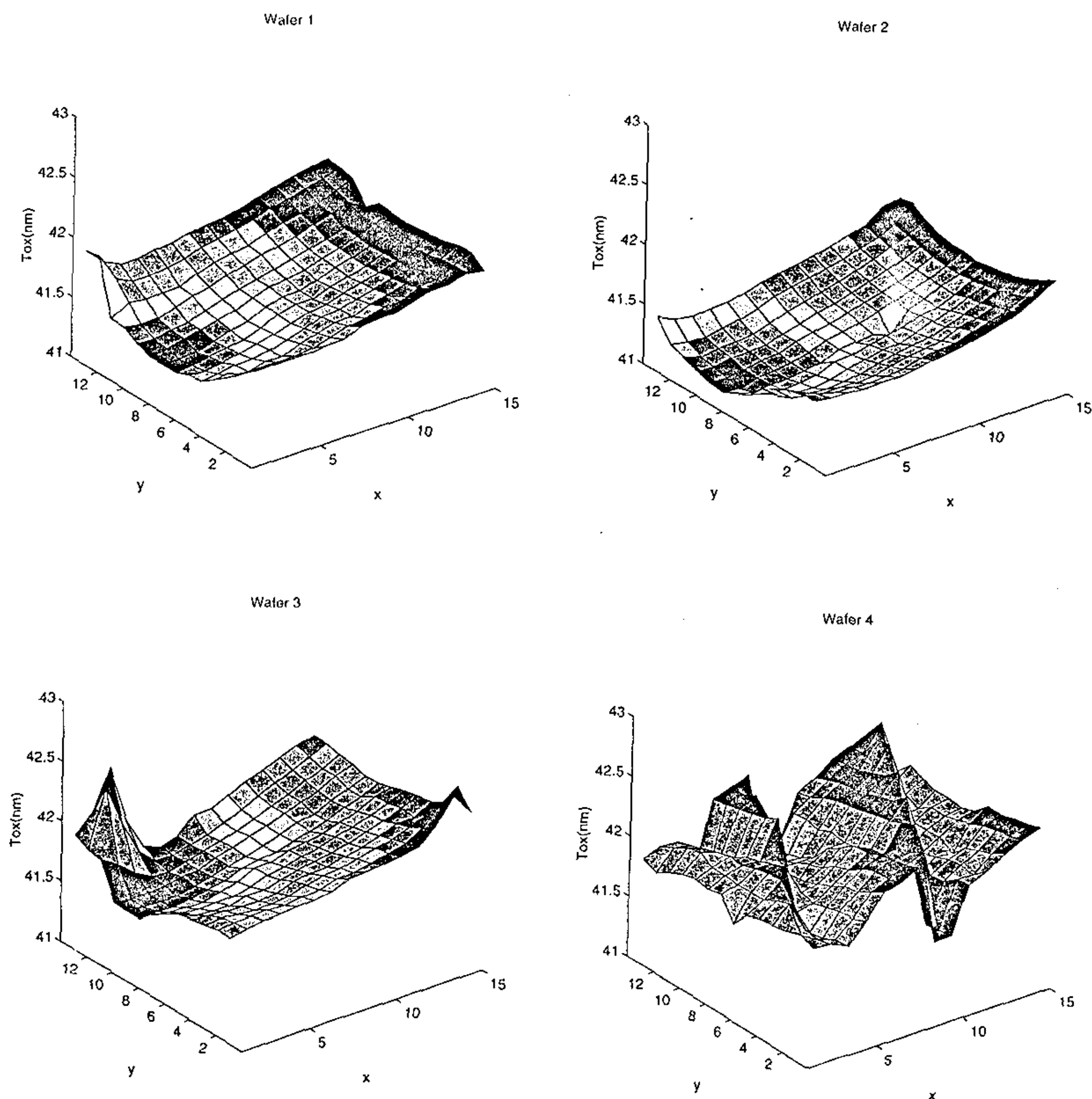


Figure 3.5: The 3-D surface of thickness vs. position (x,y).

3.2.2.3 The Measurement of Bulk Resistance

Four samples were chosen at different geographical positions for each wafer as shown in Fig.3.6.

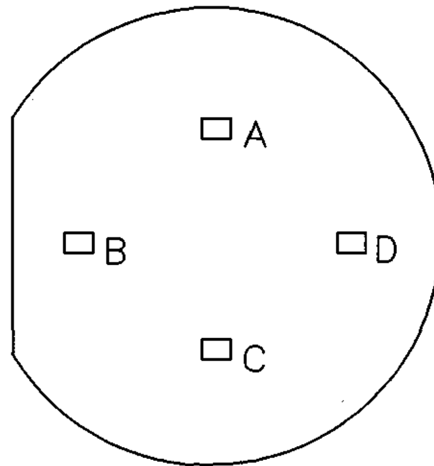


Figure 3.6: The positions of samples for bulk resistance measurement.

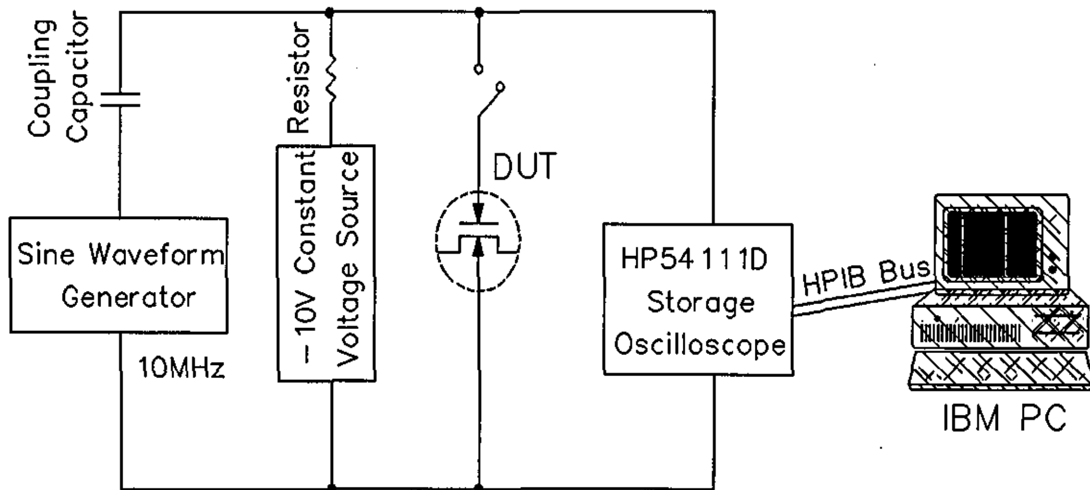


Figure 3.7: The apparatus for measuring bulk resistance of the MOS-C structures.

The apparatus shown in Fig.3.7 was used for the measurement of bulk resistance for the MOS-C structures and MOSFET, since the CR time constant is too small to be measured by L-C-R meter at 10KHz. A 10MHz Sine-wave signal was applied via a coupling capacitor to the DUT, together with a -10V d.c. bias to drive the Si surface into strong accumulation. Voltage and current profiles were captured by the HP54111D oscilloscope and downloaded to a Walters 286 computer, where the impedance Z_d of the structure was calculated. The capacitance and resistance were then extracted using the formulae:

$$R_b = \Re(Z_d) \qquad C_{ox} = \frac{1}{2\pi f |\Im(Z_d)|} \tag{3.2}$$

where f is the 10MHz sampling frequency, $\Re(Z_d)$ and $\Im(Z_d)$ are the real and imaginary parts of the impedance.

Table 3.1: Bulk resistance for MOS-C structures (Ohms).

Wafer No.		Sample A	Sample B	Sample C	Sample D	Average
1	Unimplanted	332.8	342.3	344.0	336.4	338.9
	Implanted	346.6	352.3	347.3	327.5	343.4
2	Unimplanted	250.6	239.7	248.7	244.7	245.9
	Implanted	249.3	255.3	256.9	251.3	253.2
3	Unimplanted	294.1	294.4	291.8	289.2	292.8
	Implanted	284.1	282.5	290.6	280.9	284.6

The bulk resistance in ohms are shown in Table 3.1 above, and average bulk resistance for each wafer was used in later data analysis since the difference between in-wafer measurements was negligible and well below the error in the measurement technique.

The bulk resistance of Wafer No.4 was measured as a function of temperature. The wafer was mounted on the 3" heated chuck and resistances were measured for temperatures between 20 and 150°C. The results are shown in Fig.3.8. The data appear to follow the standard linear law $R = R_0(1+\alpha(T-T_0))$ where T_0 is room temperature, R_0 is room temperature resistance and $\alpha \approx 0.0058/°C$ is the temperature coefficient of resistance.

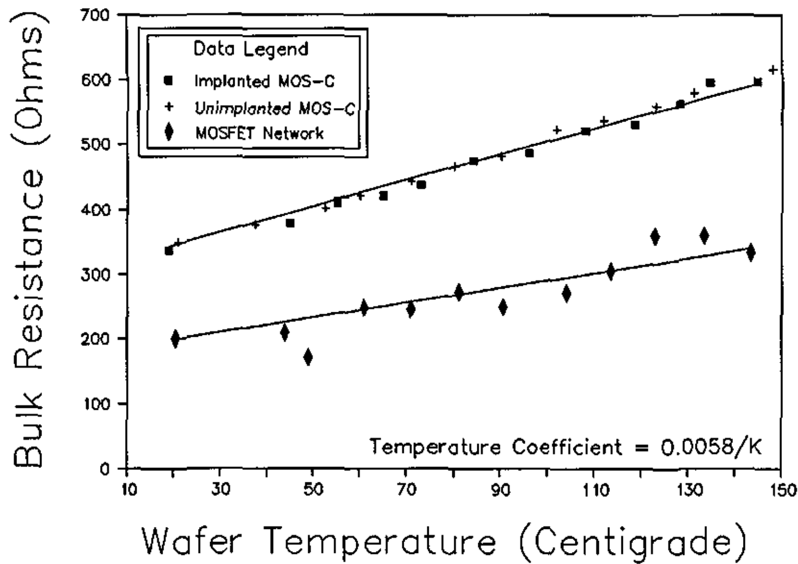


Figure 3.8: Bulk resistance for different devices on wafer No.4 as a function of temperature.

3.3 Experiments on MOS-C Structures and MOSFET Networks

3.3.1 Introduction

The general aims of these experiments were providing new data for the development of a gate-oxide dielectric breakdown model and the assessment of the MOS capacitor (MOS-C) as an oxide-integrity test vehicle. The TDDB experiments were performed on implanted and unimplanted MOS-C structures with dimensions of 215x268μm and nMOSFET networks on four different wafers. The 416x553μm dimension MOS-C structures were reserved for the measurement of gate oxide thickness. The TDDB experimental apparatus is shown in Fig.3.9.

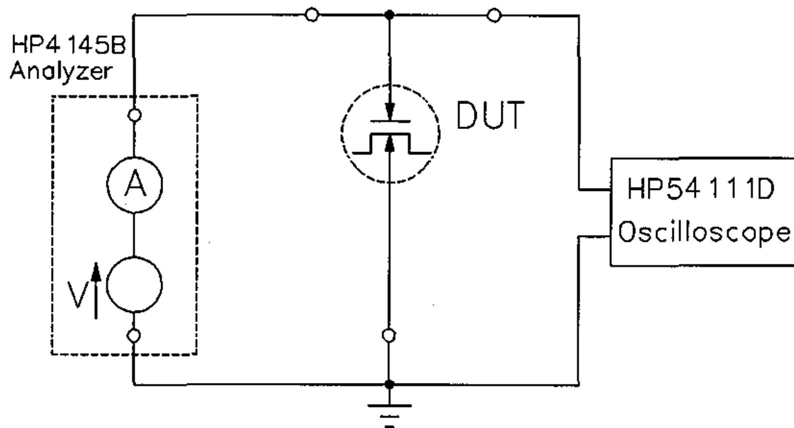


Figure 3.9: The apparatus used for TDDB experiments.

3.3.2 Field Accelerated TDDB Life Test

The MOS capacitor structures were subjected to constant voltage stress between gate and substrate at room temperature. The injection current to the gate oxide was monitored as a function of time by the HP4145B analyzer and voltage across the oxide was recorded

using the HP54111D storage oscilloscope. The oxide breakdown was easily detected as a rapid increase of injection current for large-time breakdown (t_{bd} > 1 sec.). The most accurate indication of breakdown for small-time breakdown (t_{bd} < 1 sec.) was a sudden drop of voltage across the oxide as measured by the oscilloscope. Devices stressed at each particular voltage were spread as uniformly as possible across the wafer in order to eliminate any positional dependencies.

The parameters t_{bd}, I_{max} (the maximum recorded injected current) and I_{stop} (injection current immediately prior to breakdown) were recorded for each capacitor, together with the stress voltage and the position of the die on the wafer.

3.3.3 Temperature Accelerated TDDB Life Test

MOS-C structures were subjected to -41.5V stress at four different temperatures, 50°C, 100°C, 150°C and room temperature. The 3" chuck with the temperature controlled heat source was used in the same instrument configuration as that of Section 3.3.2. The 4" wafer was initially divided into four quarters, each of which was mounted upon the 3" chuck. Die position, stress temperature, t_{bd}, I_{max} and I_{stop} were recorded for later analysis.

3.3.4 Experimental Results

This section describes the experimental results from MOS-C structures and MOSFETs. For each die, the measured results of t_{bd}, I_{max}, I_{stop}, together with the capacitance of the wide-area 416x553μm MOS-C structure were input on to a LOTUS 1-2-3 spreadsheet. Then the thickness of the die was calculated from the formula $T_{ox} = \epsilon_0 \epsilon_{ox} \cdot Area/C_{ox}$, and the oxide fields were estimated from $F_{ox} = (V_s - I_{ox}R)/T_{ox}$. The results are shown below.

3.3.4.1 Time-dependent Failure Distributions

The failure distributions of the unimplanted MOS-C structures for the four wafers are shown in Fig.3.10 together with the failure distributions for the implanted MOS-Cs and

the MOSFETs from wafer 2.

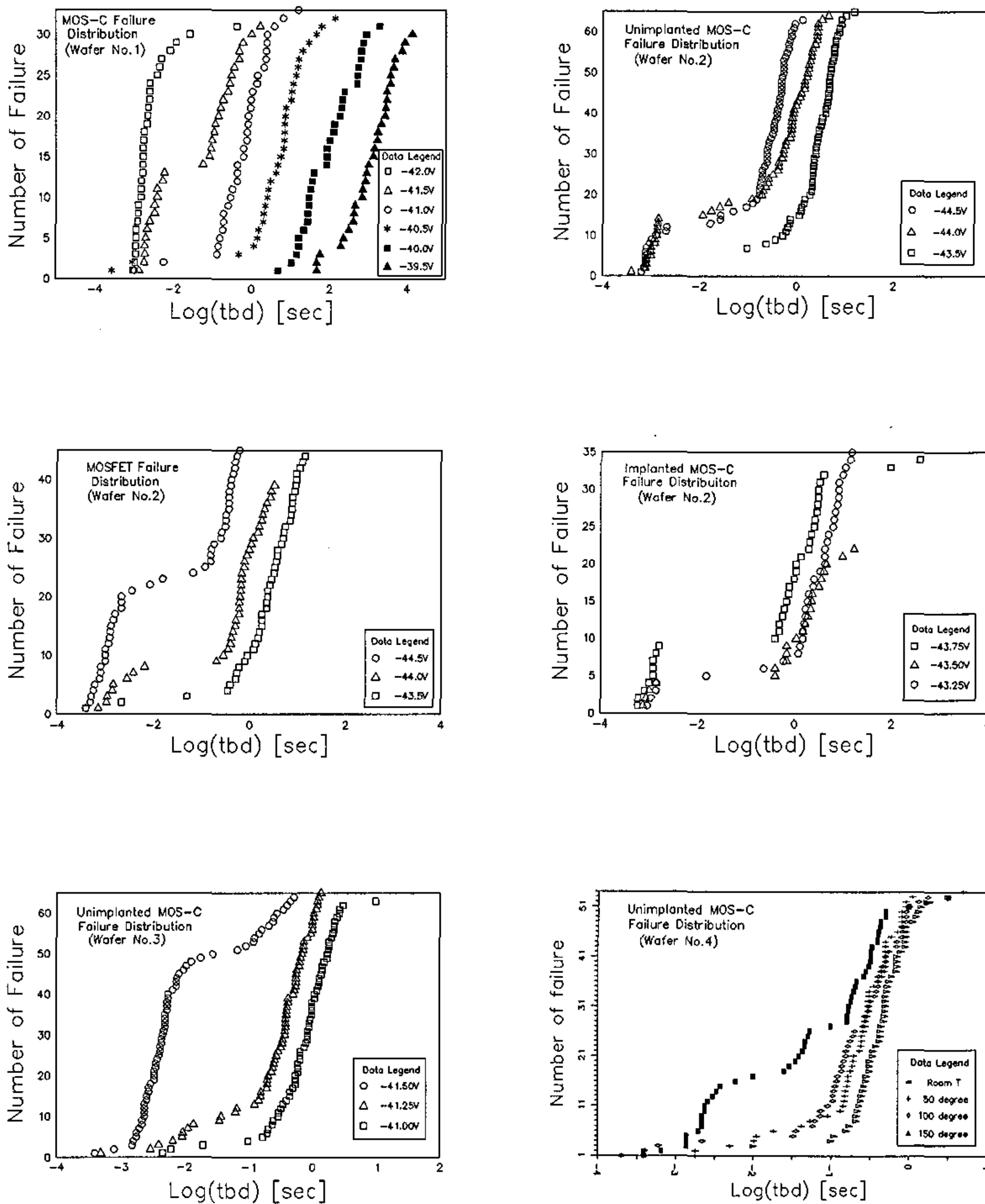


Figure 3.10: Failure distributions for unimplanted MOS-C structure (wafer No.1,2,3,4) under constant voltage stress, implanted MOS-C structures and MOSFETs (wafer No.2 only).

It is clearly shown that the failure distributions are bi-modal, consisting of 'early' failures (t_{bd} < 100ms) and 'late' failures (t_{bd} > 100ms). The ratio of the numbers of 'early' and 'late' failures generally increases with the stress voltage.

3.3.4.2 Injection Current

The injection current profiles were monitored using HP4145B parametric analyzer during the stresses. Typical current profiles are shown in Fig.3.11.

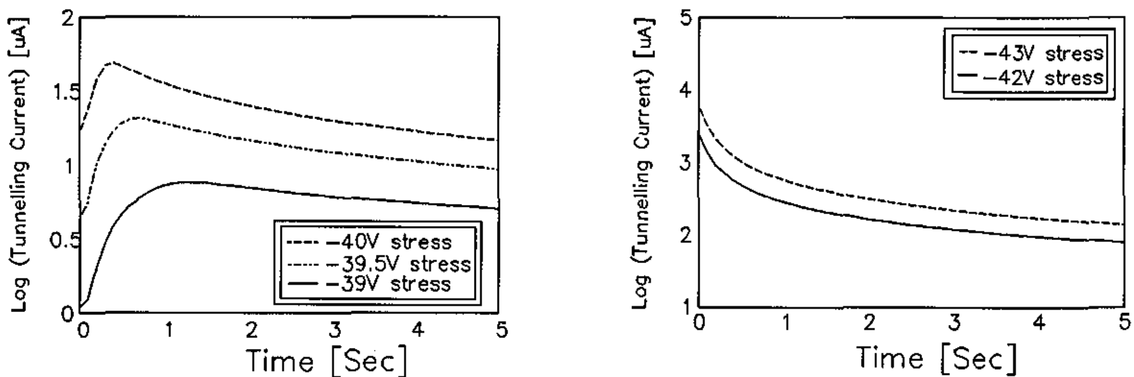
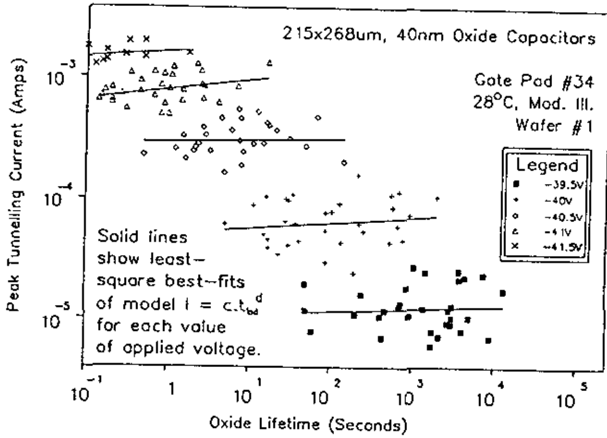


Figure 3.11: The injection current profiles of different constant voltage stresses for wafer No.2.

Fig.3.12 and 3.13 show the initial and final injection currents (the injection current at breakdown) plotted against t_{bd}. In the case of initial current data, there is little correlation between log(current) and log(t_{bd}), although both are strong function of voltage. Least-square analysis suggested a slight increase in current with t_{bd}, although the correlation fails to meet the criterion for the 95% confidence level. The final current data can be satisfactorily fitted to a model of the form $I_{bd} = a(V) \cdot t_{bd}^{-b(V)}$, where I_{bd} is the final current and a(V) and b(V) are parameters which depend only upon voltage (see inset). It is clear that the majority of the voltage dependence appears in a(V), while b(V) remains mostly between 0.7 and 0.8 throughout the voltage range.



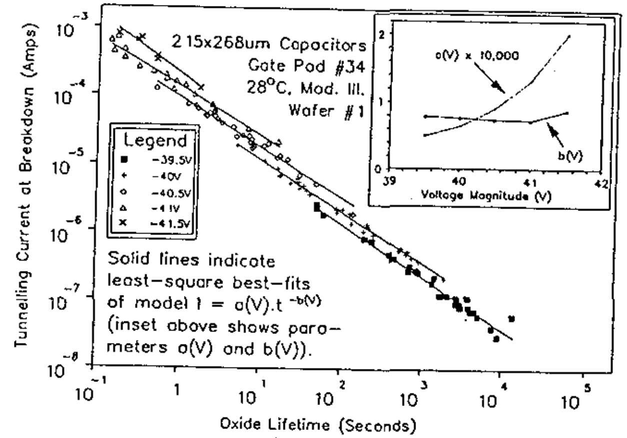


Figure 3.12 & 3.13: The initial and final injection current as a function of time-to-breakdown.

The logarithm of the maximum injection currents were found to decrease linearly with 1/F at lower fields and saturates at higher fields (Fig.3.14).

