Lehigh University
Lehigh Preserve

Theses and Dissertations

1993

Quantification of thin gate oxide impact on the switching characteristics of MOSFETs

Franklin Daniel. Nkansah
Lehigh University

Follow this and additional works at: http://preserve.lehigh.edu/etd

Recommended Citation
Nkansah, Franklin Daniel., "Quantification of thin gate oxide impact on the switching characteristics of MOSFETs" (1993). *Theses and Dissertations.* Paper 185.

This Thesis is brought to you for free and open access by Lehigh Preserve. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of Lehigh Preserve. For more information, please contact preserve@lehigh.edu.

AUTHOR:

Nkansah, Franklin Daniel

TITLE:

Quantification of Thin Gate Oxide Impact on the Switching Characteristics of MOSFETs

DATE: May 30, 1993

QUANTIFICATION OF THIN GATE OXIDE IMPACT ON THE SWITCHING CHARACTERISTICS OF MOSFETs

by

Franklin Daniel Nkansah

A Thesis
Presented to the Graduate Committee
of Lehigh University
in Candidacy for the Degree of
Master of Science
in
Electrical Engineering

Lehigh University
1993

Certificate of Approval

This thesis is accepted and approved in partial fulfillment of the requirements for the degree of Master of Science

3/9/93

Date

Professor in Charge

Chairman of Department

ACKNOWLEDGEMENTS

I would like to thank Professor D.R. Young, W.Y. Yarbrough, W.T. Cochran, J.A. Shimer and I.C. Kiziyalli for their genial support and guidance during this thesis work. I am especially grateful to my wife Mercy and daughter Claribel for their love and patience. In addition I would like to acknowledge the encouragement of K.R. Olasupo, Y.S. Obeng and D. Alugbin. I would like to sincerely thank R. Dyas, T. Kook and R. Key their assistance in growing of the thin oxides, etching and electrical chracterizations.

I would also like to thank AT&T Bell Laboratories for providing the funds and silicon resources that made this work a reality. Finally, I dedicate this thesis to my parents for their long and hard struggle in Liberia.

TABLE OF CONTENTS

THESIS TITLE PAGE i

CERTIFICATE OF APPROVAL ii

ACKNOWLEDGEMENTS iii

LIST OF TABLES viii

LIST OF FIGURES ix

ABSTRACT 1

1. INTRODUCTION 2

1.1 SCOPE OF THESIS 2

1.2 REVIEW OF OXIDATION MODEL 4

1.2.1 SPACE CHARGE EFFECTS 4

1.2.2 OXIDE STRUCTURAL EFFECTS 7

1.2.3 STRESS IN SILICON DIOXIDE 9

1.2.4 OXYGEN SOLUBILITY IN THE OXIDE 10

1.3 SILICON OXIDATION IN THE THIN REGIME 11

1.4 THIN OXIDE EFFECTS ON MOSFETs 13

1.4.1 EFFECTS ON MOBILITY 14

1.4.2 EFFECTS ON OXIDE CHARGE 15

1.4.3 BORON PENETRATION THROUGH THIN GATE OXIDES 18

2. THEORY 21

2.1 THEORY OF THIN DIELECTRIC OXIDATION 21

2.1.1 BASIC OXIDATION MODEL 21

2.2 EFFECTS OF FLUORINE 26

2.2.1 INTERACTION WITH INTERFACE STATES 27
2.2.2 FLATBAND VOLTAGE EFFECTS 30
2.2.3 Q_{BD}, V_{BD} AND AGING EFFECTS 31
2.3 EFFECTS OF NITRIDED OXIDES ON DEVICES 33
2.3.1 KINETICS OF NITRIDATION 33
2.3.2 NITRIDATION EFFECTS 35
2.4 EFFECTS OF POST OXIDATION ANNEAL (POA) 36
2.5 CHARACTERIZATION OF THIN OXIDES 38
2.5.1 INTERFACE TRAP CHARGE 38
2.5.2 FIXED OXIDE CHARGE 39
2.5.3 MOBILE IONIC CHARGE 40
2.5.4 OXIDE TRAPPED CHARGE 40
2.5.5 FOWLER NORDHEIM TUNNELING MECHANISM 41
2.6 TIME DEPENDENT DIELECTRIC BREAKDOWN 44
2.6.1 INTRINSIC BREAKDOWN MODEL 45
2.6.2 DEFECT RELATED BREAKDOWN MODEL 46
2.6.3 TDDB PREDICTION 50
2.6.4 RELIABILITY PREDICTION 51
2.6.5 FAILURE RATE 53
2.7 THE MOS CAPACITOR 55
2.8 MODES OF OPERATION 56
2.8.1 FLATBAND 56
2.8.2 ACCUMULATION 56
2.8.3 DEPLETION AND INVERSION 57
2.9 FIELD EFFECT AT THE SEMICONDUCTOR SURFACE 57
2.10 CAPACITANCE CHARACTERIZATION 60