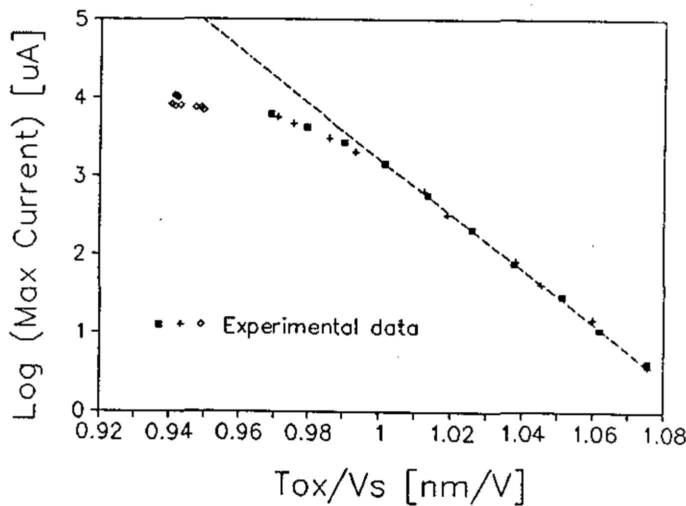


Figure 3.14: The relationship between maximum injection current and stress field.

3.4 Experiments on nMOS Inverter

The following experiments investigate the effects of the parametric drift caused by gate oxide breakdown upon circuit operation, with an aim to develop a technique for detecting its existence. The experiments were performed on the nMOS inverter (see Fig.3.3). The wafer was held on the 4" brass chuck (described in section 3.1.1) under room temperature and moderate illumination.

3.4.1 Characterisation of the Pull-down Transistor

A wafer level transistor characteristics can be easily measured by connecting the HP4145B analyzer's SMUs through a test fixture 16058A to the microprobe station. During the measurement, the drain electrode of the pull-up transistor was left floating in order to ensure current accuracy. The characterisations were performed using three SMUs connected to the source, drain and gate electrode of the pull-down transistor. The substrate of the transistor was connected to the source.

The I_G - V_G characteristics of the pull-down transistor remained below the 20pA level throughout the sweep indicating that the gate oxide worked well as a dielectric (its resistance is around 250GΩ). The I_{DS}-V_{DS} and I_{DS}-V_{GS} characteristics of the pull-down transistor nMOS inverter are shown in Fig.3.15 and Fig.3.16 respectively.

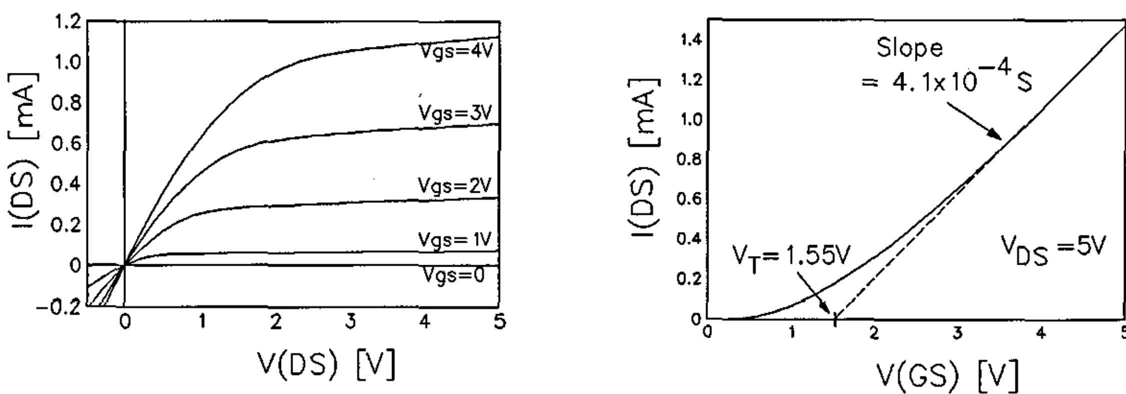


Figure 3.15 & 3.16: I_{DS} - V_{DS} & I_{DS} - V_{GS} characteristics of the pull-down transistor.

3.4.2 Characterisation of the nMOS Inverter

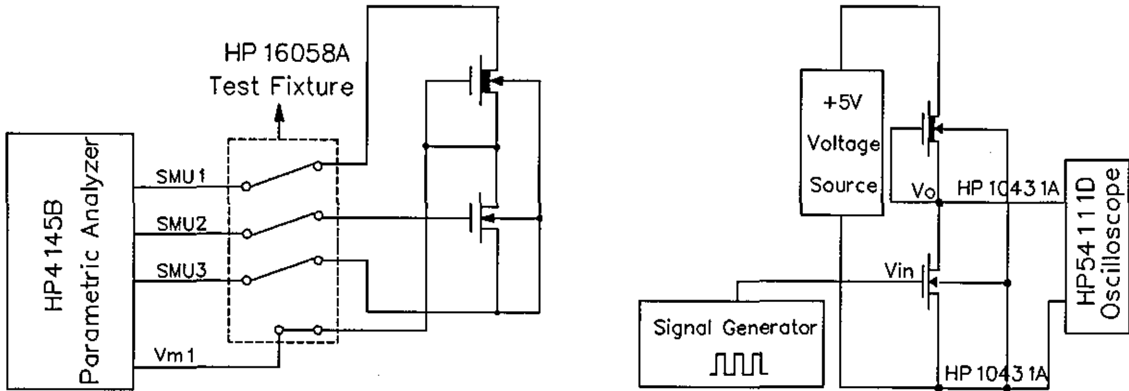


Figure 3.17 & 3.18 : The apparatus used for characterisation of the inverter.

Fig.3.17 and 3.18 show the apparatus used to extract the DC and transient characteristics of the inverter.

The DC voltage transfer characteristics and the supply current (I_{dd}) characteristics are shown in Fig.3.19 and 3.20 by the solid lines.

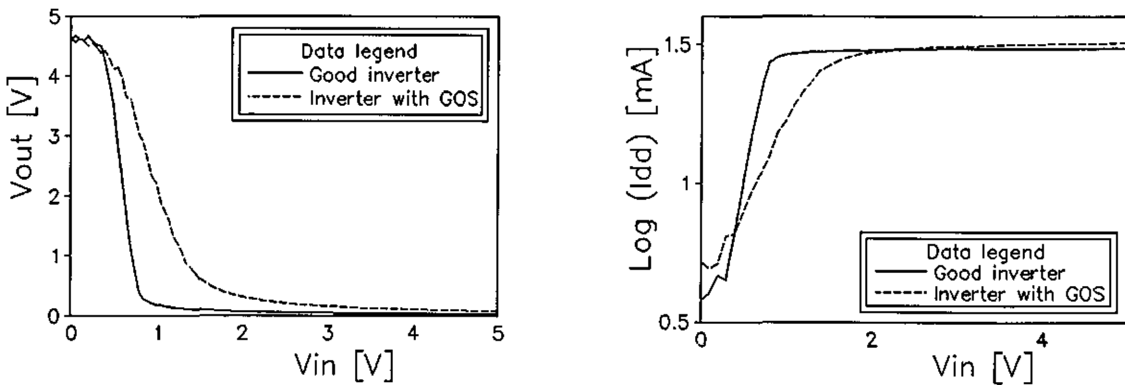


Figure 3.19 & 3.20: The DC transfer curves for the nMOS inverter.

The transient response of the inverter (whose d.c. characteristics are shown in Fig.3.19 & 3.20 by the solid lines) to a 250kHz square input waveform from the THANSAR TG102 generator was captured by HP54111D digital oscilloscope and is shown in Fig.3.21 by the solid line.

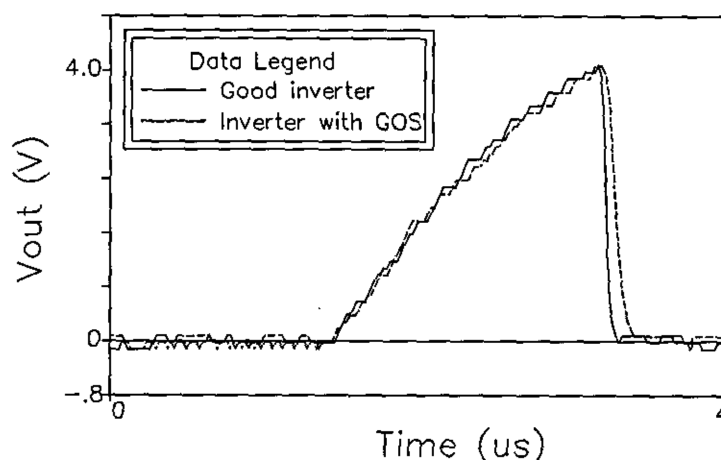


Figure 3.21: The transient response of the inverter.

3.4.3 Constant Voltage Stress on Gate Oxide of the Pull-down Transistors

The pull-down transistors were then subjected to constant voltage stress from HP4145B analyzer. The -43V stress was applied between the gate electrode and the substrate of the transistors in order to cause gate dielectric breakdown. The injection current profile was monitored by the analyzer and breakdown was detected by a rapid rise in the injection current. The stress was removed immediately after the oxide breakdown.

3.4.4 Re-characterisation of the Pull-down Transistors

After the gate oxide breakdown, the pull-down transistors were re-characterised. The occurrence of gate oxide breakdown was confirmed by the I_G - V_G characteristics shown in Fig.3.22. The existence of a tens-of-μA gate current meant that a gate oxide short was

induced during the constant voltage stress.

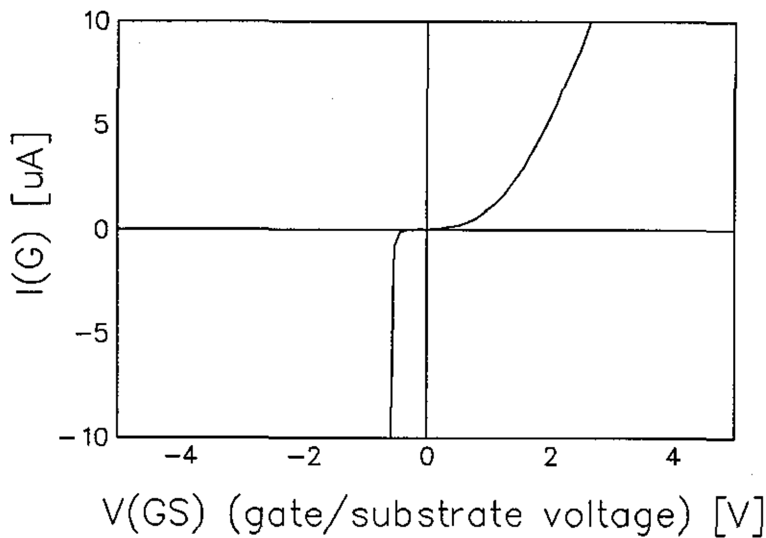


Figure 3.22: The I_G - V_G characteristics of the pull-down nMOSFET with a GOS.

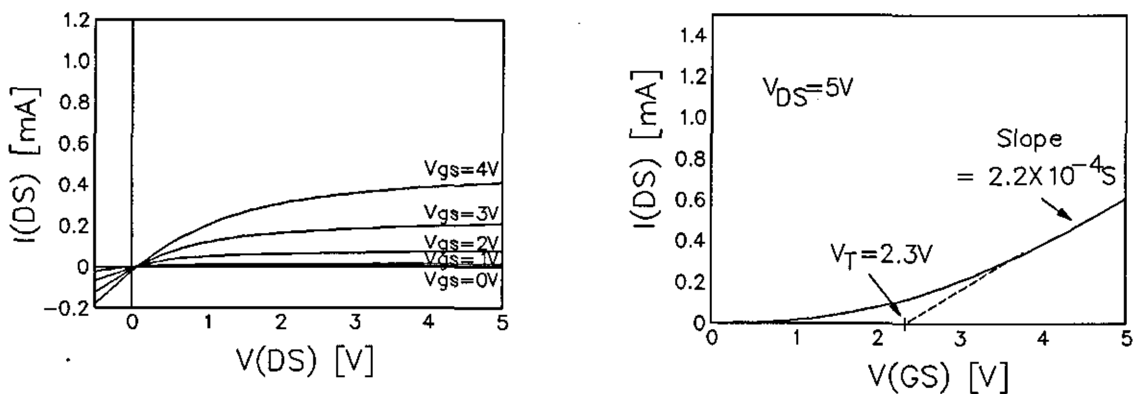


Figure 3.23 & 3.24: The I_{DS} - V_{DS} and I_{DS} - V_{GS} characteristics of the pull-down nMOSFET with a gate oxide breakdown.

The I_{DS} - V_{DS} and I_{DS} - V_{GS} characteristics of the breakdown pull-down transistors are shown in Fig.3.23 and Fig.3.24. Comparing with Fig.3.15 and Fig.3.16, the threshold

voltage V_T of the transistor increased from 1.55V to 2.3V while the mutual transconductance g_m of the transistor dropped from $4.1 \cdot 10^{-4}$(S) to $2.2 \cdot 10^{-4}$(S).

3.4.5 Re-measurement of the Characteristics of the Inverter

After the constant voltage stress removed from the pull-down transistor, the nMOS inverter's DC transfer curves and the transient response were re-measured. The results are shown in Fig.3.19, Fig.3.20 and Fig.3.21 by broken lines. An increase in the inverter's threshold voltage V_T, which related to the threshold of the pull-down transistor, was observed together with a small amount increment of power supply current.

3.5 Experimental Conclusions

Oxide breakdown were experimentally studied in this chapter. Emphasis was placed on the injection current in order to find the physical mechanism during the build-up stage of the breakdown.

The effects of the presence of a GOS upon the operation of a simple nMOS inverter circuit have been studied in this Chapter. Parametric drifts are found to occur in MOSFETs after the oxide breakdown. This gives rise to changes in the DC characteristics in of the inverter and, in addition, an extra delay in ac operation.

3.6 References

[1] Franklin,A.J., "Electrical Overstress Failure in GaAs MESFET Structures", Ph.D Thesis, Loughborough University of Technology, 1990.

[2] HP 4145B Analyzer Manual.

[3] HP 54111D User Documentation (Installation/Operating Manual).

[4] Tunnicliffe,M.J., "Electrical Overstress and Electrostatic Discharge Failure in Silicon MOS Devices", Ph.D Thesis, Loughborough University of Technology, 1993.

CHAPTER 4

Analysis of the Experimental Results

4.1 Introduction

The aims of this Chapter are threefold:

(1) To develop a mathematical model of oxide breakdown. The theory is based upon charge trapping mechanism in the dielectric, and the model predictions are then compared with the experimental data.

(2) To assess the usefulness of the MOS capacitor as a test vehicle for monitoring the oxide integrity of active MOS circuitry, and to model the distributions of oxide failure.

(3) To qualitatively explain the reason why the transconductance g_m of MOSFET dropped after the gate oxide breakdown.

4.2 Modelling of the Oxide Breakdown

4.2.1 Introduction

Numerous experiments have shown that MOS oxide breakdown is the result of a continuous irreversible 'erosion' of the SiO_2 layer. This mechanism is known as **time-dependent dielectric breakdown** or **TDDB**. According to the widely accepted model [3], the TDDB process may be divided into two stages. During the 'build-up' stage, localized high-field/current-density regions are formed as a result of charge trapping. Eventually, when the local current density or field reaches a critical value, the rapid 'runaway' stage begins. During this stage, catastrophic electrical and/or thermal processes result in breakdown. The runaway stage, once reached, is completed in a very short period of time.

Hence the time necessary to reach the runaway stage determines the lifetime of the oxide.

4.2.2 The Mechanisms of Oxide Wearout

Although numerous theories have been proposed, no consensus has ever been reached concerning the physical mechanisms involved.

4.2.2.1 The Entry of Charges into Oxide Dielectrics

Most researchers agree that the leakage current through intact oxide dielectrics is due to quantum-mechanical tunnelling of electrons through the potential barrier at the negative electrode [1-4]. This phenomena is well understood and the current-voltage dependence can be predicted by the Fowler-Nordheim (F-N) tunnelling theory [5] (Eq.(4.1)).

$$J = A F_{cat}^{2} \exp(-\frac{B}{F_{cat}}) \tag{4.1}$$

where J is the tunnelling current density, F_{cat} is the cathode electric field and A & B are constants. Although the latter constants may be computed from the Si/SiO_2 barrier height and the effective electron mass in SiO_2, the perturbatory effects of pre-existing trapped charges make it more accurate to measure the I-V characteristics and determine the constants empirically [2].

4.2.2.2 Charge Trapping in Oxides

Charge trapping in thermal oxides grown on single crystal silicon has already been studied extensively and is briefly mentioned in Section 2.4.

Electrons are believed to be captured in both the pre-existing electron traps in the oxide as well as the traps that are generated during the high field stress. It is common to use the first-order rate equation (4.2) to model electron trapping in SiO_2 films [2,9].

$$\frac{\partial N_{ot}(\sigma,\phi)}{\partial \phi} = \sigma[N_o(\sigma) - N_{ot}(\sigma,\phi)] \tag{4.2}$$

where N_{ot} is the filled trap density per unit area, N_o is the total trap density per unit area (which is assumed to be constant), σ is the trap capture cross-section and ϕ is the electron flux, i.e. the number of injected electrons per unit area per unit time. The solution to Eq.(4.2) is

$$N_{ot}(\sigma,\phi) = N_o(\sigma)(1 - e^{-\sigma\phi}) \tag{4.3}$$

Since no such saturation was observed in the electron trapping, a trap generation mechanism is suggested.

Meyer and Crook [2] postulated that the trapped charge density is an equilibrium of competing trapping and de-trapping process, which is modelled in equation (4.4).

$$N_{ot} = \frac{G}{R}(1 - \exp^{-R\phi}) \tag{4.4}$$

where G is the trap generation rate, and R is the trap recombination cross section.

Liang and Hu [9] assumed a constant trap generation rate g defined by Eq.(4.5)

$$g = \frac{\partial N_o}{\partial t} = \frac{\partial N_o}{\partial \phi}\frac{\partial \phi}{\partial t} = \frac{J}{q}\frac{\partial N_o}{\partial \phi} \tag{4.5}$$

This means that the total trap density N_o increases with time or flux. The first-order dynamic rate equation must now be replaced by a system of two differential equations

$$\frac{\partial N_{opt}(\sigma,\phi)}{\partial \phi} = \sigma_p[N_{op}(\sigma_p) - N_{opt}(\sigma_p,\phi)] \tag{4.6}$$

and

$$\frac{\partial N_{ogt}(\sigma_g,\phi)}{\partial \phi} = \sigma_g[q\frac{g}{J}\phi - N_{ogt}(\sigma_g,\phi)] \tag{4.7}$$

where N_{op} is the pre-existing trap density, N_{opt} is the density of filled pre-existing traps, and N_{ogt} is the density of the filled generated traps. The solution of Eqs.(4.6) and (4.7) subjected to the initial condition $N_{opt}(\phi=0) = N_{ogt}(\phi=0) = 0$ is

$$\begin{aligned} N_{ot}(\sigma_p,\sigma_g,\phi) &= N_{opt}(\sigma_p,\phi) + N_{ogt}(\sigma_g,\phi) \\ &= N_{op}(\sigma_p)(1-e^{-\sigma_p\phi}) + q\frac{g}{J}[\phi - \frac{1}{\sigma_g}(1-e^{-\sigma_g\phi})] \end{aligned} \tag{4.8}$$

Positive charge trapping was observed by many investigators (Section 2.4) and several physical mechanisms have been proposed. The impact ionization-recombination (IR) model is based on the assumptions that a small proportion of the injected electrons gain sufficient energy to produce impact ionization at the rate α per unit length, and that the only opposing process is recombination with electrons from the conduction band [10,11]. Under these assumptions, Klein and Solomon [11] suggested a simplified rate equation:

$$\frac{\partial p}{\partial t} = \frac{J}{q}(\alpha - \sigma p) \tag{4.9}$$

where J is the tunnelling current density, p is the positive charge density, σ is the capture cross section for recombination (which is assumed to be constant), and q is the electronic charge.

However, Jenq et al. [12] have argued that the positive charges are formed by direct tunnelling between the anode and the trap sites near the SiO_2/Si interface, although no

quantitative expression of this mechanism has been published.

4.2.2.3 Electric Field Change due to Trapped Charges in Oxides

A sheet of charge trapped at position x from the gate causes band bending. Chen et al. [3] derived expressions for the electric field profile in the oxide, based upon a two-dimensional charge-sheet model of the trapped carriers (Fig.4.1).

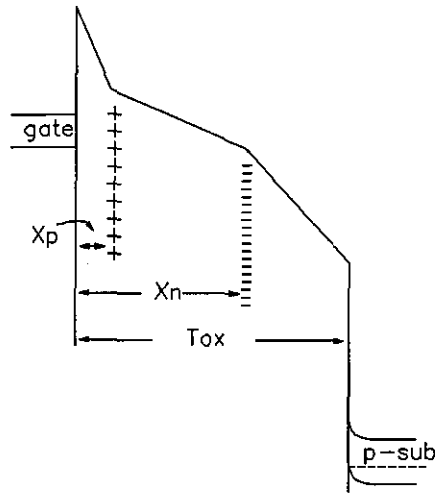


Figure 4.1: Energy-band diagram for MOS capacitor with both negative and positive charge trapped in the oxide area. x_n and x_p are the centroids of trapped negative and positive charge respectively.

In the case of constant voltage stress (negative voltage applied to gate electrode), the cathode field is given by equation (4.10)

$$F_{cat} = \frac{V_{ox}}{T_{ox}} - \left(1 - \frac{x_n}{T_{ox}}\right)\frac{Q_{ot}^-}{\epsilon_{ox}} + \left(1 - \frac{x_p}{T_{ox}}\right)\frac{Q_{ot}^+}{\epsilon_{ox}} \qquad (4.10)$$

in which Q_{ot}^+ is the magnitude of the trapped hole charge density, and Q_{ot}^- is the magnitude of the trapped electron charge density, x_p and x_n are the respective centroids of Q_{ot}^+ and Q_{ot}^- (measured from the cathode), and V_{ox} is the magnitude of the applied stress voltage.

4.2.2.4 The Breakdown of Oxide

During the rapid runaway stage, catastrophic electrical and/or thermal processes result in breakdown. This is believed to be triggered when the local current density or field reaches a critical value. The runaway stage, once reached, is completed in a very short period of time, and hence the time necessary to reach the runaway stage determines the lifetime of the oxide. Harari [1] theorized that the breakdown mechanism is intimately related to the generation of the electron traps, which increase the local internal field to the runaway limit. Chen et al. [3] provided experimental evidence which suggested that hole trapping in localized areas at the cathode is responsible for breakdown in SiO_2 films subjected to high field stress. The holes were believed to be generated by impact ionization.

4.2.3 Mathematical Model

4.2.3.1 Physical Mechanisms

In Chapter 3, the tunnelling current profiles of wafer 2 were recorded under constant gate-voltage using the HP4145B parametric analyzer. For the relatively low stress voltages (-39V, -39.5V and -40V), the resulting tunnelling currents (Fig.3.19 left) increased from their initial values to a maximum, before decaying with time. For higher stress voltages (-42V and -43V), the current appeared to decay from the very beginning (Fig.3.19 right).

The tunnelling current profiles may be characterised by two parameters: (i) the maximum current and (ii) the time at which this current was reached. The variation of these parameters with stress voltage is shown in Fig.4.2 and 4.3.

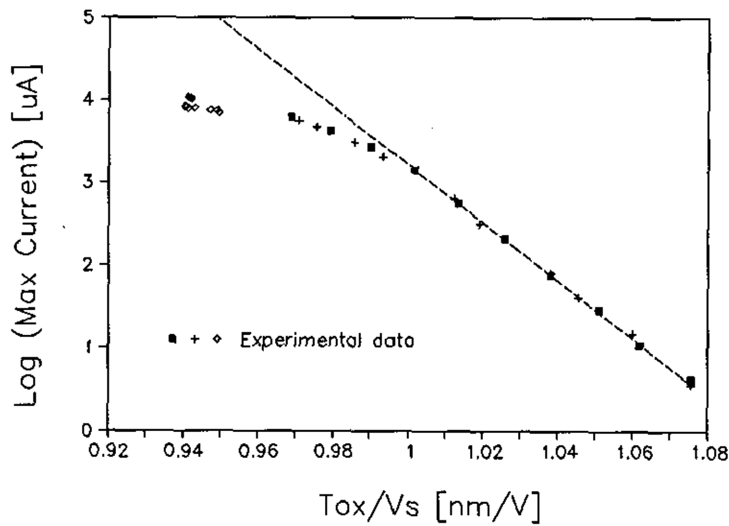


Figure 4.2: Maximum current vs. stress field.

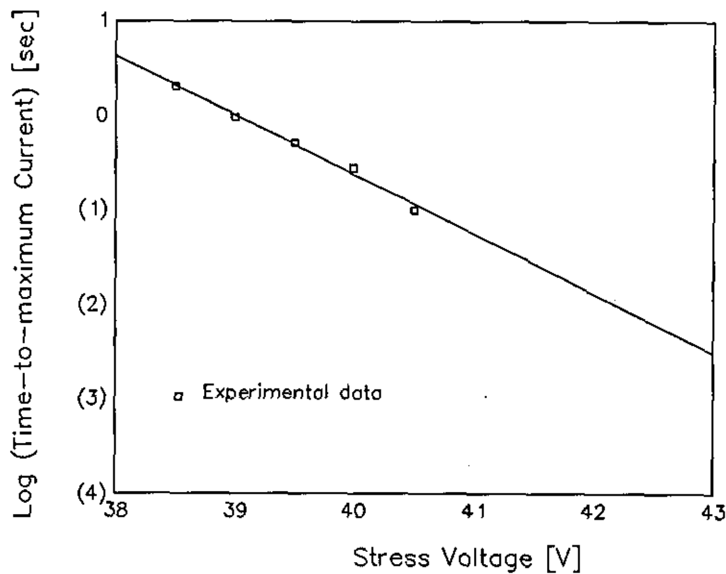


Figure 4.3: The time of maximum tunnelling current as a function of stress voltage.

The logarithm of the maximum tunnelling current varies linearly with 1/F at lower fields (in accordance with standard tunnelling theory [5]) but appears to saturate at higher fields (Fig.4.2).

The experimental results above can be qualitatively explained by the charge trapping theory proposed by Jenq et al. [12]. They suggested that two types of charges are introduced in the thin oxide during the high-field stress. One is a positive charge, believed to be formed by direct hole tunnelling from the anode into the trap sites near the SiO_2/Si interface. The other is negative, caused by the trapping of injected electrons in both pre-existing trap sites and traps that are generated during the high field stress.

When the stress is applied to the capacitor, both hole trapping and electron trapping occur. However, the hole trapping dominates initially because it is due to *direct* tunnelling from the anode to the fixed interface trap sites. Hence the net trapped charge in the oxide is initially positive, causing an increase in the cathode field, and hence the tunnelling current. However, since the tunnelling-hole trapping is such a fast process, the available hole traps are filled very quickly. Afterwards, as further electrons are trapped in the oxide and as additional electron traps are generated, the cathode field decreases, together with the tunnelling current. This leads to the profiles shown in Fig.3.19.

4.2.3.2 Quantitative Model

The oxide current is determined by the Fowler-Nordheim equation [5]:

$$J(t) = A F_{cat}^{2}(t)\, e^{-\frac{B}{F_{cat}(t)}} \tag{4.11}$$

where J(t) is the tunnelling current density, F_{cat} is the cathode field and A and B are constants.

The non-saturating tunnelling current decay observed in Fig.3.19 suggests a continuous generation of electron trap sites. From the first order rate equations (4.8), the trapped electron density Q_{ot}^{-} can be expressed as:

$$Q_{ot}^{-} = qN_{op}\left(1 - e^{-\int_0^t \sigma J dt/q}\right) + \int_0^t gJ(t)dt - \frac{qg}{\sigma_g}\left(1 - e^{-\int_0^t \sigma_g J dt/q}\right) \tag{4.12}$$

where σ and σ_g are the respective capture cross-sections of pre-existing traps and newly generated traps, N_{op} is the pre-existing trap density and g is the trap generation rate.

The increase in the oxide current seen in Fig.3.19 can be explained as a consequence of positive charge trapping. The increment in the trapped charge in time interval dt is given by:

$$dQ_{ot}^{+}(t) = J_{hole}(t)dt \tag{4.13}$$

where J_{hole} is the hole tunnelling current. The latter may reasonably be assumed to have a F-N type dependence upon the anode field F_{an} and to be proportional to the density of available hole traps in the anode, i.e.

$$J_{hole} = Ce^{-\frac{D}{F_{an}}}\left(1 - \frac{Q_{ot}^{+}}{qN_{int}}\right) \tag{4.14}$$

where N_{int} is the total density of hole traps and C and D are constants dependent upon structure of the anode/oxide/trap system. The factor in parenthesis causes the trapping to slow as the available number of states decreases, and to tend to zero as $Q_{ot}^{+} \rightarrow qN_{int}$.

The oxide field profile changes when charges are trapped within the dielectric. If we again assume sheet charges located at their respective centroids, then the electric field at cathode will be given by:

$$F_{cat} = \frac{V_{ox}}{T_{ox}} - \left(1 - \frac{x_n}{T_{ox}}\right)\frac{Q_{ot}^{-}}{\epsilon_{ox}} + \left(1 - \frac{x_p}{T_{ox}}\right)\frac{Q_{ot}^{+}}{\epsilon_{ox}} \tag{4.15}$$

The voltage across the bulk Si resistance ($\approx 250\Omega$) is much smaller than the stress voltage increment (≥ 0.5V) and is therefore ignored in the model.

The effect of trapped holes on the anode field can be expressed as:

$$F_{an} = \frac{V_{ox}}{T_{ox}} - \frac{Q_{ot}^{+}\, x_p}{\epsilon_{ox}\, T_{ox}} + \frac{Q_{ot}^{-}\, x_n}{\epsilon_{ox}\, T_{ox}} \tag{4.16}$$

4.2.3.3 Calculated Results

Simultaneous numerical solution of equations (4.11-4.16) allows the tunnelling current to be calculated. The constants A, B, C and D in Fowler-Nordheim equation (4.11) and (4.14) are extracted from the experimental results (Fig.3.11), and other parameters are given values consistent with earlier publications. The results are shown in Fig.4.4.

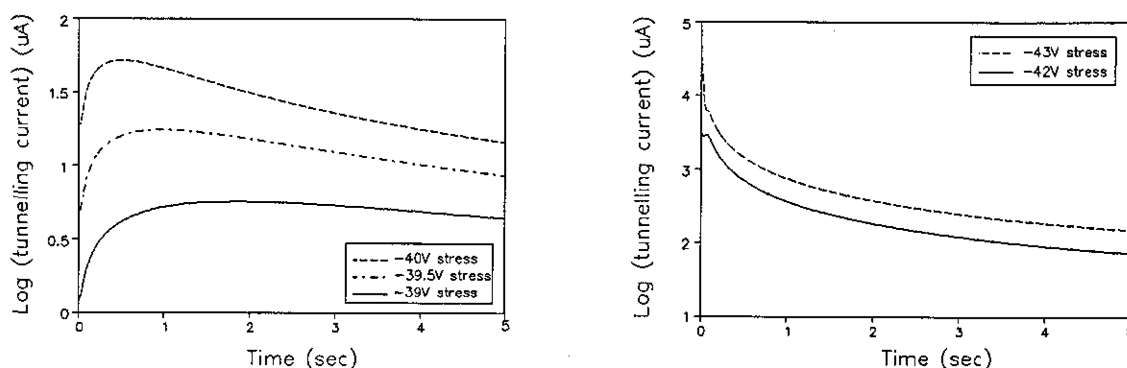


Figure 4.4: Calculated results of tunnelling current profiles for constant voltage stresses.

The results in Fig.4.2 and 4.3 can also be explained by this model. Firstly, since the density of available hole traps is fixed, the time required to fill them decreases linearly with increasing current and hence exponentially with increasing voltage. Secondly, the rapid hole-trap filling at high fields ($\sim$ 10ms for -42V stress) lies beyond the time-interval of the parametric analyzer (which was set to 100ms throughout these experiments). Hence the maximum recorded current lies within the decay-portion of the current profile, where the electron trapping compensates the trapped hole charge. The *apparent* maximum current therefore increases less rapidly with increasing field, as shown in Fig.4.2.

4.3 The Assessment of the MOS-C as An Oxide-integrity Test Vehicle

The voltage dependence of the time-to-breakdown is illustrated in Fig.4.5, which shows the average values of $\log(t_{bd})$ plotted against $1/V_{ox}$. (The error bars indicate the *standard error in the mean*, i.e. $\pm\sigma/\sqrt{}$(Number of samples)). The capacitor data are clearly consistent with a linear relationship, indicating a model of the form $\tau(F) = \tau_0 e^{\gamma/F}$ suggested by Chen et al. [3]. The transistor data may be approximately fitted to the same curve, shifted downwards by approximately half a log-cycle (i.e. a factor of three).

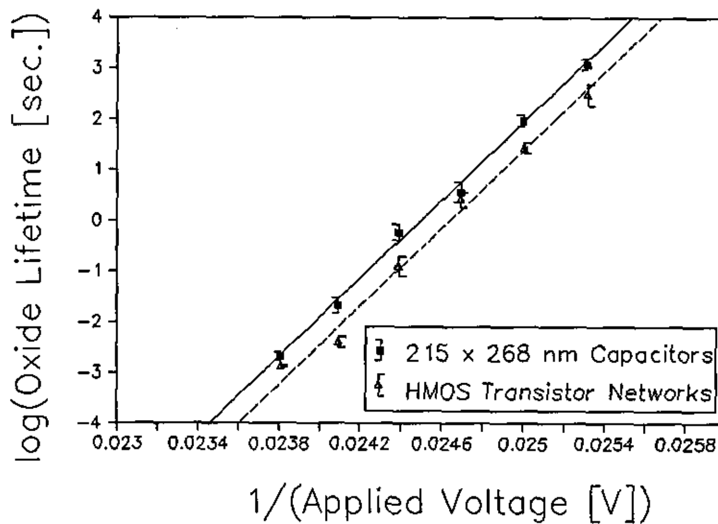


Figure 4.5: Average log (time-to-breakdown) plotted as a function of stress voltage.

From the results in Fig.4.5, MOS-C structure can be used as a test to predict the oxide lifetime of MOSFET networks, in this particular case the predicted lifetime is three times as long as the actual lifetime. However, this will not be universally true. Although the oxide area of the MOSFET network is compatible with that of the MOS-C ($3.1\cdot10^{-4}cm^2$ for MOSFET network and $5.26\cdot10^{-4}cm^2$ for MOS-C), MOSFET networks have a much larger perimeter (being composed of 34 small MOSFETs), and consequently are more likely to fail due to edge effects.

One of the aims of this study was to assess the usefulness of the MOS capacitor as a test vehicle for monitoring the oxide integrity of active MOS circuitry. This was achieved by comparing the lifetimes of transistor and capacitor structures on the same die.

Figure 4.6 compares the values of $\log(t_{bd})$ obtained from MOSFET and MOS-C structures on the same die. (Devices on any given die were always stressed at the same voltage). The broken diagonal lines indicate the results which would be expected if the failure times for a given die were equal.

Correlation does exist between the two breakdown times, but as the stress voltage is increased, a high degree of scatter occurred in the 'early' failures. Of those die which exhibited 'early' failure, only 36% suffered it in both the MOS-C and the MOSFET network. This can be easily understood by a bi-modal distribution with an extrinsic failure mechanism which is associated with defects in the oxide and located randomly with position. The full picture is not quite this clear as the tunnelling current profile shown in Fig.3.11 is accompanied by a drop in oxide voltage. This results in an apparent increase in the number of early failures (see Section 4.4).

The correlation (Fig.4.6) demonstrates the limitations of MOS-C as a test-vehicle for monitoring in-wafer variations of oxide integrity. However, with a carefully chosen low stress voltage (below most of the known activation energies for early failure) and a more compatible shape and size of the oxide, MOS-C could be used as a test vehicle of monitoring the oxide integrity of MOSFETs. This has not been studied because of the limited supply of wafers.

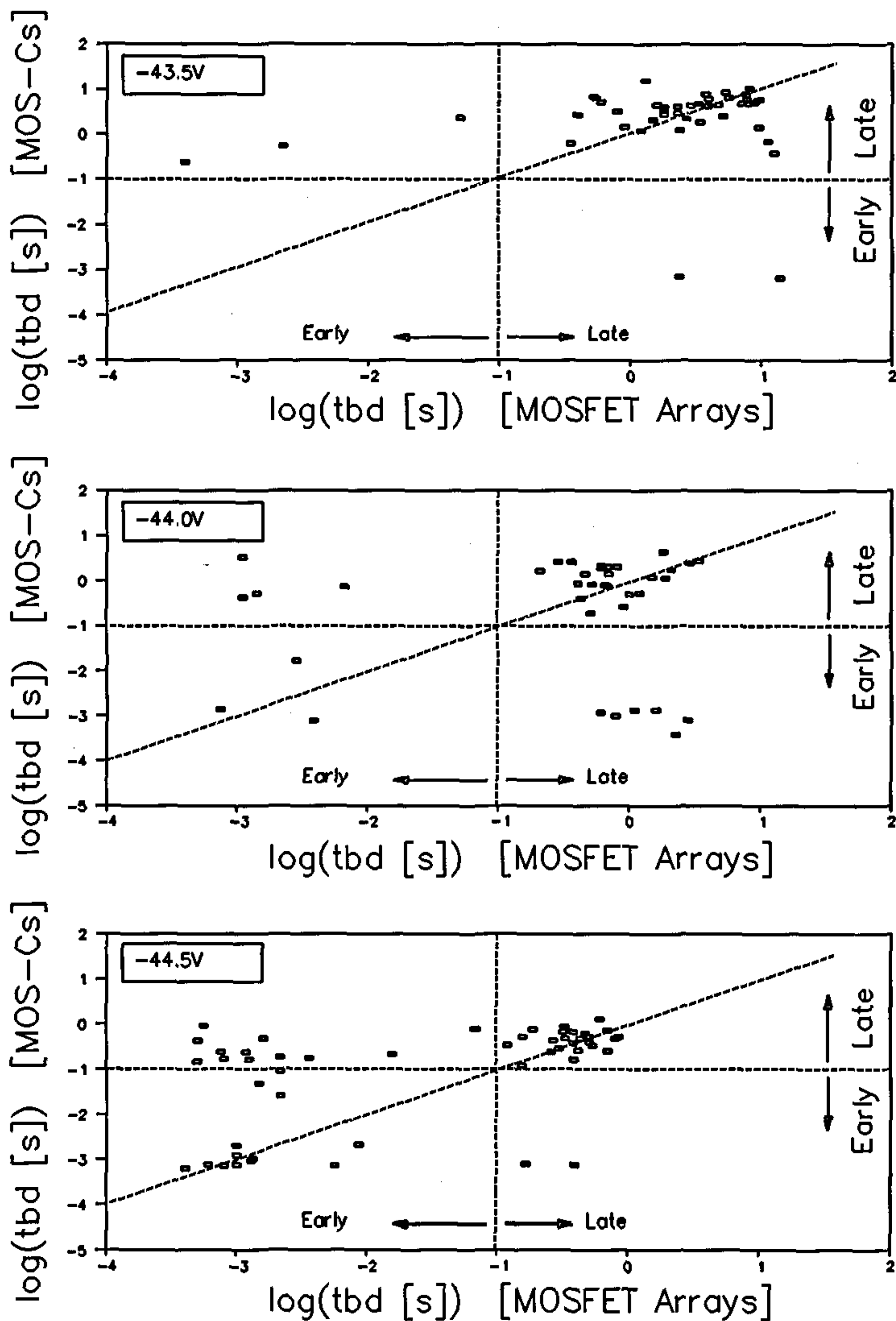


Figure 4.6: Correlation between $\log(t_{bd})$ in transistors and capacitors on the same die.

4.4 The Distributions of the Oxide Failure

As shown in Fig.3.10, the failure distributions of oxide appeared to be bi-modal, and the ratio of the numbers of 'early' and 'late' failures generally increases with the stress voltage. Bi-modal failure distributions have been observed by many investigators [17,18] and have normally been attributed to 'intrinsic' and 'extrinsic' failures. Intrinsic failures are associated with the inherent oxide properties, while extrinsic failures are caused by structural defects and/or chemical impurities. The latter occur much more rapidly than the former, and form a separate statistical distribution. It is also possible some extrinsic defects are only activated above a particular field, such that their apparent population increases with voltage.

An alternative scheme, based upon the time-dependence of the pre-breakdown tunnelling current, is presented below:

The tunnelling current may be modelled by

$$I(t) = I(0)(1+\frac{t}{t_P})^{-n}(1+\frac{t_P}{t})^{-m} \qquad (4.17)$$

where t_P is the time when the current reaches the maximum value and it is dependent upon stress voltage, m and n are constants between 0 and 1, and I(0) is the initial current.

Also the tunnelling current has to be consistent with F-N equation (4.1).

The time-dependence of the field across the oxide is determined by

$$F_{ox}(t) = \frac{V_s - I(t)R_b}{T_{ox}} \qquad (4.18)$$

where V_S is the stress voltage, R_b is the Si bulk resistance.

Oxide breakdown is caused by a continuous, non-reversible accumulation of damage inflicted during the pre-breakdown period. Breakdown occurs when the damage reaches a critical level. The rate at which oxide damage accumulates increases rapidly with the

field which is generally accepted as

$$\tau(t) = \tau_0 \exp\left(-\frac{\gamma}{F_{ox}(t)}\right) \quad (4.19)$$

in which τ represents the damage rate, $\tau_0=10^{-43}$s [3] and $\gamma=1000$MV/cm [19].

The accumulated damage can be described as

$$f(t) = \int_0^t \frac{dt}{\tau(t)} \quad (4.20)$$

Assuming that the failure distribution is lognormal, the percentage of failures can be written as

$$F[\log(t)] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \exp\left(\frac{-f^2[\log(t)]}{2}\right) d[\log(t)] \quad (4.21)$$

Calculated results are shown in Fig.4.7, and they agree reasonably well with the experimental results. In general there will be both intrinsic and extrinsic failures, however, the above analysis show that many of the early failure may in fact be due to variations in the stress voltage caused by trap filling (Section 4.2). This is important for manufacturers who probably use early failure as an indictor for oxide quality.

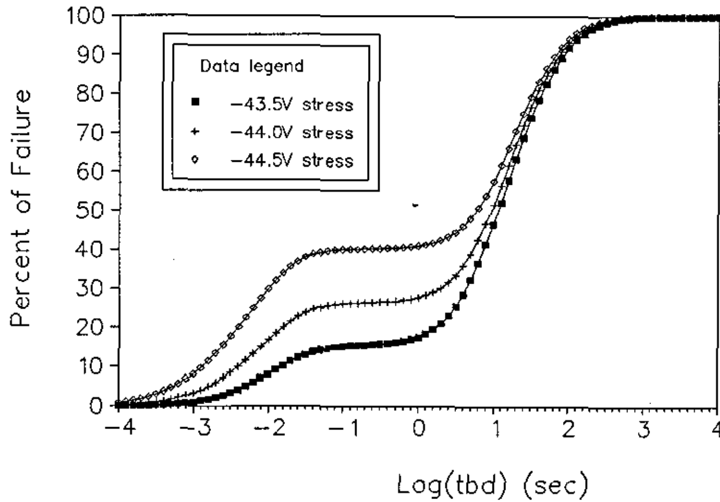


Figure 4.7: Calculated failure distributions.

The temperature acceleration TDDB distribution of Wafer 4 (see Fig.3.10) can also be explained by this model. Although all the devices on Wafer 4 were subjected to the same voltage, at higher temperature the actual voltage across the oxide was smaller, since the bulk resistance increased with temperature (see Fig.3.8), and consequently the failure distributions of high temperature have longer mean breakdown time.

4.5 Degradation of Transconductivity of MOSFET

As mentioned in Section 3.4.4, the mutual transconductance g_m of the transistor dropped nearly 50% after the gate oxide breakdown. This can be qualitatively explained by scattering theory.

The conduction band of Si has six equivalent ellipsoids located along the Δ lines (these are the (100) axes) about 85% of the way to the zone edge at X [14]. Scattering within each ellipsoid is limited to acoustic phonon and impurities, as the intravalley optical process is forbidden.

4.5.1 The Important Scattering Equations

Phonon distort the crystal lattice and create three kinds of potential energy changes: *deformation potential*, *piezoelectric potential* and *polar optic potential*.

The scattering rate for deformation potential is given by [15]:

$$\frac{1}{\tau_{tot}^{dp}} = \frac{\Xi_1^2 k_B T (2m^*)^{3/2}}{2\pi\hbar^4 \rho v_s^2} \mathscr{E}^{1/2} \qquad (4.22)$$

where Ξ_1 is deformation potential, ρ is crystal density, $v_s = \omega_q/q$ ($\hbar\omega_q$ is the phonon energy) and τ_{tot}^{dp} is the relaxation time associated with the deformation potential scattering.

Piezoelectric potential scattering occurs predominately at relatively low temperature and can be largely ignored at room temperature.

The optical scattering rate can be expressed as:

$$\frac{1}{\tau_{tot}^{opd}} = \sqrt{\frac{m^*}{2}} \frac{m^* D^2}{\pi\rho\hbar^3\omega_0} \left[N_q \sqrt{\mathscr{E}_{\bar{k}} + \hbar\omega_0} + (N_q+1)\sqrt{\mathscr{E}_{\bar{k}} - \hbar\omega_0}\, \mu_0(\mathscr{E}_{\bar{k}} - \hbar\omega_0) \right] \qquad (4.23)$$

in which D is a macroscopic deformation field, and N_q is the phonon occupation number [15, p.89].