2.10.1 C_{GS} DEPENDENCE ON GATE VOLTAGE 60
2.11 FRREQUENCY DEPENDENCE OF CAPACITANCE 61
2.11.1 C_{GS} AT LOW FREQUENCY 61
2.11.2 C_{GS} AT HIGH FREQUENCY 62
2.12 THE MOS TRANSISTOR 63
2.12.1 BAND STRUCTURE AT DRAIN EDGE 64
2.12.2 CHANNEL CROSS-SECTION OF NMOSFET 66
2.12.3 LONG CHANNEL NMOSFET MODEL 67
2.12.4 SHORT CHANNEL NMOSFET MODEL 69
2.12.5 NARROW WIDTH EFFECT 71
2.12.6 DRAIN-INDUCED BARRIER LOWERING 73
2.12.7 SUBTHRESHOLD REGIME 74
2.13 HOT CARRIER EFFECTS ON RELIABILITY 76
2.13.1 SUBSTRATE CURRENT 77
2.13.2 SYMPTOMS OF DEGRADATION 79
2.14 GATE OXIDE THICKNESS LIMITATIONS FOR ULSI 81
2.14.1 STATIC LOGIC LIMITATIONS 82
2.14.2 GATE OXIDE TUNNELING 85
2.15 MOS CAPACITANCE EFFECTS ON SWITCHING CHARACTERISTICS 88
2.15.1 DEVICE CAPACITANCE 88
2.15.2 PARASITIC CAPACITANCE 92
2.16 INVERTER DELAY CHARACTERIZATION 93
2.16.1 FALL TIME DETERMINATION 94
2.16.2 RISE TIME DETERMINATION 96
3. EXPERIMENT 97

3.1 INTRODUCTION 97

3.2 DEVICE SIMULATION 98

3.3 CMOS DEVICE FABRICATION SEQUENCE 99

3.3.1 NTUB FORMATION 99

3.3.2 PTUB FORMATION 100

3.3.3 DEVICE ISOLATION 101

3.3.4 THRESHOLD ADJUST IMPLANT 102

3.3.5 GATE DEFINITION 103

3.3.6 LDD AND SPACER FORMATION 104

3.3.7 SELECTIVE N^{+}/P^{+} S/D FORMATION 105

3.3.8 SELF-ALIGNED SILICIDATION 106

3.3.9 INTERLEVEL DIELECTRIC, CONTACT, METALLIZATION AND PASSIVATION 107

3.4 TEST STRUCTURES 109

4. RESULTS AND CONCLUSIONS 113

4.1 DELAY CHARACTERIZATION 113

4.2 ELECTRICAL CHARACTERIZATION 118

4.3 DEVICE AGING CHARACTERIZATION 132

REFERENCES 143

LIST OF TABLES

TABLE 1: Summary of C-V measurement data for various gate oxide thicknesses 123

TABLE 2: Ion/Ioff data for NMOS of $L_{CODED} = 0.75\mu m$ as a function of gate oxide thickness. 131

TABLE 3: Ion/Ioff data for PMOS of $L_{CODED} = 0.85\mu m$ as a function of gate oxide thickness. 131

TABLE 4: Summary of Device degradation for NMOS of $L_{CODED} = 0.75\mu m$ as a function of V_{ds} and gate oxide thickness 141

Figure 1: Enhanced transport of ionized oxidizing species during the initial stages of oxidation in dry oxygen by field enhanced diffusion.