The impurity scattering rate can also be shown to be:

$$\frac{1}{\tau_{tot}^{I}} = \frac{16\pi^2 m^* e^4 k}{\hbar^3 V_{ol} \epsilon^2 \epsilon_0^2} \int_0^1 \frac{w\,dw}{(2k^2 w^2 + q_s^2/\epsilon)^2} \qquad (4.24)$$

The rate given in eq.(4.24) is for a single impurity (also single charged). In real situations, a multiplying factor N_I (which is equal to the impurity number) should be introduced.

In the channel of a MOSFET, surface roughness scattering must also be considered. The scattering rate is determined by:

$$\frac{1}{\tau_{tot}^{sr}} = \frac{m^* \Delta^2 L^2 e^4}{2\hbar^3 \epsilon_s^2} \left[I_0(\frac{k^2L^2}{2}) - I_1(\frac{k^2L^2}{2}) \right] e^{(-\frac{k^2L^2}{2})} \qquad \mathbf{(4.25)}$$

in which Δ is the r.m.s. height of the fluctuation in the interface, L is the correlation length for the fluctuations (i.e. average distance between 'bumps' in the interface), and I_0 and I_1 are modified Bessel functions.

4.5.2 Scattering Rates and Mobility

The relationship between scattering rate and the effective mobility can be written as:

$$\mu = \frac{|e| \tau_{tot}^{all}}{m^*} \qquad \mathbf{(4.26)}$$

where $1/\tau_{tot}^{all}$ is the overall scattering rate for all important scattering mechanisms. If these scattering mechanisms are independent, then

$$\frac{1}{\tau_{tot}^{all}} = \frac{N_I}{\tau_{tot}^{I}} + \frac{1}{\tau_{tot}^{ac}} + \ldots \qquad \mathbf{(4.27)}$$

The gate oxide breakdown would be expected to change the states of the channel. Firstly, more impurity scattering centres are induced in the channel and increase the impurity scattering. Secondly, the correlation length for the fluctuations L decreases and hence the surface roughness scattering increases. Thirdly, a deformation potential increase is expected after the Si/SiO_2 interface is damaged, causing an increase in the deformation

potential scattering.

It is not clear which of these factors contributes most to the decrease in mobility. A proper discussion of events is beyond the scope of this thesis.

The total scattering rate therefore increases after the oxide breakdown. Thus the effective mobility decreases, together with the conductivity of the transistor. This effect is clearly observed in the data in Fig.3.16 and 3.24.

4.6 References

[1] Harari,E, "Dielectric Breakdown in Electrically Stressed Thin Films of Thermal SiO_2", J. Appl. Phys., **49(4)**, April, pp.2478-2489, 1978.

[2] Meyer,W.K., Crook,D.L., "Model for Oxide Wearout due to Charge Trapping", IEEE/IRPS, pp.242-247, 1983.

[3] Chen,I-C., Holland,S.E., Hu,C, "Electrical Breakdown in Thin Gate and Tunnelling Oxides", IEEE Trans. on Electron Devices, **ED-32(2)**, February, pp.413-422, 1985.

[4] Tunnicliffe,M.J., Dwyer,V.M., "Monitoring the Integrity of MOS Gate Oxides: A Preliminary Study", presented at EPMS'93 Conference, Sheffield Hallam University, Sheffield, and accepted for the International Journal of Electronics, 1993.

[5] Lenzlinger,M, Snow,E.H., "Fowler-Nordheim Tunnelling into Thermally Grown SiO_2", J.Appl.Phys., **40(1)**, pp.278-283, 1969.

[6] Young,D.R., "Characterization of electron traps in SiO_2 as influenced by processing parameters", J. Appl. Phys., **52(6)**, pp.4090-4094, June, 1981.

[7] Young,D.R., Irene,E.A., DiMaria,D.,J., Dekeersmaecker,R.F., "Electron Trapping in SiO_2 at 295 and 77 Degrees K", J. Appl. Phys., **50(10)**, pp.6366-6372, 1979.

[8] Itsumi,M, "Positive and negative charging of thermally grown SiO_2 induced by Fowler-Nordheim Emission", J. Appl. Phys., **52(5)**, pp.3491-3497, May 1981.

[9] Liang,M-S, Hu,C, "Electron Trapping in Very Thin Thermal Silicon Dioxides", IEEE/IEDM Tech. Dig., pp.396-399, 1981.

[10] Distefano,T.H., Shatzkes,M, "Impact Ionization Model for Dielectric Instability and Breakdown", Appl. Phys. Lett., **25**, pp.685-687, 1974.

[11] Klein,N, Solomon,P, "Current Runaway in Insulators Affected by Impact Ionization and Recombination", J. Appl. Phys., **47**, pp.4364-4372, 1976.

[12] Jenq,C.S., Ranganath,T.R., Huang,C.H., Jones,H.S., Chang,T.T.L., "High-field Generation of Electron Traps and Charge Trapping in Ultra-thin SiO_2", IEEE/IEDM Tech. Digest, pp.388-391,

1981.

[13] Lee,J.C., Chen,I-C, Hu,C, "Modelling and Characterization of Gate Oxide Reliability", IEEE Trans. on Electron Devices, **35(12)**, pp.2268-2278, December, 1988.

[14] Ferry,D.K., "Chapter 7: Electron-Phonon Interaction", Semiconductors, Macmillan Publishing Company, 1991.

[15] Hess,K, "Chapter 8: Scattering Theory", Advanced Theory of Semiconductor Devices, Prentice-Hall International Editions, 1988.

[16] Dong,L, Tunnicliffe,M.J., Dwyer,V.M., "Pre-Breakdown Charge Trapping in High Field Stressed Oxides", 5th ESREF, Glasgow, 1994.

[17] Wolters,D.R., Van Der Schoot, J.J., "Dielectric Breakdown in MOS Devices", Phillips J. Res., pp.115-193, 1985.

[18] Vollertsen,R.-P., "Statistical Modelling of Time Dependent Oxide Breakdown Distributions", Microelectron. Reliab., **33(11/12)**, pp.1665-1677, 1993.

[19] Tunnicliffe,M.J., Dwyer,V.M., Campbell,D.S., "Experimental and Theoretical Studies of EOS/ESD Oxide Breakdown in Unprotected MOS Structures", Proc. 12th. EOS/ESD Symposium, pp.162-168, 1990.

CHAPTER 5

PSPICE Simulation

5.1 Introduction

This chapter is devoted to studying the effects of Gate Oxide Short (GOS) upon a simple circuit, and to understand the operation of damaged circuits with help of computer simulation. Finally a possible mechanism for detecting the presence of a GOS is proposed.

5.2 MOSFET Theory

5.2.1 Introduction

The Metal-Oxide-Semiconductor Field Effect Transistor (MOSFET) is a three-terminal device in which the current between two terminals is controlled by a control input applied to the third. Unlike the bipolar transistor, however, field-effect devices are controlled by a voltage rather than by a current. They are unipolar device, i.e., the current involves only majority carriers.

MOSFETs are characterised by a high input impendence, since the control voltage is applied across an insulator. Hence, these devices are particularly well suited for controlled switching between a conducting state and a nonconducting state, and are therefore useful in digital circuits. Millions of MOSFETs are commonly used together in today's electronic equipment.

5.2.2 Basic Operation of MOSFET

The basic MOS transistor is illustrated in Fig.5.1. This particular device is nMOS

enhancement mode, in which an n-type channel formed on a p-type Si substrate. The n^+ source and drain regions are diffused or implanted into the relatively lightly doped p-type substrate, and a thin oxide layer separates the Al metal gate from the Si surface. No current flows from drain to source without a conducting n channel between them, since the drain-substrate-source combination includes oppositely directed p-n junctions in series.

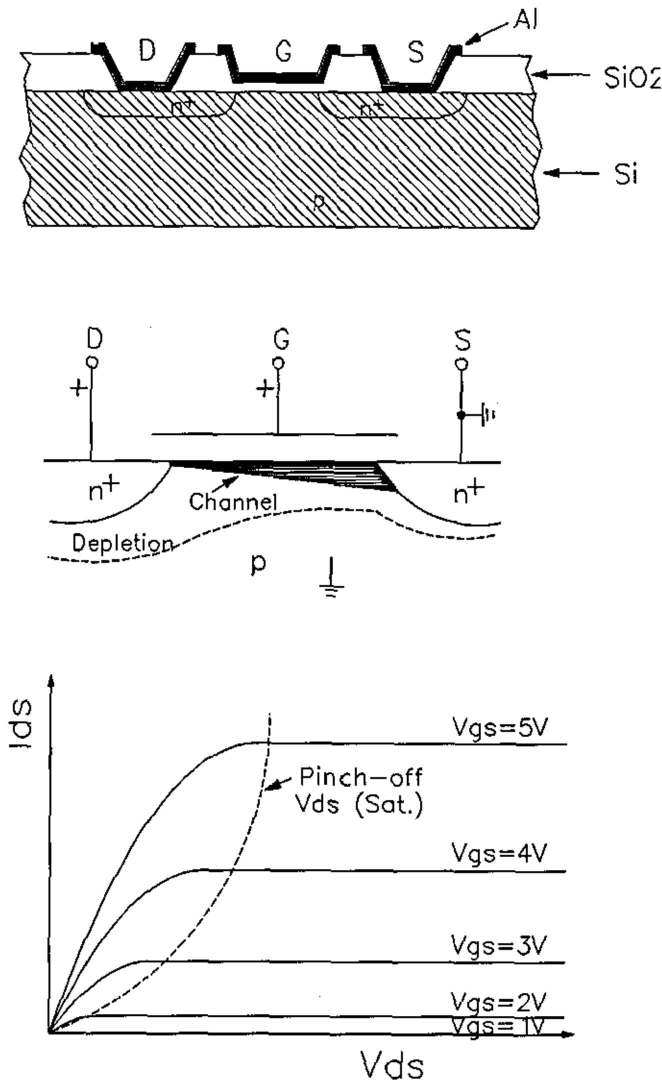


Figure 5.1: An enhancement-type n-channel MOS transistor: cross section of the device, schematic illustration of the induced n-channel and the depletion region near pinch-off and drain current-voltage characteristic as a function of gate voltage.

When a positive voltage is applied to the gate relative to the substrate (which is connected to the source in this case), positive charges are in effect deposited on the gate metal. In response, negative charges are induced in the underlying Si, by the formation of a depletion region and a thin surface 'inversion' region containing mobile electrons. These induced electrons form the channel of the FET, and allow current to flow from drain to source. As Fig.5.1 suggests, the effect of the gate voltage is to vary the conductance of this induced channel for low drain-to-source voltages. For a given value of V_{GS} there will be some drain voltage V_{DS} for which the channel 'pinches off' at the drain. Beyond this point the current becomes saturated, and it remains essentially constant.

In general, the positive gate voltage of an n-channel device must be larger than some value V_T before a conducting channel is induced (such as that shown in Fig.5.1), however, some n-channel devices have an implanted channel with zero gate voltage, and in fact a negative gate voltage is required to turn the device off. Such a 'normally on' device is called a **depletion-mode** transistor while the more common 'normally off' device is called a **enhancement-mode** transistor.

5.2.3 A Mathematical Model of the MOSFET

5.2.3.1 Threshold Voltage V_T

An important parameter of a MOS transistor is the threshold voltage V_T, which is the minimum gate voltage required to induce the channel. The zero substrate bias threshold voltage V_{T0} (i.e. V_{BS}=0) can be expressed as Equation 5.1 below

$$V_{T0} = \Phi_{ms} - \frac{Q_i}{C_i} - \frac{Q_d}{C_i} + 2\phi_F \qquad \textbf{(5.1)}$$

where Φ_{ms} is the effective work-function difference between gate and substrate, Q_i is the total space-charge in the insulator, ϕ_F is the Fermi-potential in the bulk Si and Q_d is the total depletion layer charge when the surface potential equals to $2\phi_F$.

Thus the voltage required to create strong inversion must be large enough to first

achieve the flat band condition (Φ_{ms} and Q_i/C_i terms), then accommodate the charge in the depletion region (Q_d/C_i), and finally to induce the inverted region ($2\phi_F$). Eq.5.1 accounts for the dominant threshold voltage effects in typical MOS devices.

For the case of $V_{BS}\neq0$, threshold voltage becomes

$$V_T = \Phi_{ms} - \frac{Q_i}{C_i} + \left[\frac{-2Nq\epsilon_s(2\phi_F - V_{BS})}{C_i^2}\right]^{\frac{1}{2}} + 2\phi_F \qquad (5.2)$$

in which the third term is the voltage required to accommodate the charge in the depletion region (Q_d/C_i).

5.2.3.2 Drain Current

It is impossible to obtain an analytical solution for MOSFET drain characteristics. However, based on the **depletion approximation**, an approximate expression can be obtained:

The mobile charge in semiconductor, i.e. inversion charge can be expressed as

$$Q_n = -C_i\left[V_{GS} - \left(V_{FB} + \phi_s - \frac{Q_d}{C_i}\right)\right] \qquad (5.3)$$

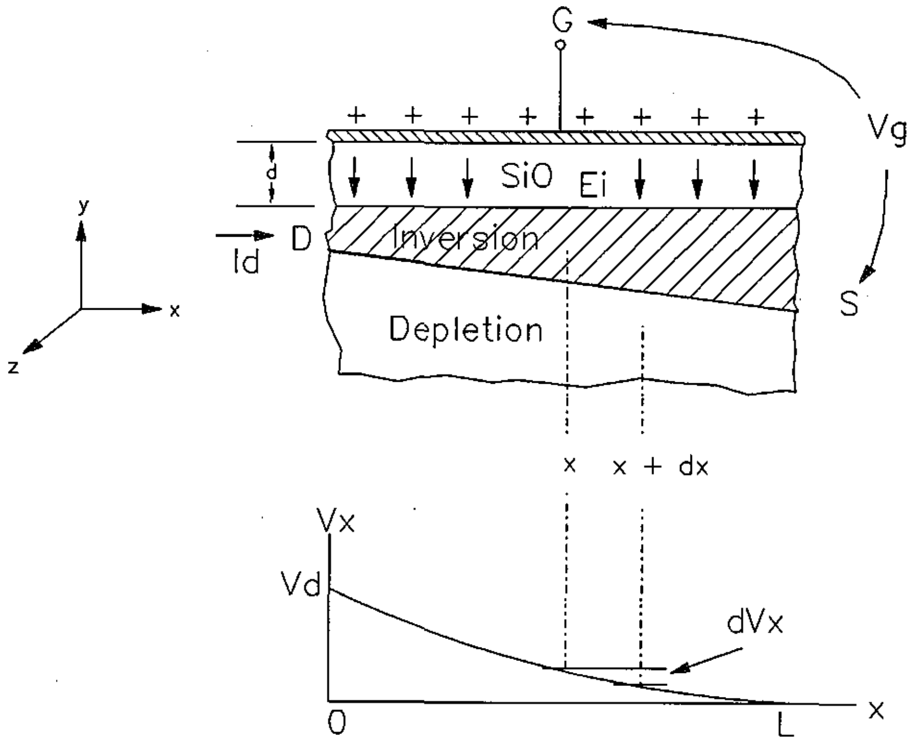


Figure 5.2: Schematic view of the n-channel region of a MOS transistor under bias below pinch-off, and the variation of voltage V_x along the conducting channel.

With a voltage V_{DS} applied, there is a voltage drop V_x from each point x in the channel to the source (Fig.5.2). Thus the surface potential $\phi_s(x)$ is that required to achieve strong inversion ($2\phi_F$) plus the voltage V_x, and the mobile charge in the channel at point x is:

$$Q_n = -C_i\left[V_{GS} - V_{FB} - 2\Phi_F - V_x - \frac{1}{C_i}\sqrt{2q\epsilon_s N_a(2\Phi_F + V_x)}\right] \qquad (5.4)$$

The conductance of the differential element dx is $\mu_n^* Q_n(x)Z/dx$, where Z is the channel width and μ_n^* is the effective mobility of carriers in the inversion and independent of field. At point x:

$$I_{DS}\, dx = \mu_n^* Z Q_n(x)\, dV_x \tag{5.5}$$

Integrating from drain to source,

$$I_{DS} = \frac{\mu_n^* Z C_i}{L}\left((V_{GS}-V_{FB}-2\phi_F-\frac{1}{2}V_{DS})V_{DS} - \frac{2}{3}\frac{\sqrt{2\epsilon_s q N_a}}{C_i}\,[(V_{DS}+2\phi_F)^{\frac{3}{2}} - (2\phi_F)^{\frac{3}{2}}]\right) \tag{5.6}$$

The drain characteristics resulting from the above equations are shown in Fig.5.1. For low V_{DS}, as the drain voltage is increased, the voltage across the oxide decreases near the drain, and Q_n becomes smaller there. As a result the channel becomes pinched off at the drain end, and the current saturated. The saturation condition is approximately given by

$$V_{DS}(sat.) \simeq V_{GS} - V_T \tag{5.7}$$

The drain current at saturation remains essentially constant for larger values of drain voltage. Substituting Eq.(5.7) into Eq.(5.6):

$$I_{DS}(sat.) \simeq \frac{1}{2}\mu_n^* C_i \frac{Z}{L}(V_{GS} - V_T)^2 = \frac{Z}{2L}\mu_n^* C_i V_{DS}^2(sat.) \tag{5.8}$$

5.2.3.3 Transconductance and Conductance in MOSFET

The gate transconductance, which is defined by:

$$g_m = \left.\frac{\partial I_{DS}}{\partial V_{GS}}\right|_{V_{DS}} \quad (5.9)$$

can be evaluated from Eq.(5.6). For $|V_{DS}| \leq |V_{GS}-V_T|$ it is given by

$$g_m = \frac{C_i \mu_n^* Z}{L} V_{DS} \quad (5.10)$$

The transconductance in the saturation range ($|V_{DS}| \geq |V_{GS}-V_T|$) can be obtained approximately by differentiating Eq.(5.8) with respect to the gate voltage:

$$g_m(sat.) = \frac{\partial I_{DS}(sat.)}{\partial V_{GS}} \simeq \frac{Z}{L} \mu_n^* C_i (V_{GS} - V_T) \quad (5.11)$$

The drain-source conductance is defined by

$$g_{ds} = \left.\frac{\partial I_{DS}}{\partial V_{DS}}\right|_{V_{GS}} \quad (5.12)$$

and

$$g_{ds} = \frac{C_i \mu_n^* Z}{L} \left(V_{GS} - V_{FB} - 2\phi_F - V_{DS} - \left[\frac{2Nq\epsilon_s}{C_i^2} (V_{DS} + 2\phi_F) \right]^{\frac{1}{2}} \right) \quad (5.13)$$

For zero drain-source voltage, Eq.(5.13) reduces to

$$g_{ds}\Big|_{V_{DS}=0} = \frac{C_i \mu_n^* Z}{L}(V_{GS} - V_T) \qquad \textbf{(5.14)}$$

These equations represent an ideal 'first-order' model of a MOSFET. More advanced 'second-order' models are available in the PSPICE library described later.

5.3 MOSFET Models in PSPICE

5.3.1 Introduction to PSPICE Simulator

PSPICE is a member of the SPICE (Simulation Program with Integrated Circuits Emphasis) family of software packages. The programs in this family come from the SPICE2 circuit simulation program developed at the University of California at Berkeley during the early 1970's. PSPICE, the first SPICE-based simulator available on the IBM-PC, was commercially released in 1984.

PSPICE is used to simulate and test circuits containing both analog and digital components, eliminating the need to build hardware prototypes. The temporal response to different inputs, of varying frequency, together with other details such as noise can be conveniently calculated for any specified circuit.

The input circuit file normally includes three parts: device descriptions, model definitions and analysis commands. Each device in the circuit is represented in the input file by one line and the format is:

1) the device name,
2) 2 or more nodes,
3) a model name (not all devices have this),
4) 0 or more parameter values

Many devices used models to assign values to the various parameters which describe the device. The .MODEL statements have the form

```
.MODEL name name type (parameter=value parameter=value ...)
```

Analysis commands specify the analyses which PSPICE will perform on the described circuit. All the commands start with a '.' note. Various analyses are available in PSPICE including: DC sweep of an input voltage (.DC), frequency response calculation (.AC), DC sensitivity calculation (.SENS), transient response (.TRAN) etc. An example of a analysis command is:

```
.DC Vgs -5V 5V 0.1V
```

In PSPICE, a nested sweep is available as

```
.DC Vds 0 10 V 0.1V Vgs 0 5V 1V
```

In this statement, the first sweep would be the 'inner' loop: the entire first sweep would be done for each value of the second sweep. Thus the I_{ds} - V_{ds} characteristics could be printed out after this command.

Running PSPICE is straightforward. Execute PSPICE by using the following command format:

```
PSPICE <input file> [<output file> [<Probe data-file>]] [/B] [/PS=<com port>]
```

When the running is complete, PSPICE generates an output file with all the simulation results in it. There are several ways to output the simulation results: direct output, print tables and plots. The print tables could be saved as an ASCII file and imported to LOTUS 1-2-3 spreadsheet or Freelance for further data analysis or graphical presentation.

5.3.2 MOSFET Models

As shown in Fig.5.3, the MOSFET is modelled as an intrinsic MOSFET with ohmic resistances in series with the drain, source, gate and bulk (substrate). There is also a shunt resistance (RDS) in parallel with the drain-source channel. The various capacitors (C_{gb}, C_{gd}, etc.) represent the capacitances associated with oxide, depletion layers and packaging.

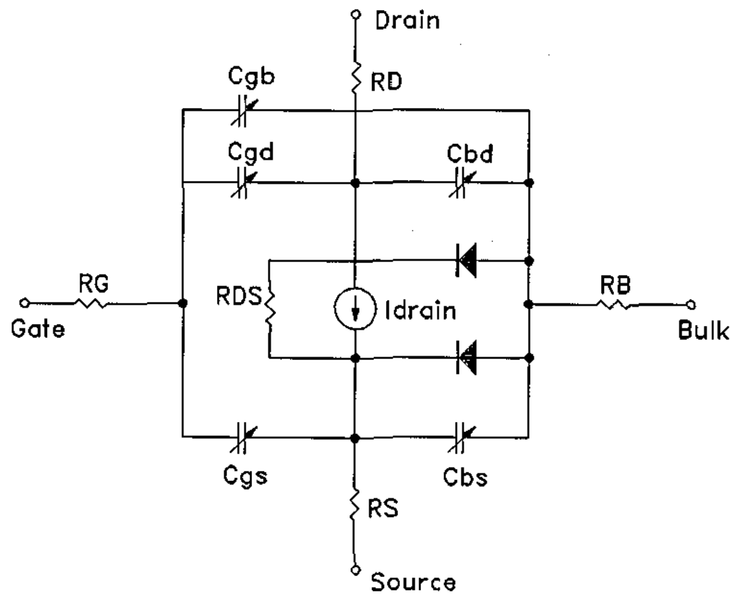


Figure 5.3: MOSFET model in PSPICE.

PSPICE provides a choice of four MOSFET device models, which differ in the formulation of the I-V characteristic.

Level 1 model is based on an model which developed first by H.Shichman and D.A.Hodges [4]. This first order model is particularly useful for the analysis of monolithic integrated switching circuit. The model is equally applicable to both enhancement-mode and depletion-mode devices, whether p-channel or n-channel. Simulation results are valid for any combination of terminal voltages that do not cause electrical breakdown in the actual device.

The DC equations for the devices are the same as described in Section 5.2.3. Average values of device capacitances (rather than voltage-dependent representations) are used in the gate-channel capacitors C_{gs},C_{gd} and C_{gb}.

Second-order effects [3] become important when the channel lengths are below 5μm, especially if the supply voltages are kept at the 5V level. An analytical one-dimensional model which incorporates most of the second-order effects of small-size devices is available as the level 2 MOS model in PSPICE [5].

The threshold voltage can be expressed as

$$V_T = V_{BIN} + \gamma_S \sqrt{2\phi_F - V_{BS}} \tag{5.15}$$

where

$$V_{BIN} = V_{FB} + 2\phi_F + DELTA\frac{\pi\epsilon_s}{4C_iW}(2\phi_F - V_{BS}) \tag{5.16}$$

is the corrected built-in voltage for narrow channel [5], and γ_S is the corrected GAMMA for short-channel with static drain-to-gate feedback [5].

$$\gamma_S = GAMMA(1 - \alpha_S - \alpha_D) \tag{5.17}$$

where α_D and α_S are the correction factors for the depletion charge at the drain and source respectively. DELTA and GAMMA are parameters of the model in PSPICE.

The drain current of the level 2 model is slightly modified and includes the effect of the bulk charge and charge-oriented models based on the actual distribution of charge in the MOS structure and its conservation is used for the capacitances.

A small geometry MOSFET, defined as a transistor with $L \leq 2\mu m$ and $W \leq 2\mu m$ is characterized by the following features: threshold voltage sensitivity to the length and the width of the device due to the two-dimensional nature of potential distribution; threshold voltage sensitivity to the drain voltage due to the drain induced barrier lowering; relaxed transition between linear and saturation regions, and reduced saturation voltage/current due to the velocity saturation of hot electrons.

Level 3 MOS model has been developed to address the above features and the computational efficiency.

An expression for V_T which sums up the above features is formulated as

$$V_T = V_{FB} + 2\phi_F - \sigma V_{DS} + \gamma F_S\sqrt{2\phi_F - V_{BS}} + F_N(2\phi_F - V_{BS}) \qquad \textbf{(5.18)}$$

where σ is the coefficient of static feedback and

$$\sigma = ETA\frac{\Omega}{C_i L^3} \qquad \textbf{(5.19)}$$

and Ω is the empirical constant which is $8.15 \cdot 10^{-22}$ (F·m) and F_S is the correction factor of short channel effect.

Level 4 MOS model parameters are all obtained from process characterisation and can be generated automatically [6].

5.4 Simulation Results for MOSFETs with Gate Oxide Short

5.4.1 Models for MOSFET with Gate Oxide Short (GOS)

Two different equivalent circuits for the MOSFETs with GOS have been developed in the past few years. One is proposed by J.A.Segura et al. [7] and shown in Fig.5.4.

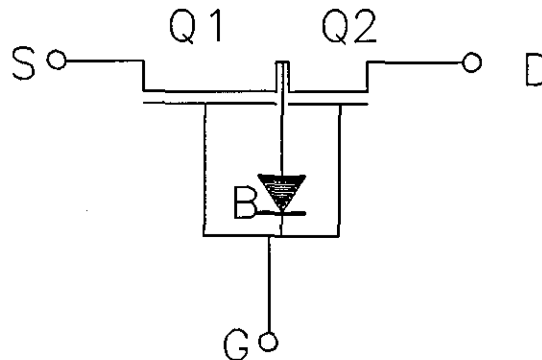


Figure 5.4: A circuit level model for Gate Oxide Short defects in the channel zone in an n-channel MOSFET, proposed by J.A.Segura et al.

In this approach (hereafter referred to as model 1), the GOS is represented by a rectifying barrier, B, between the gate and the channel. The barrier B is characterised by a threshold voltage, a breakdown voltage and forward and breakdown resistances (V_γ, V_{BR}, R_F, R_{BR}). The PSPICE description for this GOS-fault nMOS device is:

```
.SUBCKT Faulty_NMOS 3 1 0 0
MOSN1 2 1 0 0 NMOS
MOSN2 3 1 2 0 NMOS
DIODE 2 1 D
.ENDS SUBCKT
```

M.Syrzycki presented another approach (model 2) for both n-channel and p-channel MOSFETs with GOS [8]. In the case of n-channel MOS transistor, the gate oxide short structure is shown in Fig.5.5(a) and may be represented by a model shown in Fig.5.5(b). The conducting path between the gate and the channel is represented by the gate oxide short resistor R_{gos}. The junction between the n^+ spot and the substrate is modeled by the diode D_1, and the resistors R_1, R_2, R_3, R_4 represent the resistance of the n^+ spot in the direction parallel to the drain current flow. The latter resistors model the size and the net doping level of the n^+ spot. They may be of a significant value for relatively large spots of missing gate oxide, although they may also be negligible for most tiny GOS caused by the gate oxide breakdown.

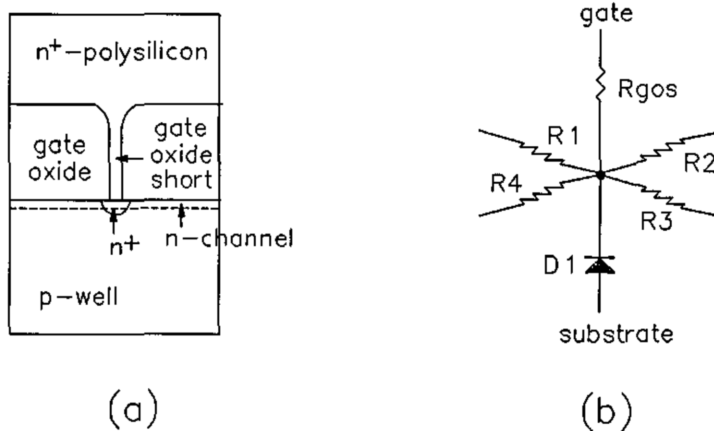
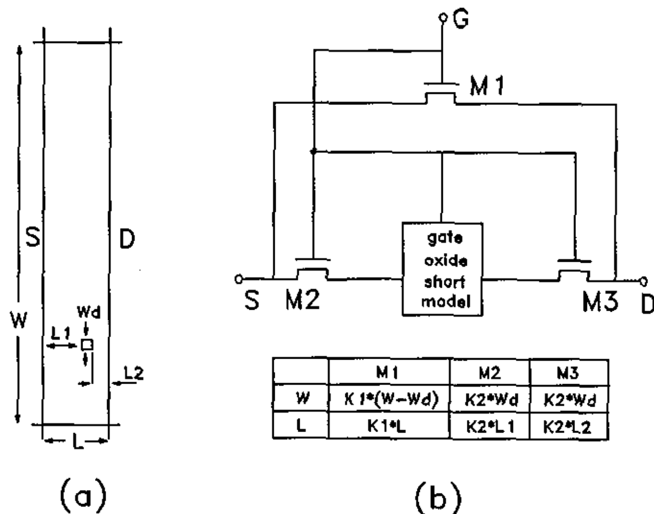


Figure 5.5: The structure of the gate oxide short in: (a) an n-channel MOS transistor and (b) the model of this gate oxide short.

If one assumes a simplified lumped-element model of the MOSFET, a MOS transistor with rectangular gate and W/L >> 1 which contains a GOS can be represented by the equivalent circuit of Fig.5.6.



	M1	M2	M3
W	K1*(W-Wd)	K2*Wd	K2*Wd
L	K1*L	K2*L1	K2*L2

Figure 5.6: A simplified lumped-element model of a rectangular gate MOSFET with gate oxide short. (K1 and K2 are scaling constants which take into account short-channel effects in the defective transistors, if necessary.)

Combining the gate oxide short structure together with the simplified lumped-element model, an n-channel MOSFET with GOS can be modeled by the circuit shown in Fig.5.7.

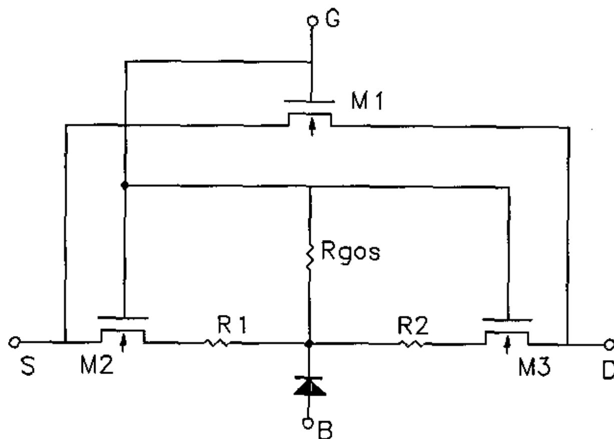


Figure 5.7: Model of an n-channel MOS transistor with gate oxide short.

Using the same procedure as for n-channel MOSFET above, the model for p-channel MOS transistor can be derived (see Fig.5.8).

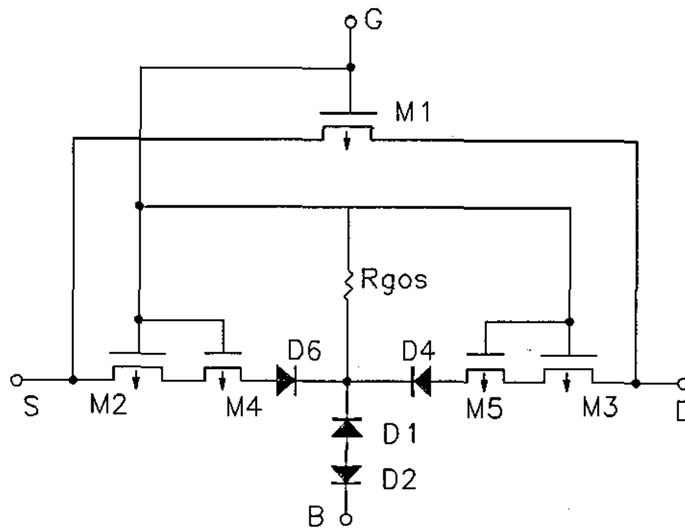


Figure 5.8: Model of the p-channel MOS transistor with gate oxide short.

5.4.2 Simulation Results for Transistors with GOS

Simulations were performed for both models described above. Three circuits were described in each PSPICE input file, the first of which represented an undamaged MOSFET, whose parameters were selected to fit the characteristics of the MOSFET prior to breakdown. The second circuit represented a MOSFET with a GOS modeled by Syrzycki's model. The third and final circuit described a MOSFET with a GOS modeled by Segura's model.

Initially, I_g - V_g characteristics of the three MOSFET models were simulated. The parameters of the devices in the equivalent circuits were chosen to fit the experimental results obtained from MOSFETs in Chapter 3 (Fig.3.22). The simulation results were produced by linearly sweeping the gate voltage V_g and recording the gate current I_g.

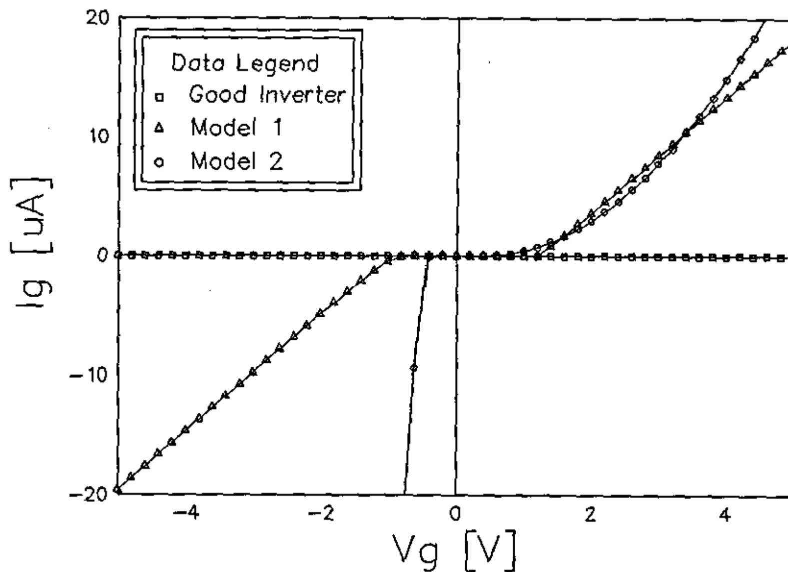


Figure 5.9: Simulation results of I_g - V_g for a good MOSFET, a MOSFET with a GOS modeled by the Segura model (model 1), a MOSFET with a GOS modeled by the Syrzycki model (model 2).

For model 1, the reverse breakdown voltage and series resistance were crucial parameters which defined the shape of gate leakage characteristic. A 200KΩ resistor was inserted in series with the intrinsic diode in order to produce good agreement with the experimental positive gate current. Although this required a sacrifice in the accuracy of the negative current characteristics, such agreement was not important in this case. The reverse breakdown voltage of the diode was set to 1V.

For model 2, the threshold voltages of M2 and M3 and the value of R_{gos} were sensitive to the shape of the gate current characteristic. A very good agreement could be reached by carefully choosing these parameters. The three simulated I_g - V_g curves are all shown in Fig.5.9.

Next, the I_{ds} - V_{ds} characteristics were simulated in PSPICE using the same parameter values determined in the former simulation. However, in order to agree well with the experimental results, the transconductance coefficient **KP** was reduced from $40 \cdot 10^{-6}$ (amp/volt^2) to $17 \cdot 10^{-6}$ (amp/volt^2) during breakdown. This is an important effect which has been explained in Chapter 4. The sizes of M2, M3 in model 2 and Q1, Q2 in model 1, all of which were determined by the position and the size of the short, were found to

be sensitive to the intersection of the drain current curves.

Simulation results of the good MOSFET are shown in Fig.5.10, the results for MOSFET with GOS are shown in Fig.5.11 (left: model 1, right: model 2). The source and substrate electrodes of the three MOSFETs were grounded in this simulation.

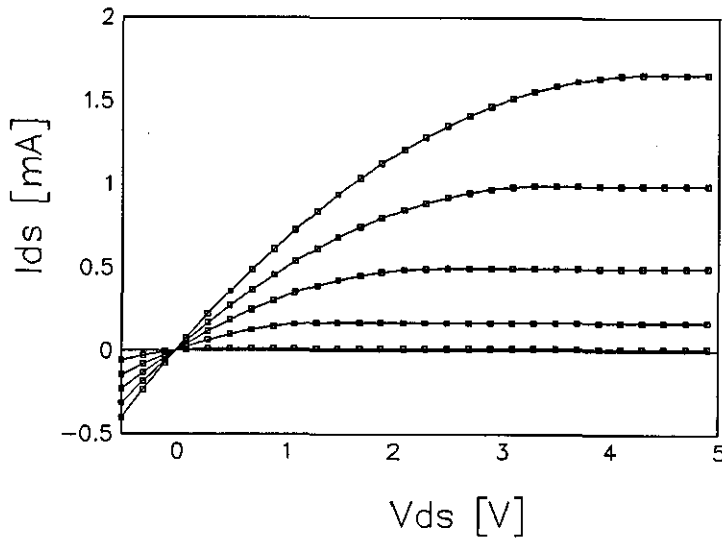


Figure 5.10: Simulation results of I_{ds} - V_{ds} for a good MOSFET.

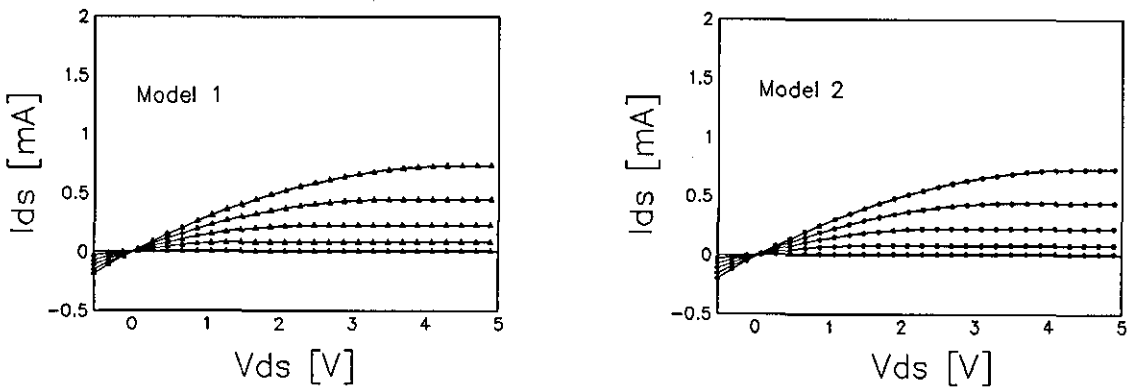


Figure 5.11: Simulation results of I_{ds} - V_{ds} for a MOSFET with GOS.

5.5 Simulation Results of the nMOS Inverter with GOS

5.5.1 Introduction to Inverter

The inverter is amongst the simplest of all logic circuit, and usually consists of two transistors. The pull-up transistor supplies charge to the load capacitor when the output goes from '0' to '1', while the pull-down transistor draws the charge away when the output goes from '1' to '0'. If both of the transistors are n-channel, the circuit is an nMOS inverter, while if both were p-channel it would be called pMOS inverter. A CMOS (Complementary MOS) inverter has one n-channel transistor and one p-channel transistor.

Since the inverter is one of the most commonly used logics, forming the basis of a large number of circuits, it is important to understand how its parameters would drift as a result of breakdown in one of the transistor's gate oxides.

5.5.2 Delay and Power

Gate propagation delays are produced because the voltage across the capacitance does not rise or fall instantaneously. It will take a period of time $\mathbf{t_c}$ (**charge delay**) or $\mathbf{t_d}$ (**discharge delay**) for the voltage across a capacitor to rise or fall respectively to the required voltage. For a simple RC series circuit, the quantity RC (which has the dimension of time) is called the **time constant** and is equal to the time for capacitor to charge to 63% of its maximum value.

MOS networks are more complicated than simple RC circuits. Firstly, the resistance of a MOS transistor is a nonlinear function of its terminal voltages, the type of the transistor (p-, n-, enhancement or depletion), and the context of its use (pull-up, pull-down, or pass transistor). Secondly, static and dynamic values of the resistances need to be distinguished in order to perform precise calculations. Thirdly, capacitances in MOS circuits also depend nonlinearly upon the terminal voltages.

However, the relationship in Eq.(5.20) is always true for the case of an nMOS inverter circuit low-to-high transition shown in Fig.5.12.

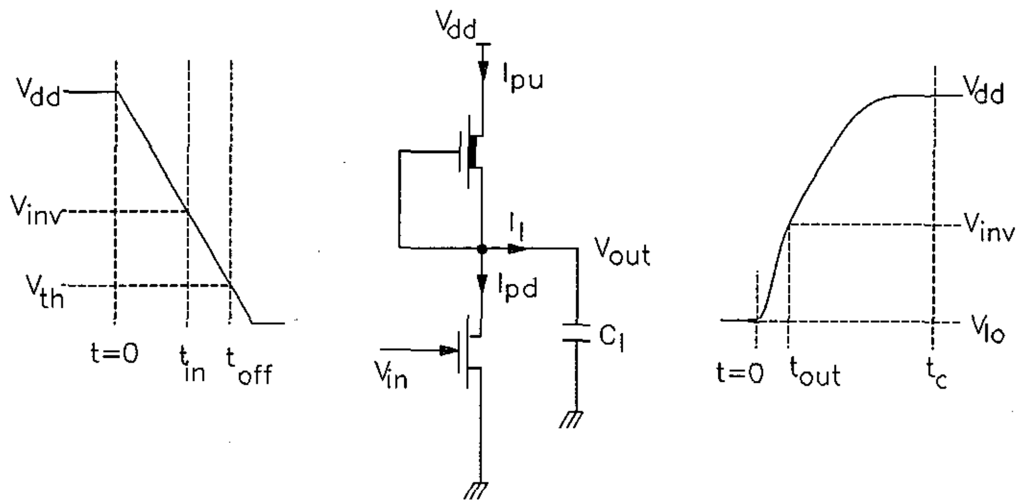


Figure 5.12: Low-to-high transition in an nMOS inverter circuit.

$$C_l(V_{dd} - V_{lo}) = \int_0^{t_c} I_l \, dt \tag{5.20}$$

where $I_l = I_{pu} - I_{pd}$.

By making several simplifying assumptions [9], the charge delay time t_c can be derived as

$$t_c = C_l\left(\frac{V_{dd} - V_{lo}}{I_{max}}\right) + \frac{1}{2}(t_{in} + t_{off}) \tag{5.21}$$

where t_{out} is the time when the output voltage V_{out} of the circuit crosses a predefined output threshold voltage V_{inv}, and similarly t_{in} is the time when the input voltage crosses V_{inv}. V_{inv} is defined as the output voltage of an inverter when the output and input of the inverter are connected together, i.e., $V_{out} = V_{in} = V_{inv}$.

Similarly, the discharge delay t_d for high-to-low transition can be written as

$$t_d = 2C_l\left[\frac{V_{dd} - V_{lo}}{\beta_d[(V_{dd} - V_{Td} - V_{inv}/2)V_{inv} - (V_{dd} - V_{Td} - V_{lo}/2)V_{lo}]}\right] \quad (5.22)$$

where V_{lo} is the low (logic '0') output voltage, $\beta_d = (\mu\epsilon_{ox}/T_{ox})(W/L)$ for the pull-down transistor, and V_{Td} is the threshold voltage of the pull-down transistor.

Very simple expressions for time delays in a CMOS inverter have been derived [9] as

$$t_d = \frac{8L^2V_{dd}}{\mu_n(V_{dd} - V_{Tn})^2} \quad (5.23)$$

and

$$t_c = \frac{8L^2V_{dd}}{\mu_p(-V_{dd} + |V_{Tp}|)^2} \quad (5.24)$$

Power consumption is another important parameter in integrated circuits. In an nMOS inverter, the instantaneous power is $V_{dd}\cdot I_{ds}$, where I_{ds} may be found by applying the appropriate drain current equation to the pull-up transistor. In an nMOS inverter, there is no significant current drawn from the supply when the input is low (logic '0'), since the pull-down transistor is off. However, a sizeable current flows in the inverter when the input is high (logic '1') since both transistors are on. If a depletion MOSFET is employed as pull-up transistor, the static current draw from power is $I_{ds} = \beta_u\cdot(-V_{Tu})^2$, and hence the total static power in this nMOS inverter is $\beta_u\cdot(-V_{Tu})^2\cdot V_{dd}$. Dynamic power is more complicated since the drain current of the pull-up transistor is a non-linear function of the input voltage.