Figure 2a: Micropores in thin oxides affecting the transport oxygen species to the interface.

Figure 2b: Modeling oxidation-rate enhancement in the thin regime by adding additional processes in parallel to those described by the linear-parabolic model.

Figure 3: Interface states effect on device mobility.

Figure 4a: Drain current versus drain voltage—experiment. Solid lines: pre-stress; dashed lines: post-stress normal; dashed-dotted lines: post-stress reverse.

Figure 4b: Measured substrate current. Solid lines: pre-stress; dashed lines: post-stress normal; dashed-dotted lines: post-stress reverse.

Figure 4c: Drain current versus drain voltage—experiment. Solid lines: pre-stress; dashed lines: post-stress normal; dashed-dotted lines: post-stress reverse stress conditions.

Figure 4d: Measured substrate current. Solid lines: pre-stress; dashed lines: post-stress normal; dashed-dotted lines: post-stress reverse.

Figure 5c: Flat-band voltages for boron- and BF_2-implanted devices as a function of 900 C anneal time.

Figure 5a: Boron penetration through the gate oxide.

Figure 5b: V_{FB} as a function of gate oxide thickness for n-channel and BF_2-implanted p-channel capacitors.

Figure 6: Oxide-thickness dependence of the excess in growth rate for the oxidation of lightly doped (100) silicon in dry oxygen in the 800 –1000 C range.

Figure 7c: Si dangling bond, Si-Si bond, and O vacancy level at the interface. These energy levels move in the lower half of the Si band gap with changing bond length d. Open and closed circles denote O and Si atoms respectively.

Figure 7a: Basic model constructed with amorphous SiO_2 represented by Bethe-lattice and Si substrate with (111) orientation. Open and closed circles denote O and Si atoms respectively.

Figure 7b: Perfect interface and oxygen dangling bond at the Si-SiO_2 interface having no energy level in the range between 0.5 eV below the top of the valence band and 0.5 eV above the bottom of the conduction band of Si. Open and closed circles denote O and Si atoms respectively.

Figure 8b: Impurity at the Si-SiO_2 interface. If any of H, OH, Cl, and F is bonded to the Si dangling bond, no energy level exists in the energy range between 0.5 eV below the top of the valence band and the 0.5 eV above the bottom of the conduction band of Si. Open and closed circles denote O and Si atoms respectively.

Figure 8a: Commonly observed U-shaped distribution of interface-trap state density in the forbidden gap of Si.

Figure 9: Change in flat-band voltage of fluorine enriched SiO_2 vs. the fluorine implant dose.

Figure 10a: Breakdown field of fluorine rich oxides as a function of implant dose.

Figure 10b: Charge-to-breakdown of fluorine rich oxides as a function of implant dose.

Figure 11: Transistor lifetime until 10% degradation in g_m vs. reciprocal of drain voltage.

Figure 12: Charge to breakdown of nitroxide dielectrics.

Figure 13a: Breakdown field for MOS capacitors with and without post-oxidation anneals.

Figure 13b: The effect of gate dielectric formation conditions on surface mobility. OX: pure oxide grown at 1050 C for 35 seconds. NO: nitroxide formed at 1150 C for 24 seconds in a 25% NH_3 ambient. ONO: NO reoxidized at 1150 C for 20 or 40 seconds. AONO 30: ONO annealed an additional 30 seconds in argon at 1150 C. AONO 60: ONO annealed an additional 60 seconds in argon at 1150 C.

Figure 13c: Hot electron lifetime of annealed NO MISFETs.

Figure 14: Fowler-Nordheim Tunneling DiagramFigure 15a: Dependence of the leakage current on the oxide thickness. The oxide thickness was: 5.1 nm (curve a), 6.0 nm (curve b), 7.5 nm (curve c), and 9.7 nm (curve d).

Figure 15a: Dependence of the leakage current on oxide thickness. The oxide thickness was: 5.1 nm (curve a), 6.0 nm (curve b), 7.5 nm (curve c), and 9.7 nm (curve d).