5.5.3 Simulation Results of nMOS Inverter

Simulations of an nMOS inverter similar to that used in the experiments were performed. The dimensions were set to the nominal values of the real devices, while the other parameters for the good devices were extracted from the experimental results (see Fig.3.15 & 3.16). The parameters for the transistor with GOS were the same as they were in the discrete transistor simulations.

The DC simulation results are shown in Fig.5.13.

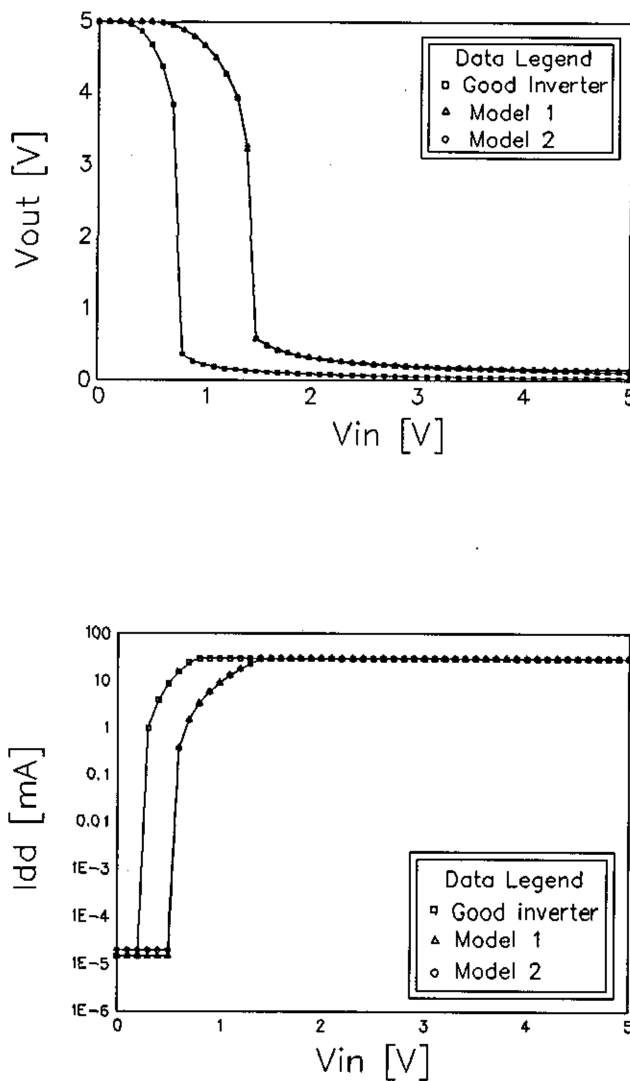


Figure 5.13: DC simulation results for the nMOS inverter.

If these curves are compared with the experimental voltage transfer characteristics in Chapter 3 (Fig.3.19), a good correlation can be seen to exist.

Transient characteristics of the inverter were examined by applying a square waveform to the inputs of the inverter and the resulting outputs are shown in Fig.5.14.

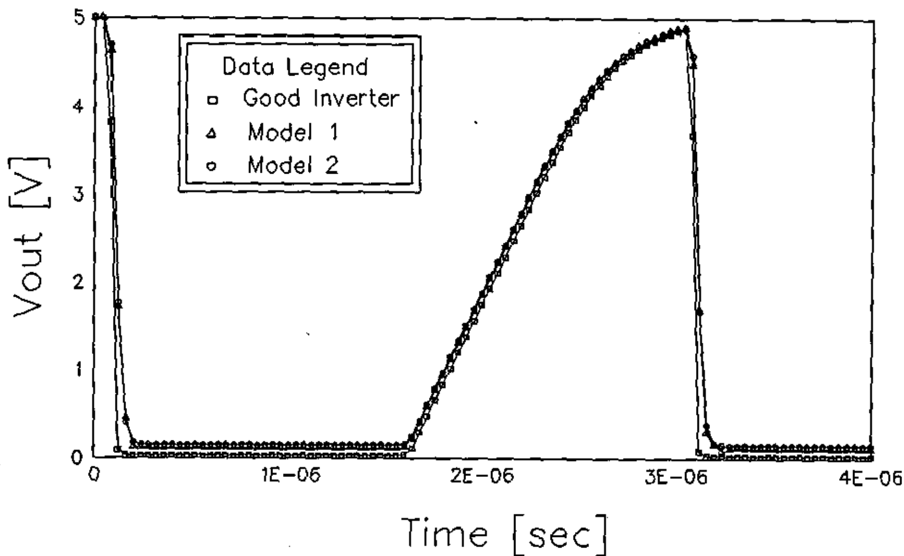


Figure 5.14: Transient characteristics of the nMOS inverter from the PSPICE simulation.

Again, a very good agreement was found between the simulations and the experiments.

5.5.4 Conclusion

Although in transistor level simulation results from model 2 are more accurate than those from model 1, the results for circuit level simulations are exactly the same (even for circuit shown in Fig.5.17). Consequently, although Syrzycki's model is more accurate for transistor-level simulations, Segura's model is more suitable for circuit level simulations due to its simplicity.

5.6 Test Consideration of MOS circuits with GOS

5.6.1 CMOS Inverter

A CMOS inverter consists of a series-connected complementary pair of transistors connected between the power supply rails, V_{dd} and V_{ss}. Both gates of the transistors are connected together to form the input of the inverter, while the drains are similarly connected to form the output. The structure is shown in Fig.5.15.

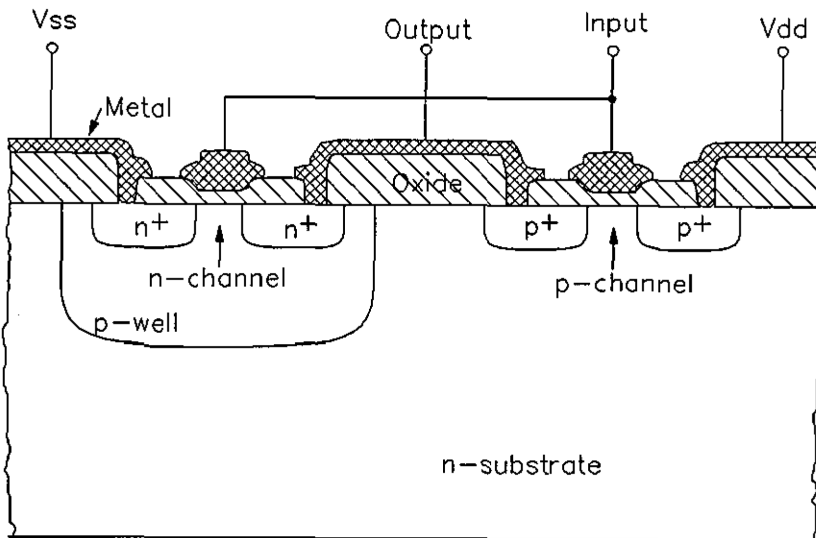


Figure 5.15: The cross-section of a metal-gate CMOS inverter circuit.

In a CMOS inverter, in either of its static states, one transistor is always in the 'OFF' state, and hence the static current from power supply is only the leakage current of the OFF transistor (of the order of 10^{-11}A [10]).

When a GOS occurs in either the n-channel or the p-channel transistor, a short occurs between the gate and the channel, source or drain. In the case of two cascaded CMOS inverters as shown in Fig.5.16, where the GOS is in the N2 transistor, the extra current drawn from the supply would contribute to the static power current when the input is logic '0'. Many experimental measurements of this extra current have been published [8,10],

and are generally in the range $10^{-9} \sim 10^{-4}$A. The static power current is several orders of magnitude greater than that of the same CMOS structure without the GOS fault. Thus measuring the static I_{dd} is a very easy and effective way to detect GOS in CMOS logic circuits.

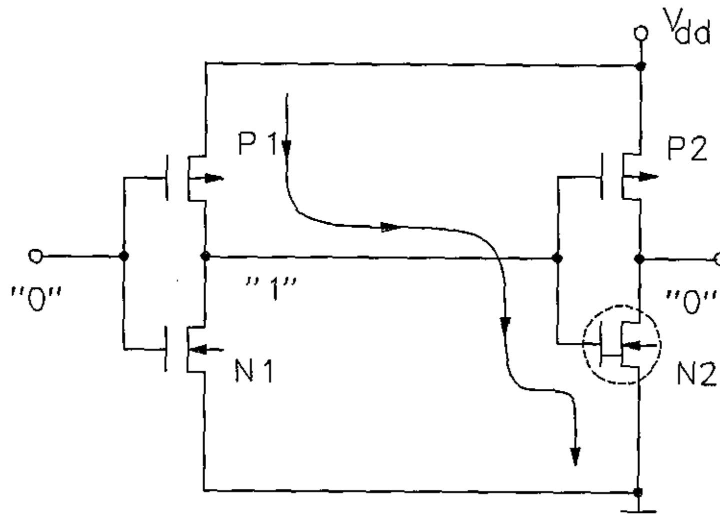


Figure 5.16: Two cascaded CMOS inverters with a GOS in one of the transistors.

5.6.2 NMOS Inverter

In nMOS inverter circuits, however, there is a continuous current drawn from power for a GOS-free circuit when input is logic '1'. In a cascaded pair of nMOS inverters, the extra current induced by a GOS is smaller than the order of magnitude GOS-free power current. Hence, no conclusion about the existence of GOS could be reached by measuring the I_{dd}.

According to the experimental results in Section 3.4.4, significant changes occurred in the transistor's drain current and threshold voltage after gate oxide breakdown. From the circuit delay theory in Section 5.5.2, a visible change in delay time would be expected, and has indeed been observed in the experimental data of Fig.3.21.

4.6.3 Other CMOS Logic Circuits

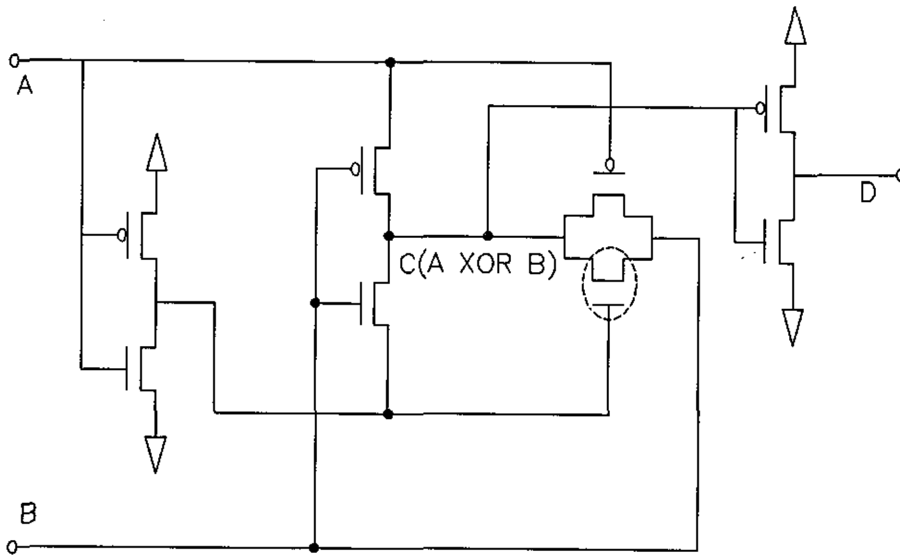


Figure 5.17: Transmission gate intensive NXOR logic gate.

PSPICE simulation was then performed on a transmission gate intensive NXOR logic gate with one of the transistors suffering from a GOS. The circuit is shown in Fig.5.17 and the GOS-fault transistor is circled with a broken line.

Simulation results are shown in Fig.5.18. The solid lines represent the case where all the devices are GOS-free, while the broken lines represent the case in which one of the transistor has a GOS. The simulation from both models are the same.

The simulation results in Fig.5.18 clearly show that although the difference in output 'low' level at node C is lost at node D, the time delay passes to the next gate and ultimately reaches the primary output.

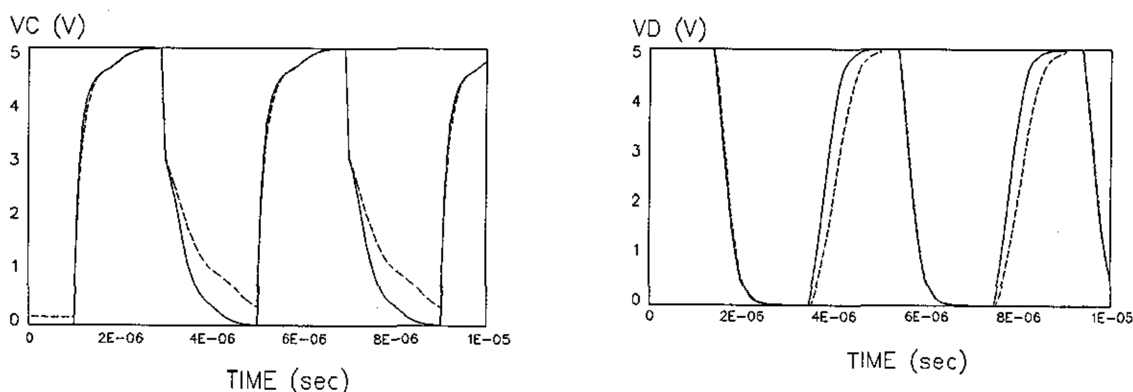


Figure 5.18: PSPICE simulation results (left: output waveform at node C where the logic output is A XOR B; right: output waveform at node D where the logic output is A NXOR B.).

5.6.4 Conclusion

Although the measurement of static current from power supply is a simple way to test for the presence of GOS, it does not work for all the cases. However, a GOS introduces an extra delay in signal transition from input to output and consequently the transient response of the output is an effective alternative test for the existence of a GOS.

5.7 References:

[1] MicroSim Corporation, "PSpice Circuit Analysis", Version **4.04**, July, 1990.

[2] Cobbold, R.S.C., "Theory and Applications of Field-Effect Transistors", Wiley-interscience, 1970.

[3] Milne, A.D. (editor), "MOS Devices - Design and Manufacture", Edinburgh University Press, 1982.

[4] Shichman,H, Hodges,D.A., "Modelling and Simulation of Insulated-Gate Field-Effect Transistor Switching Circuits", IEEE Journal of Solid-State Circuits, Sept., 1968, pp.285-289.

[5] Vladimirescu,A, Liu,S, "The Simulation of MOS Integrated Circuits Using SPICE2", Memorandum No.UCB/ERL **M80/7**, Electronics Research Lab., College of Engineering, University of California, Berkeley, Feb., 1980.

[6] Pierret,J.R., "A MOS Parameter Extraction Program for the BSIM Model", Memorandum No.**M84/99** and **M84/100**, Electronics Research Lab., College of Engineering, University of California, Berkeley, Nov., 1984.

[7] Segura,J.A., Figueras,J, Rubio,A, "Approach to the Analysis of Gate Oxide Shorts in CMOS Digital Circuits", Microelectron. Reliab., **32(11)**, pp.1509-1514, 1992.

[8] Syrzycki,M, "Modelling of Gate Oxide Shorts in MOS Transistors", IEEE Trans. on Computer-Aided Design, **8(3)**, pp.193-202, 1989.

[9] Mukherjee,A, "Introduction to nMOS & CMOS VLSI System Design", Prentice-Hall, Englewood Cliffs, 1986.

[10] Greason,W.D., Kucerovsky,Z, Chum,K.W.K., "Latent Effects Due to ESD in CMOS Integrated Circuits: Review and Experiments", IEEE Trans. on Industry Applications, **29(1)**, pp.88-97, 1993.

[11] Soden,J.M., Hawkins,C.F., "Test Considerations for Gate Oxide Shorts in CMOS ICs", IEEE Design and Test, August, pp.56-64, 1986.

[12] Favalli,M, Dalpasso,M, Olivo,P, Riccò,B, "Analysis of Resistive Bridging Fault Detection in BiCMOS Digital IC's", IEEE Trans. on VLSI Systems, **1(3)**, pp.342-355, 1993.

[13] Dong,L, Dwyer,V.M., "A Comparison of Static and Dynamic Techniques for the Detection of Gate Oxide Shorts in Digital MOS Circuit", 5th ESREF, Glasgow, 1994.

CHAPTER 6

Conclusions

MOS gate oxide breakdown is studied in this thesis. The experiments were limited to a single 4" p-type silicon wafer. The structures of the device under test (DUT) include: MOS capacitors, MOSFET networks and nMOS inverter circuits. After characterisation of the devices, constant voltage stresses were applied to the gates of the DUT to cause dielectric breakdown of the oxides.

The experimental results may be summarised by:

(i) The failure distributions appear to be bi-modal, with distinct 'early' and 'late' failure categories. The ratio of 'early' failures to 'late' failures increased with increasing stress voltage. The distributions are lognormal in shape.

(ii) Limited correlation exists between the times-to-breakdown of MOSFET and MOS-C structures fabricated upon the same die.

(iii) For enhancement n-channel MOSFETs, several parameters drift after the gate oxide breakdown. The threshold voltage increases from 1.55V to 2.30V while the mutual transconductance g_m of the transistor drops from $4.1 \cdot 10^{-4}$ (Ω^{-1}) to $2.2 \cdot 10^{-4}$(Ω^{-1}), this is likely to be due to an increase in scattering in the channel.

(iv) An extra time delay is found in an nMOS inverter circuit, together with some slight changes in DC transfer curves, after the pull-down transistor is subjected to the constant voltage stress.

One of the main objectives of this work was to understand the time profile of the oxide tunelling current. It is found that this profile strongly influences the observed time-to-failure distributions and may be understood by considering trap dynamics.

Both electron and hole trapping are observed to occur during stressing. Trapped electrons are believed to be formed by the capture of injection electrons both in the pre-existing electron traps in the oxide and in the traps that are generated during the high field

stress. Positive trapped charges are believed to be formed by means of a direct tunnelling between the anode and the trap sites near the SiO_2/Si interface. Based on the above charge trapping mechanisms, a mathematical model is proposed which shows good agreement with the experimental results.

An important side effect of the current profile is a time variation of the 'constant' oxide voltage. This can significantly affect the apparent ratio of 'extrinsic' to 'intrinsic' failures.

The second objective of the work was to investigate some of the consequences of defective devices reaching the market place.

PSPICE simulations were performed on MOSFETs and on nMOS inverter circuits. Two models for a MOSFET with GOS were used in the simulations. Syzrycki's model is more accurate in transistor level simulation, however, Segura's model is preferred in circuit level simulation for its simplicity.

In order to detect a GOS in an MOS circuit, the measurement of the static current from the power supply is a simple way for the test of GOS, although it does not work in all the cases. However, a GOS also introduces an extra delay in signal transition from input to output and consequently the transient response at output is an effective way to test the existence of GOS.

APPENDIX

Reproductions of Published Papers

This appendix reproduces a series of papers, written by the author, which relate to the work of this thesis.

Paper I

Dong, L., Tunnicliffe, M.J., Dwyer, V.M.,

"Pre-breakdown Charge Trapping
in High Field Stressed Oxides",

Proc. 5th. European Symposium
on Reliability of Electron Devices,
Failure Physics and Analysis
(ESREF)

Glasgow, Scotland, 1994.

PRE-BREAKDOWN CHARGE TRAPPING IN HIGH FIELD STRESSED OXIDES

L Dong, M J Tunnicliffe and V M Dwyer

Department of Electronic and Electrical Engineering
Loughborough University of Technology
LEICS, LE11 3TU, U.K.

Tel: +44 509 222856, Fax: +44 509 222854

ABSTRACT

Constant voltage stresses were applied to MOS capacitor structures and the resulting tunnelling current profiles were recorded. It was found that hole trapping occurred only during the first few seconds of the stress, while the electron trapping occurred throughout the stress period. A quantitative model for charge trapping during the build-up stage is presented and is shown to agree well with the experimental current profiles.

1.INTRODUCTION

Gate dielectric breakdown is a major cause of MOS circuit failure, especially in very large scale circuits. It has been widely accepted that oxide breakdown is triggered when the accumulation of trapped holes in the oxide reaches a critical value. A proper understanding of charge trapping is therefore vital if an accurate breakdown model is to be developed.

The build-up stage of the dielectric breakdown starts when electrons enter the oxide conduction band by means of the Fowler-Nordheim tunnelling mechanism. The tunnelling current is exponentially dependent upon the cathode electric field, which is affected by the trapped charges. The investigation reported here focuses on the tunnelling current profile, *from which the charge trapping model* is proposed.

The phenomenon of the charge trapping in thermally grown SiO_2 has been a subject of many investigations. Liang and Hu (Ref.1) presented a mathematical model of electron trapping and trap generation in silicon dioxide. Chen et al. (Ref.2) reported that holes were trapped after they were generated by high-energy electrons, while Jenq (Ref.3) suggested hole trapping was the result of direct tunnelling between the anode and the trap sites.

New experimental data is presented in this paper and a mathematical model of charge trapping in SiO_2 is proposed.

2. EXPERIMENTAL RESULTS

Experimental studies were limited to a single 4" <100> p-type Si wafer, containing 215x268μm MOS capacitor structures. The oxide thickness of the capacitors had a mean value of 41.42nm and a standard deviation of 0.161nm.

Constant voltage stress was generated by a Hewlett Packard HP4145B parametric analyzer and applied to the devices-under-test (DUT) between the gate and substrate terminals. Throughout the experiments, the gates were stressed negatively with respect to substrate and the oxide current was monitored by the analyzer as a function of time.

During the first set of experiments, relatively low stress voltages (-39V,-39.5V and -40V) were applied to the devices. The resulting tunnelling currents (Fig.1) increase from their initial values to a maximum, before decaying with time.

Subsequent experiments, performed using higher stress voltages (-42V and -43V), showed tunnelling currents which appeared to decay from the very beginning (Fig.2).

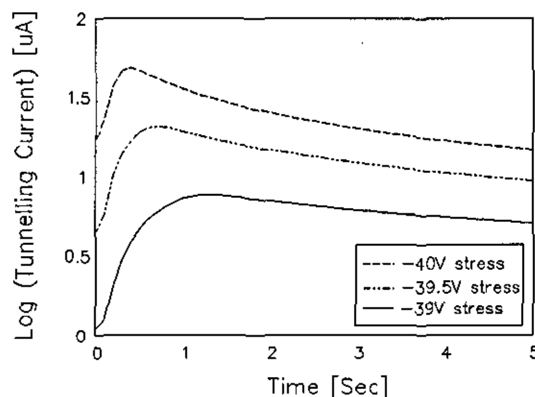


Figure 1: Tunnelling current profiles for stresses -39V, -39.5V and -40V.

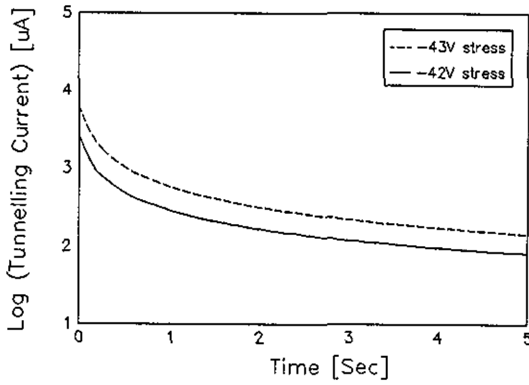


Figure 2: The oxide tunnelling current profiles for stress -42V and -43V.

The tunnelling current profiles may be characterised by two parameters: (i) the maximum current and (ii) the time at which this current is reached. The variation of these parameters with stress voltage is shown in Figs.3 and 4.

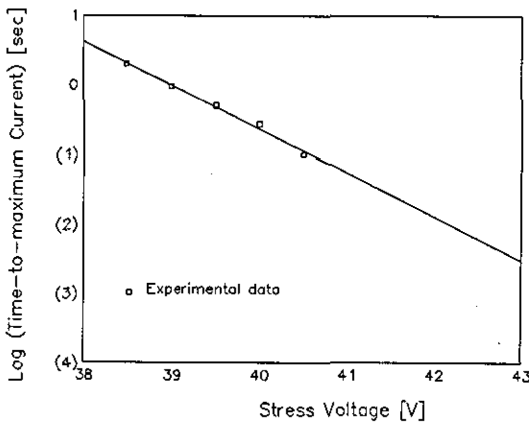


Figure 3: The time of maximum tunnelling current as a function of stress voltage.

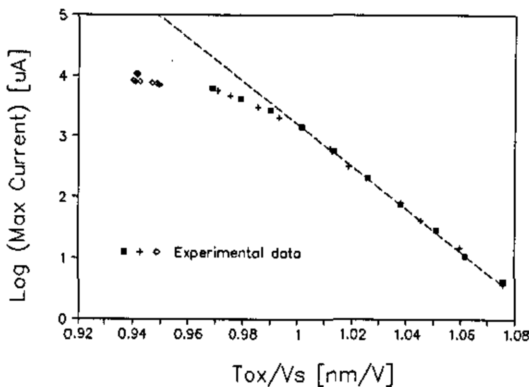


Figure 4: Maximum current vs. stress field.

The logarithm of the maximum tunnelling current varies linearly with 1/E at lower fields (in accordance with standard tunnelling theory (Ref.4)), and saturates at higher fields (Fig.4).

3. MATHEMATICAL MODELLING

The oxide current is determined by the Fowler-Nordheim equation (Ref.4):

$$J(t) = A E_{cat}^2(t)\, e^{-\frac{B}{E_{cat}(t)}} \qquad (1)$$

where J(t) is the tunnelling current density, E_{cat} is the cathode field and A and B are constants. The profile of tunnelling current in Fig.1 can be understood from a consideration of the charge trapping mechanism and its effect upon $E_{cat}(t)$.

3.1 Negative charge trapping

The non-saturating tunnelling current decay observed in Figs.1 and 2 suggests a continuous generation of electron trap sites. From first order rate equation, the trapped electron density Q_{ot}^- can be determined using the model of Liang and Hu (Ref.1)

$$Q_{ot}^- = qN_{op}\left(1 - e^{-\int_0^t \sigma J dt/q}\right) + \int_0^t gJ(t)dt - \frac{qg}{\sigma_g}\left(1 - e^{-\int_0^t \sigma_g J dt/q}\right) \qquad (2)$$

where σ and σ_g are the respective capture cross-sections of pre-existing traps and newly generated traps, and, N_{op} is the pre-existing trap density and g is the trap generation rate.

3.2 Positive charge trapping

The increase of oxide current in Fig.1 can be explained as a consequence of positive charge trapping. Since the current increase is very rapid, it is believed to be formed by direct hole tunnelling between the anode and the trap sites near the SiO_2/Si interface (Ref.3). The increment of trapped charge in time interval dt is given by

$$dQ_{ot}^{+}(t) = J_{hole}(t)dt \qquad (3)$$

where J_{hole} is the hole tunnelling current. The latter may reasonably be assumed to have a F-N type dependence upon anode field E_{an} and to be proportional to the density of available hole traps in the anode (Ref.5), i.e.

$$J_{hole} = Ce^{-\frac{D}{E_{an}}}\left(1 - \frac{Q_{ot}^{+}}{qN_{int}}\right) \qquad (4)$$

where N_{int} is the total density of hole traps and C and D are constants dependent upon structure of the anode/oxide/trap system. The factor in parenthesis causes the trapping to slow as the available number of states decreases, and to tend to 0 as $Q_{ot}^{+} \rightarrow qN_{int}$.

The oxide field profile changes when charges are trapped within the dielectric. If we assume sheet charges located at their respective centroids, then the electric field at cathode will be given by:

$$E_{cat} = \frac{V_{ox}}{T_{ox}} - \left(1 - \frac{x_n}{T_{ox}}\right)\frac{Q_{ot}^{-}}{\epsilon_{ox}} + \left(1 - \frac{x_p}{T_{ox}}\right)\frac{Q_{ot}^{+}}{\epsilon_{ox}} \qquad (5)$$

in which Q_{ot}^{+} and Q_{ot}^{-} are the magnitudes of the trapped hole and electron densities, x_p and x_n are the respective centroids of Q_{ot}^{+} and Q_{ot}^{-} and T_{ox} is the oxide thickness. The voltage across the bulk Si resistance ($\approx 250\Omega$) is much smaller than the stress voltage increment ($\geq 0.5V$) and was therefore ignored in the model.

The effect of trapped holes on the anode field can be expressed as:

$$E_{an} = \frac{V_{ox}}{T_{ox}} - \frac{Q_{ot}^{+}}{\epsilon_{ox}}\frac{x_p}{T_{ox}} + \frac{Q_{ot}^{-}}{\epsilon_{ox}}\frac{x_n}{T_{ox}} \qquad (6)$$

Simultaneous numerical solution of equations (1-6) allows the tunnelling current to be calculated. Parameters were given values consistent with earlier publications, with the exception of N_{int}, which was adjusted in order to obtain an optimal correlation with the experimental data. The results are shown in Fig.5 and Fig.6.

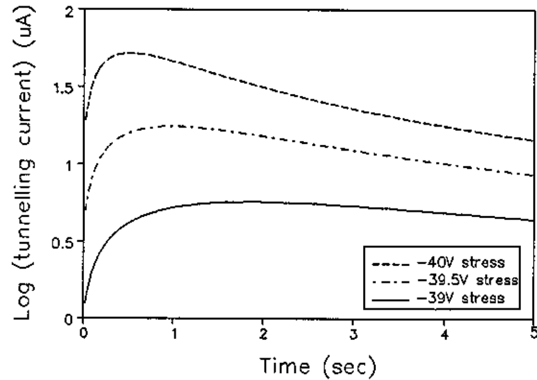


Figure 5: Modelling results of tunnelling current profiles for lower voltage stresses.

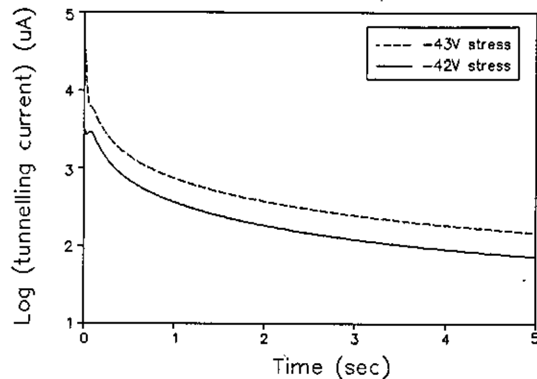


Figure 6: Modelling results of tunnelling current profiles for higher voltage stresses.

The results in Fig.2 and Fig.3 can also be explained by this model. Firstly, since the density of available traps is fixed, the time required to fill them decreases linearly with increasing current and hence exponentially with increasing voltage (Eqn.4). Secondly, the rapid hole-trap filling at high fields (~10ms for -42V stress) lies beyond the time-interval of the parametric analyzer (which was set to 100ms throughout these experiments). Hence the maximum recorded current lies within the decay-portion of the current profile, where the electron trapping compensates the trapped hole charge. The *apparent* maximum current therefore increases less

rapidly with increasing field, as shown in Fig.3.

4. CONCLUSION

Dielectric breakdown in MOS devices appears to be triggered by the accumulation of holes in oxide trap states. The build up of these trapped holes may be monitored by studying the current through the oxide. Fig.1 clearly show the effects of charge trapping.

This paper presents a simple physical model to describe such effects. In the model, the initial current rise is dominated by the tunnelling of holes from the silicon valence band directly into trap states located near the Si/SiO_2 interface. This mechanism soon appears to saturate, after which the slower electron trapping begins to dominate, causing the observed tunnelling current decay.

Despite the fact that hole tunnelling into such traps is poorly understood, and the relative simplicity of the model, a reasonable agreement is found between calculated and measured currents.

5. REFERENCES

1. M-S.Liang and C.Hu, "Electron Trapping in Very Thin Thermal Silicon Dioxides", IEDM Tech. Dig. (Washington D.C), pp.396-399, 1981.

2. I.C.Chen, S.E.Holland and C.Hu, "Hole Trapping and Breakdown in Thin SiO_2", IEEE Trans. on Electron Device Letters, Vol.EDL-7, No.3, pp.164-167, 1985.

3. C.S.Jenq, T.R.Ranganath, C.H.Huang, H.S.Jones and T.T.L.Chang, "High-field Generation of Electron Traps and Charge Trapping in Ultra-thin SiO_2", IEDM Tech. Dig. (Washington D.C), pp.388-391, 1981.

4. M.Lenzlinger and E.H.Snow, "Fowler-Nordheim Tunnelling into Thermally Grown SiO_2", J. Appl. Phys., Vol.40, No.1, pp.278-283, 1969.

5. V.M.Dwyer, M.J.Tunnicliffe, L.Dong, to be published.

Paper II

Dong, L., Dwyer, V.M.,

"A Comparison of Static and Dynamic Techniques For the Detection of Gate Oxide Shorts in Digital MOS Circuits"

Proc. 5th European Symposium on Reliability of Electron Devices, Failure Physics and Analysis (ESREF)

Glasgow, Scotland, 1994.

A COMPARISON OF STATIC AND DYNAMIC TECHNIQUES FOR THE DETECTION OF GATE OXIDE SHORTS IN DIGITAL MOS CIRCUITS

L Dong and V M Dwyer

Department of Electronic and Electrical Engineering
Loughborough University of Technology
Loughborough
LEICS, LE11 3TU, U.K.

Tel: +44 509 222856, Fax: +44 509 222854

ABSTRACT

Gate oxide shorts (GOS) represent a reliability threat in MOS based ICs and commonly used static tests for such shorts are not always effective. In this paper, a GOS was intentionally induced in an inverter circuit resulting in an extra time delay which is passed on to the output pin. Using a simple model for the defective oxide PSPICE simulations were performed and a good agreement was achieved between the experimental and simulation results. As a result, a simple dynamic test, measuring the signal delay through the circuit also may indicate the presence of a GOS.

1. INTRODUCTION

Gate oxide shorts in MOS transistor circuits have long been recognised as a serious reliability problem, especially for scaled-down devices (Ref.1). They can be caused either during manufacturing process or by electrostatic discharge (ESD) after packaged. A large amount of research has been carried out, both experimentally and theoretically.

GOS defects cannot be detected by electrical test programs based on stuck-at fault models since the circuits with GOS defects generally do not lose their function. However, electrical parametric changes do occur after GOS present in circuits.

GOS fault can be easily detected in CMOS circuits by measuring the static power current, however, this method does not work in nMOS circuits. In this paper, a simple dynamic test is proposed and its effectiveness is evaluated.

2. STATIC TEST FOR GOS

The most obvious characteristics for a MOS transistor with GOS is that there is gate current when the transistor is on. Since there is no static power dissipation in a GOS-free CMOS circuit (the leakage current is in the order of 10^{-11}A (Ref.2)), the presence of GOS gives rise to a conductive path between the power supplies as shown in Fig.1.

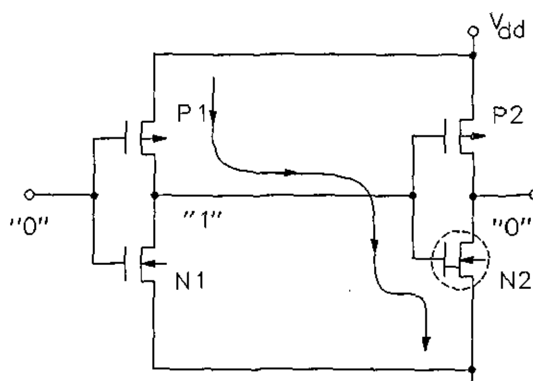


Figure 1: Schematic diagram of current drawn from power supply for a two-cascaded CMOS inverters with GOS defect in one of the transistors.

This current varies from 10^{-9}A to 10^{-3}A (Refs.1,2) and depends on the size and the extent of the oxide failure. Even the lightest GOS defect will increase the static power consumption up to 2 orders. Hence, static power consumption is a very effective way to detect GOS faults in CMOS inverter circuits.

It would not be so simple for a very complicated logic circuits. However, it can be solved by measuring the power supply during the logic test procedure.

However, problems arise for the detection of nMOS circuit with GOS since there is a static power dissipation even for a GOS-free circuit. In a two-cascaded nMOS inverter GOS-free circuit, for any of the static states, both of the pull-up transistors are on and one of the pull-down transistors is on. This

determines the static power current is the 'on' current of pull-up transistor. While the presence of GOS only adds a gate leakage current (in the order of μA) to the power current. Consequently, if any of the transistors contain a GOS, no noticeable difference is detected.

3. DYNAMIC TEST FOR GOS

3.1 Theory of circuit delay

In MOS circuits, the outputs at any internal node connect to the inputs of the next gate. The oxide between gate and channel in MOS transistors forms a capacitor C_l between the gate and substrate, such that the load at the internal node appears as C_l. This internal circuit can be modelled as a circuit shown in Fig.2.

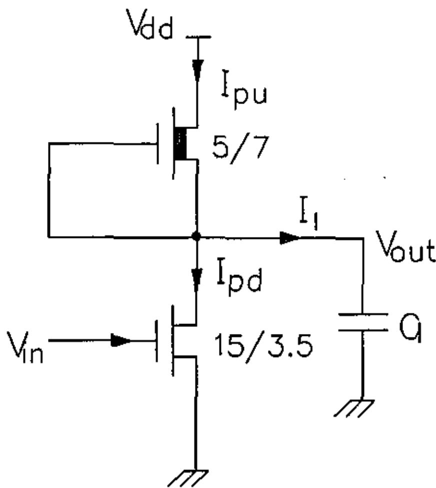


Figure 2: A nMOS inverter.

The charge and discharge times t_c and t_d can be determined by equation (1,2).

$$C_l(V_{dd} - V_{lo}) = \int_0^{t_c}(I_{pu} - I_{pd})dt \qquad (1)$$

$$C_l(V_{dd} - V_{lo}) = \int_0^{t_d}(I_{pd} - I_{pu})dt \qquad (2)$$

Normally the transistor current characteristics degrades upon creation of a GOS, and since the delay time of a gate depends on both the load capacitance and the charge or discharge current, a change in time delay in expected.

3.2 Test samples

A 4" wafer containing a nMOS inverter as shown in Fig.2 on each die was used throughout the experiments. The wafer was held upon a brass chuck by a vacuum pump, and four adjustable microprober were used to access the device bond-pads. The pull-up transistor is a depletion-mode transistor with a dimension of 5/7 (μm), while the pull-down transistor is a enhancement-mode transistor with a dimension of 15/3.5 (μm).

3.3 Experimental procedure and results

3.3.1 Characterisation of the pull-down transistors and nMOS inverters

Measurement of the I_G - V_G and I_{DS} - V_{DS} characteristics of the pull-down transistors using the HP4145B parametric analyzer: The gate current is only a noise in the order of nA magnitude and the drain current characteristics are shown in Fig.3.

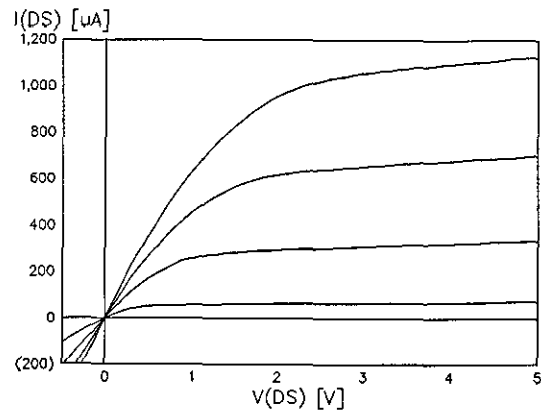


Figure 3: I_{DS} - V_{DS} characteristics of the pull-down transistor.

Measurement of the DC transfer curve of the inverter: Both the output voltage and the supply current were recorded, and are shown in Fig.4 and Fig.5 (solid lines).

Measurement of the transient characteristics of the inverter: A 300KHz square waveform from a THANSAR TG102 pulse generator was applied to the input of the inverter and a HP54111D storage oscilloscope was configured to capture the output signal. The result is shown in Fig.6 (solid line).

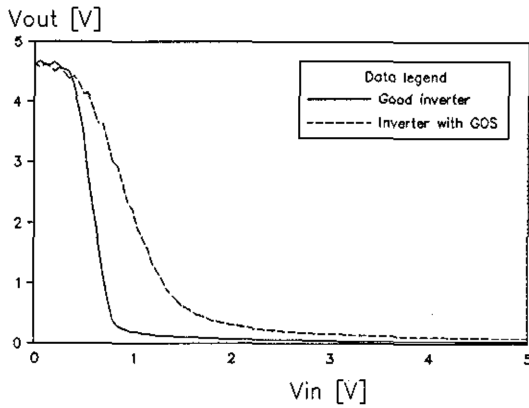


Figure 4: DC transfer curve of the nMOS inverter.

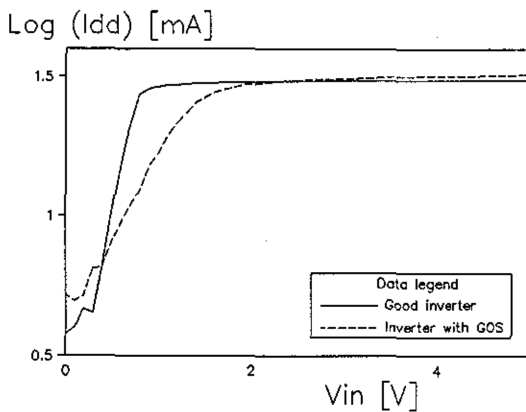


Figure 5: I_{DD} - V_{IN} characteristics of the nMOS inverter.

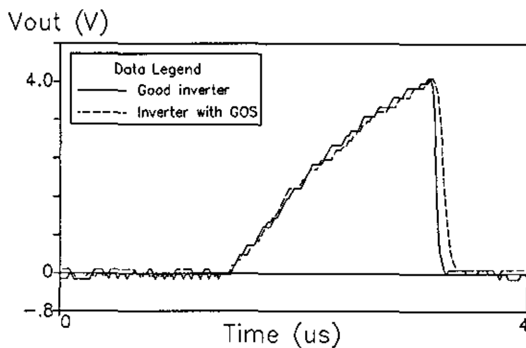


Figure 6: Transient response of the inverter to a 300KHz square waveform.

3.3.2 Oxide breakdown of the pull-down transistors

In order to introduce the gate oxide short, a constant voltage stress (-44V) from the HP4145B parametric analyzer was applied to the gate pad of the pull-down transistor. The stress was removed immediately after the oxide breakdown.

3.3.3 Re-characterisation of the pull-down transistors and the nMOS inverters

The oxide breakdown was confirmed by the gate current characteristics shown in Fig.7.

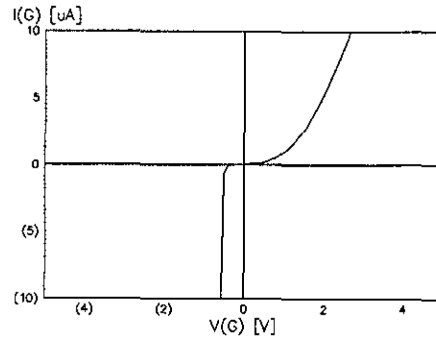


Figure 7: I_G - V_G characteristics of the pull-down transistor after the gate oxide breakdown.

After the measurement of gate current, all the other characteristics of the pull-down transistors and nMOS inverters were re-measured and are shown in Fig.4, Fig.5 and Fig.6 with dashed lines.

3.3.4 Experimental conclusion

We still consider the two-cascaded nMOS inverters circuit: in the case of GOS-free circuit, the static power current is 32.43+4.26 = 36.69 (μA). While with one of the pull-down transistor with GOS, the static power current is 32.1+5.197 = 37.297 (μA). Thus, no conclusion can be made on the static current results. Although there is a noticeable difference in the output 'low' logic level, it will be lost through the next gate. However, from the transient response of the square waveform, the difference in the discharge time is clearly shown in Fig.6, also this will not disappear through successive gates and will ultimately reach the external output.

4. SIMULATION RESULTS FROM PSPICE

Syrzycki's model (Ref.3) as shown in Fig.8 for

MOSFET with GOS was used in the simulations.

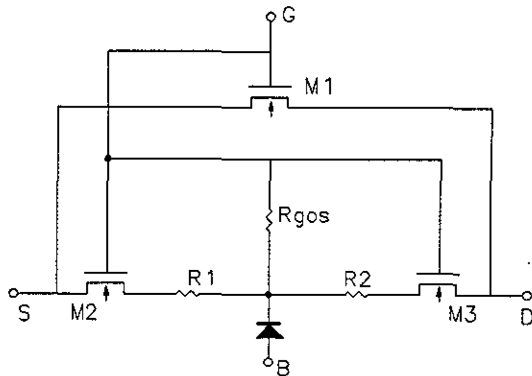


Figure 8: *Syrzycki's model for nMOSFET* with GOS.

The parameters of the transistors and resistors in Fig.8 were chosen to fit the experimental results *of the gate current and drain current characteristics* of the pull-down transistor after the inducing of GOS (Fig.5 & 6). Then the simulations for the inverter were performed based on the same parameters. The simulation results are shown in Fig.9-11 and in which the solid lines represent the case where all transistors are good devices, while the broken lines represent the case where the pull-down transistor has a GOS-fault.

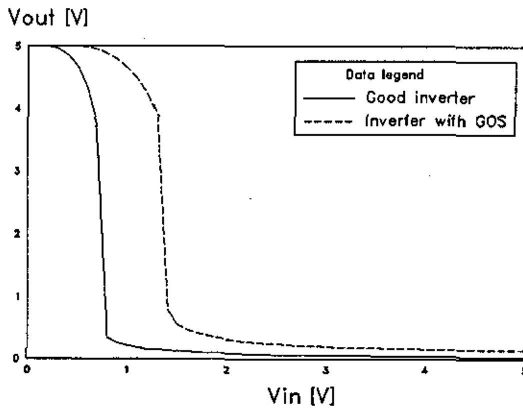


Figure 9: DC transfer curve of the nMOS inverter from PSPICE simulation.

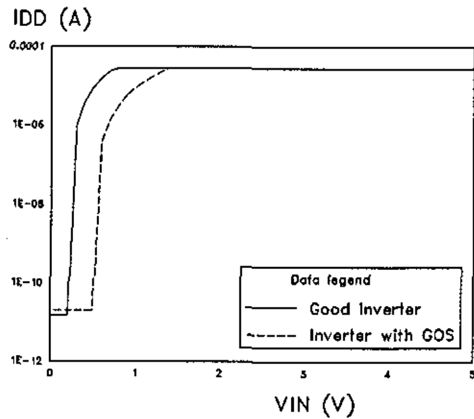


Figure 10: I_{DD} - V_{IN} curve of the nMOS inverter from PSPICE simulation.

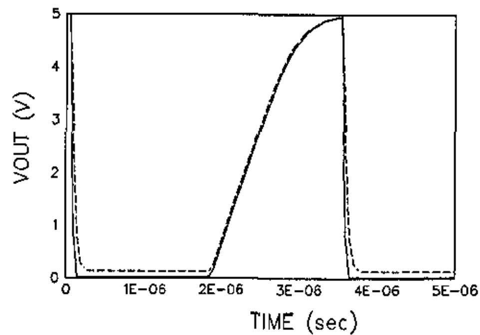


Figure 11: Transient response from simulation.

The simulation results agree very well with the experimental ones. This indicates that Syrzycki's model can not only model the DC effect of GOS in circuits, but also the transient parametric drift of the circuits.

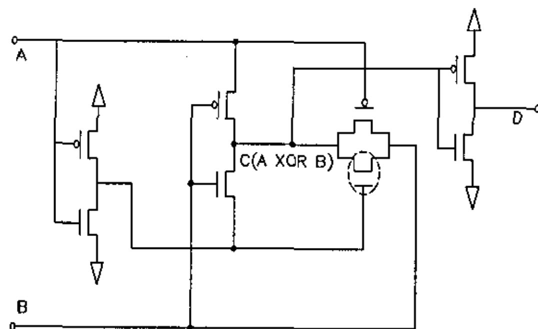


Figure 12: Transmission gate intensive NXOR logic gate.

Following the successful results above, simulations for a more complicated logic circuit as shown in Fig.12 were performed. The GOS-fault transistor is circled with broken line. The results are shown in Fig.13 and Fig.14 and the solid lines represent GOS-free circuit and broken lines represent the circled transistor is with GOS.

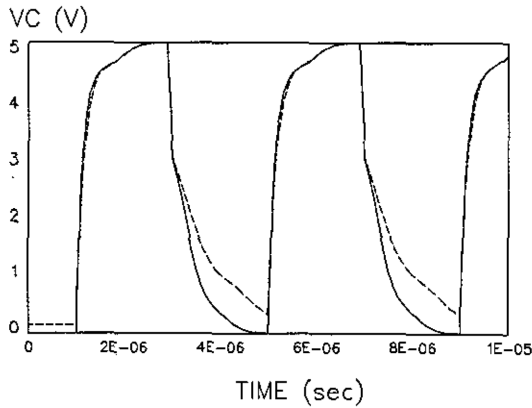


Figure 13: Output waveform at node C where the logic output is A XOR B.

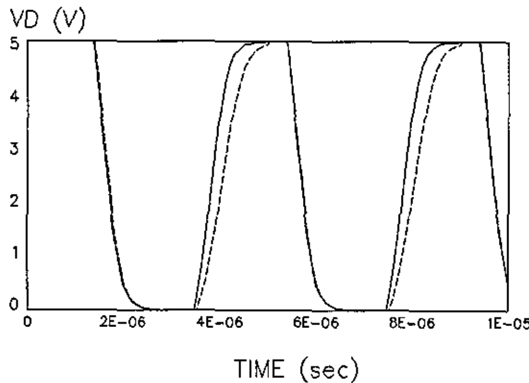


Figure 14: Output waveform at node D.

The simulation results clearly show that although the difference of output 'low' level at node C has lost at node D, the time delay passes the next gate and will reach the primary output at last.

5. CONCLUSIONS

Whilst the measurement of static current from power supply is a simple method of testing MOS circuits for gate oxide shorts, it does not work for all cases. However, a GOS also introduces an extra delay in signal transmission from input to output and consequently the delay test in those cases could be an effective alternative.

6. REFERENCES

1. J.M.Soden and C.F.Hawkins, "Test Considerations for Gate Oxide Shorts in CMOS ICs," IEEE Design and Test, August, 1986, pp.56-64.

2. W.D.Greason, Z.Kucerovsky and K.W.K.Chum, "Latent Effects Due to ESD in CMOS Integrated Circuits: Review and Experiments," IEEE Transaction on Industry Applications, No.29, 1993, pp.88-97.

3. M.Syrzycki, "Modelling of Gate Oxide Shorts in MOS Transistors," IEEE Transaction on Computer-Aided Design, No.8, 1989, pp.193-202.

Paper III

Tunnicliffe, M.J., Dong, L., Dwyer, V.M.,

"Monitoring the Integrity of MOS Gate Oxides"

Journal of Electroceramics, 1993.

Monitoring the Integrity of MOS Gate Oxides

by

Martin J. Tunnicliffe, Liqin Dong and Vincent M. Dwyer

International Electronics Reliability Institute,
Department of Electronic and Electrical Engineering,
Loughborough University of Technology,
Loughborough, Leicestershire, LE11 3TU.

Telephone: (0509)222807, Fax:(0509)222854.

Abstract

Inter-sample variability of time dependent dielectric breakdown in MOS gate oxides was experimentally examined. MOS−C and MOSFET structures were subjected to constant-voltage stress and the resulting times-to-breakdown were recorded. Failure distributions were found to be bi-modal, with distinct 'early' and 'late' failure categories. The ratio of 'early' failures to 'late' failures increased with increasing stress voltage. The 'late' failure distributions were approximately Weibull in shape, indicating a weak-link failure mechanism. Although pre-breakdown tunnelling was a function of electric field, the average time-to-breakdown appeared to depend solely upon voltage. Little correlation existed between the times-to-breakdown of MOSFET and MOS-C structures fabricated upon the same die. Analysis suggests that the oxide breakdown is in fact principally uni-modal, and that the apparent bi-modal behaviour is caused by interference from the transient voltage drop across the series substrate resistance.

1. Introduction

This paper describes part of an ongoing study of the nature and mechanisms of time-dependent dielectric breakdown (TDDB) in metal-oxide-semiconductor (MOS) gate-oxide layers. The general aims are:

1. The advancement of dielectric breakdown physics,
2. The development of an accurate dielectric breakdown model, and
3. The development a quick and efficient oxide characterization test for use during integrated-circuit fabrication.

So far, the investigations have focused on the times-to-breakdown (t_{bd}) of MOS devices subjected to constant voltage stress. Preliminary results, reported earlier this year [1], showed that:

1. Very little correlation exists between the times-to-breakdown of MOS field effect transistors (MOSFET's) and MOS capacitors (MOS-C's) fabricated on the same die.
2. Variations in the time-to-breakdown and the pre-breakdown oxide tunnelling current are statistically unrelated.

This latter observation is important since it conflicts with the generally-accepted model of TDDB: Namely that some form of 'damage' is continuously inflicted upon the oxide by the tunnelling current. Breakdown occurs when this 'damage' reaches some critical level [2,3]. This picture has become widely accepted, and has formed the basis for several elaborate failure models [e.g.4]. Any doubt about its validity must therefore be thoroughly investigated.

The present paper seeks to extend these earlier studies, by the acquisition of further experimental data and the development of a theoretical model to explain the results.

2. Test Samples and Apparatus

The device samples used throughout these experiments were small-dimension enhancement/depletion mode MOSFETs and wide-area (215x268μm) MOS−C structures, fabricated

on a 4" <100> p-type silicon wafer. All devices had gate-oxides of nominal thickness 40nm, and n^+-doped polysilicon gate electrodes.

The MOSFETs on each die were interconnected with common source and gate contacts and individual drain terminals, such that gate-stress was applied to the entire network at once. The total gate-area A of the network was approximately $3.10^{-4}cm^2$. The MOS-C gate inputs were isolated from all other devices on the die.

The variation in local oxide thickness (T_{ox}) for each die was gauged by measuring the capacitance (C_{ox}) of an additional wide-area MOS-C structure (416x553μm) which was reserved for this purpose. Measurement was performed at 10kHz using a Wayne-Kerr 420 LCR meter, with a -10V bias potential to ensure carrier accumulation at both oxide surfaces. The oxide thicknesses, estimated from the formula $T_{ox}=\epsilon_0\epsilon_{ox}.Area/C_{ox}$ (where $\epsilon_0\epsilon_{ox}$=oxide permittivity), had a mean value of 41.42nm and a standard deviation of 0.161nm.

The series substrate resistances (R_b) for the MOSFET and MOS-C structures were obtained by measuring the device impedances at 10MHz and extracting the in-phase components. R_b was found to equal $154\pm10\Omega$ for the MOS-Cs and $79\pm10\Omega$ for the MOSFET networks.

The wafers were mounted on a 4" brass chuck and held in place by a vacuum suction system. A 20μm-tip manually-adjustable microprober was used to probe the device gate bond pads under microscopic observation. Constant-voltage stress was supplied by a Hewlett Packard 4145B parametric analyzer and applied to the device-under-test (DUT) between gate and substrate. The stress-voltage was accurate to within $\pm$10mV and had a rise-time of approximately 800μs. Throughout the experiments, the gates were stressed negatively with respect to the substrate.

The HP4145B was configured to monitor the oxide current as a function of time and the device voltage was independently monitored by a Hewlett Packard 54111D digital storage oscilloscope. For long time scales ($t_{bd}>1$sec), breakdown was easily detected by the HP4145B as a rapid rise in the injection current. For shorter time-scales ($t_{bd}<1$sec), the most accurate indication of breakdown was a sudden drop in the device voltage as measured by the oscilloscope.

3. Experimental Procedures and Results

A 15x13 matrix of die was selected in the centre of the wafer, and results are reported for those devices only. Approximately equal numbers of MOS-C's and MOSFET networks were subjected to constant-voltage stresses of −43.5V, −44.0V and −44.5V. Devices stressed at each particular voltage were spread as uniformly as possible across the wafer surface in order to eliminate any positional dependencies. The time-to-breakdown (t_{bd}) was recorded for each device, together with the maximum tunnelling current density (J_{max}) and the tunnelling current density at breakdown (J_{bd}). (Note: Current measurement was only possible for $t_{bd}>0.1$s due to the finite integration time of the parametric analyzer). Room temperature (28°C) was maintained throughout all the experiments, and the devices were constantly illuminated by the microscope illumination system.

The time-to-breakdown results are shown in Fig.1, plotted in the Weibull format, i.e. $ln(-ln(1-F(t)))$ vs. $log(t)$ (where $F(t)$=failure probability at time t). The distributions are clearly bi-modal, consisting of 'early' failures ($t_{bd}<100$ms) and 'late' failures ($t_{bd}>100$ms). The ratio of the numbers of 'early' and 'late' failures generally increases with the stress voltage (V).

The 'late' failure distributions are generally linear, indicating a 'Weibull' distribution of the form: $F(t) = 1 - exp[-(t/\tau)^\beta]$ (where β is the distribution slope and τ is the *characteristic lifetime*). The Weibull distribution is related to extreme-value statistics, and may be used to model failures at weak spots in a device [2,5]. The nature of the 'early' distribution is difficult to determine since its lower tail is dominated by the stress voltage risetime.

Fig.2 compares the values of $log(t_{bd})$ obtained from MOSFET and MOS-C structures on the same die. (Devices on any given die were always stressed at the same voltage). The broken diagonal lines indicate the results which would be expected if the failure times for a given die were equal. The high degree of scatter indicates that this is not generally true. Of those die who exhibited early failure, only 36% suffered it in both MOS-C and MOSFET network. This (together with the results of earlier studies [1]) demonstrates the limitations of the MOS-C as a test-vehicle for monitoring in-wafer variations of oxide integrity.

Figs.3 shows the values of $log(J_{max})$ measured for the MOSFET and MOS-C structures plotted against the corresponding values of $log(t_{bd})$. Although both depend upon voltage, no obvious relationship appears to exist between them within a given data set.

Fig.4 shows $log(J_{max})$ and $log(t_{bd})$ plotted against the corresponding electric fields $(V - A.J_{max}R_b)/T_{ox}$ and $(V - A.J_{bd}R_b)/T_{ox}$. The data can be analyzed either as a single data set or as a series separate data sets, each associated with a

particular stress voltage. In the former case, current clearly rises and t_{bd} falls with increasing field. In the latter case, although a distinct current rise is observed with increasing field, no corresponding t_{bd} decrease is obvious. If anything, the time-to-breakdown appears to *increase* with field (this is most obvious in the -44.0V data). However, insufficient data is available to confirm such a trend.

4. Discussion and Modelling of Results

Bi-modal failure distributions have been observed by many earlier workers [e.g.2,5] and have normally been attributed to 'intrinsic' and 'extrinsic' failures. Intrinsic failures are associated with the inherent oxide properties, while extrinsic failures are caused by structural defects and/or chemical impurities. The latter occur much more rapidly than the former, and form a separate statistical distribution.

The application of this model [2] to Fig.1 requires a voltage-dependent extrinsic defect density. It is possible that some extrinsic defects are only activated above a particular field, such that their apparent population increases with voltage. An alternative scheme, based upon the time-dependence of the pre-breakdown tunnelling current, is presented below.

Consider first the pre-breakdown tunnelling current transient. Fig.5 shows typical waveforms obtained from four different MOS-C structures at four different voltages. (Currents were monitored by measuring the voltage across a 771.4Ω resistor, connected in series with the oxide.) In all four cases, the oxide current magnitude ($I_{ox}(t)$) increases from zero to a maximum value at about 100ms, beyond which it decays with time. Since this current generates a voltage drop across the bulk resistance (R_b=250Ω), the *oxide* voltage magnitude ($V_{ox} = V - I_{ox}R_b$) experiences a 'dip' at approximately 100ms.

In order to continue, we make the following reasonable assumptions:

1. Oxide breakdown is caused by a continuous, non-reversible accumulation of damage inflicted during the pre-breakdown period. Breakdown occurs when the damage reaches a critical level (hereafter called the 'damage-to-breakdown'). This assumption is consistent with previous models [2,3].

2. The rate at which oxide damage accumulates increases rapidly with oxide voltage and is independent of the oxide tunnelling current. Although this differs from existing models (which generally require *damage rate* ∝ *exp(-constant/Field)* [3]), it is justified by the results in Figs.3 and 4 which show t_{bd} dependent on V only.

3. The damage-to-breakdown varies randomly between devices according to a uni-modal distribution. This eliminates the need to invoke any voltage-dependent 'extrinsic' defects.

When time<<100ms, the oxide voltage magnitude is high, together with the rate of oxide wearout. Many of the devices at the lower end of the damage-to-breakdown distribution fail in this region, creating the 'early' failure mode. As time increases to ~100ms, the voltage magnitude decreases together with the wearout rate. Very few devices fail in this region. When the current decays and the voltage rises once more, the remainder of the devices break down to form the 'late' failure distribution.

The above model would clearly account for the voltage dependence of the 'early' failure distribution: As the stress voltage increases, a greater number of oxides reach their critical damage level in the t<100ms region and fewer survive to fail in the 'late' distribution.

Finally, it is necessary to consider the physical implications of this work. The fact that t_{bd} depends upon voltage rather than field suggests that wearout is governed by the *absolute* energy of the charge carriers, rather than their rate of acceleration as was previously believed [2]. In addition to this, the apparent independence of the maximum current as a function of t_{bd} suggests that the injected electrons are not responsible for the oxide wearout. Damage may therefore be caused by some other mechanism such as mobile ion drift or substrate holes injection.

5. Conclusions

This paper presents an experimental study on the inter-sample variability of oxide wearout, together with a qualitative theoretical model to explain the observed failure distributions. The fundamental conclusions are as follows:

1. Time-dependent dielectric breakdown under constant voltage stress exhibits a bi-modal distribution. The 'late' (i.e. long

time-scale) mode is governed approximately by the Weibull model. The ratio of 'early' (short time-scale) failures to 'late' (long time-scale) failures increases with the stress voltage *V*.

2. While the pre-breakdown oxide tunnelling current is a function of electric field, the time-to-breakdown appears to depend solely upon the absolute voltage. For a given voltage, the tunnelling current and the time-to-breakdown are statistically *unrelated*.

3. There is little or no correlation between the times-to-failure of MOSFET and MOS-C structures fabricated on the same die. This supports earlier claims that the MOS-C has limited use as an oxide-integrity test vehicle [1].

A framework has been developed to explain the apparently bi-modal time-to-breakdown distributions in terms of a single uni-modal damage-to-breakdown distribution. Although the model is qualitative, a quantitative version is currently under development.

6. References

[1] TUNNICLIFFE,M.J., DWYER,V.M., presented at EPMS'93 Conference, 31 August to 3 September, 1993, Sheffield Hallam University, Sheffield, U.K. and accepted for the International Journal of Electronics, (1993).

[2] WOLTERS,D.R., VAN DER SCHOOT,J.J., (1985), *Phillips J. Res.*, **40**, pp 115-92.

[3] CHEN,I.C., HOLLAND,S, HU,C., (1985), *IEEE Trans. Electron Dev.*, **ED32**, pp 413-22.

[4] CHAN,C.K., CAREY,M.B., (1992), *IEEE Trans. on Reliability*, **41**, pp 414-20.

[5] VOLLERTSEN,R.-P., (1993), *Microelectron. Reliab.* **33**, pp 1665-77.

List of Figures

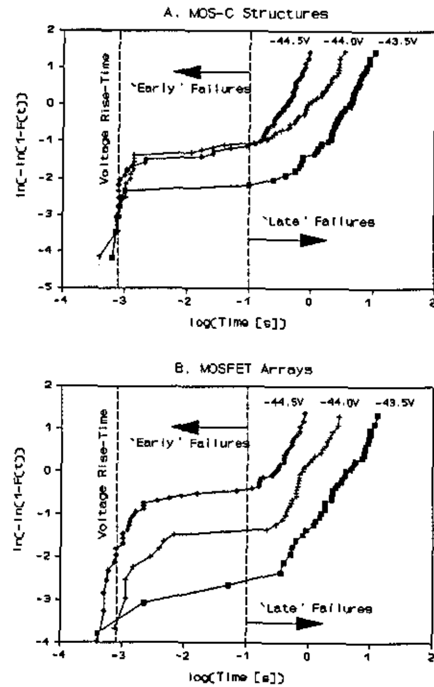


Figure 1: Time-to-breakdown distributions for MOS-C and MOSFET structures.

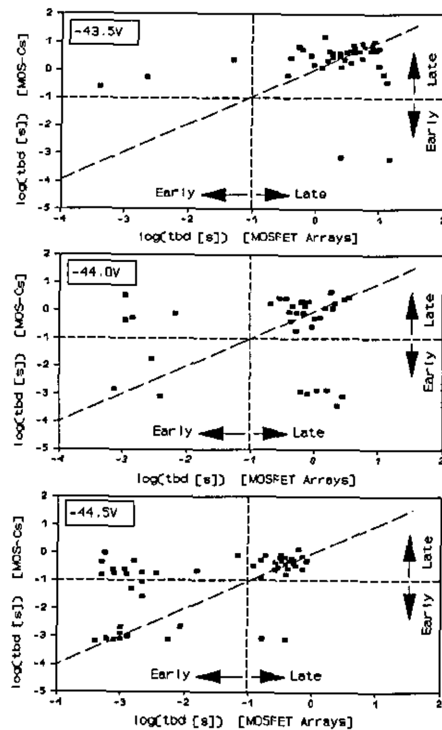


Figure 2: Comparison between breakdown-times of MOSFET and MOS-C structures fabricated on the same die.

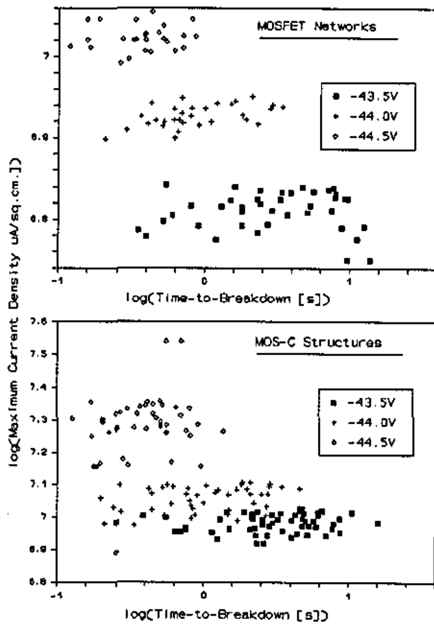


Figure 3: Maximum injection current vs. time-to-breakdown for MOSFET and MOS-C structures.

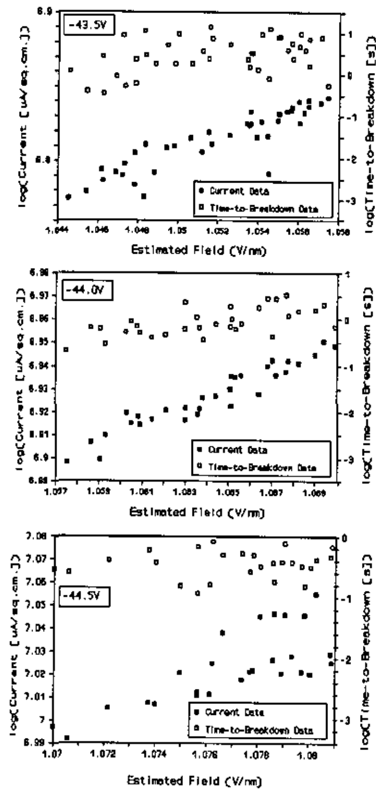


Figure 4: Field dependencies of maximum injection current and time-to-breakdown in MOSFET structures.

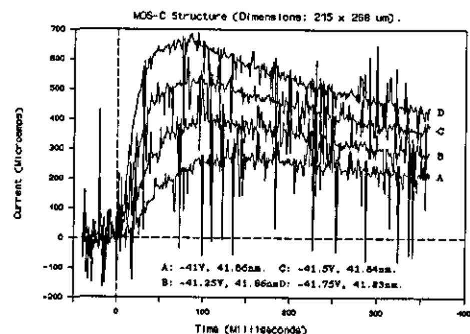


Figure 5: Typical injection-current waveforms for MOS-C structures.