Figure 15b: Increase of the low-field leakage current as a function of electron fluence for a capacitor. The stress was: 2.5 V (curve a), 3 V (curve b), 3.5 V (curve c), 4 V (curve d). [70], and 4.5 V (curve e).

Figure 16: log (t_{BD}) versus 1/Eox for a 79Å gate oxide.

Figure 17a: Oxide defects modeling as effective oxide thinning by DXox.

Figure 17b: An illustration of the dependence t_{BD} on the severity of defects DXox as a function of the oxide field Eox.

Figure 18a: Time-dependent dielectric breakdown (TDDB) data for an oxide of 107Å under 13.2 V plotted in cumulative percent failure versus t_{BD} or equiva lently versus DXox.

Figure 18b: TDDB data for an oxide of 107Å under various stress voltages. Good agreement is obtained with the ΔXox model (solid curves) through the equation $t_{BD2} = t_{BD1}(t0/t_{BD1})(1\text{-}Vox1/Vox2)$.

Figure 19: Defect density as a function of ΔXox, D(ΔXox) using Gamma distribution for the data in Figure 18a. Once D(ΔXox) is derived for a given process, many reliability parameter such as yield, failure rate, and optimum screen condition can be determined.

Figure 20: Cumulative percentage failure versus time for (a) different areas and (b) different stress voltages of 107Å oxide. Solid curves are derived from the defect density and DXox models.

Figure 21: Failure rate versus time for (a) different areas and (b) different stress volt ages of 107Å oxide. Solid curves are derived from the defect density and ΔXox models.

Figure 22: Theoretical curves of failure rate and cumulative percent failure for different areas at 5 V operation. These curves are used for screening and screen yield optimization.

Figure 23a: An MOS (Metal Oxide Silicon) capacitor.

Figure 23b: Energy band diagrams for ideal MOS capacitors for: (b) accumulation; (c) depletion; and (d) inversion.

Figure 24: MOS capacitance-voltage curves. (a) low frequency; (b) high frequency; and (c) deep depletion.

Figure 25: An MOS Field Effect Transistor (FET).

Figure 26: Energy band diagrams near the drain of NMOSFET with different gate and drain bias conditions: (a) Vg > Vt, Vd = 0; (b) Vg < Vt(Vd), Vd > 0; and (c)Vg > Vt(Vd), Vd > 0.

Figure 27: Cross-section of N-channel band structure.

Figure 28: Graphical representation of "Charge Sharing Phenomena".

Figure 29: Realistic picture for a device with bird's beak effect on active area

Figure 30: Shape of the depletion region in a short-channel MOSFET at zero and high drain bias. The grey areas indicate the charge shared by the gate and the junctions at zero drain bias.

Figure 31: A cross-sectional schematic of a transistor illustrating subthreshold conduction.

Figure 32a: Schematic of hot-electron effects in an n-channel MOSFET.

Figure 32b: Schematics of channel potential and horizontal electric field of a MOSFET biased at saturation region.

Figure 33a: ΔV_{th} increases as t_n with n ~0.65 in this figure.

Figure 33b: Degradation characteristics after stressing.

Figure 33c: Lifetime as a function of substrate current, lifetime defined as 10 mV shift in threshold voltage (V_{th}).

Figure 34: Lateral electric field distributions associated with LDD structures.

Figure 35a: Transfer voltage characteristics of a CMOS inverter with the critical volt ages V_{OH}, V_{IL}, V_{IH}, and V_{OL} labeled where V_E is the only point at which the NMOS and PMOS transistors are both saturated and the output voltage is equal to Vdd/2.

Figure 35b: Noise margin definitions.

Figure 36: Dependence of Noise Margin on Leakage Current I_L/I_E.

Figure 37: Gate voltage for given current densities and oxide thicknesses.

Figure 38a: Effect of leakage current on the voltage-transfer characteristics of an un-loaded inverter.

Figure 38b: Simulated effect of leakage current on the switching speed of an inverter. The gate has a fan-out of 3 and each gate is driving a 1 pF line capacitance.

Figure 39: Definitions of device and parasitic capacitances.

Figure 40: Small-signal capacitances vs V_{DS} for a MOSFET in linear and saturation regions: (a) $V_{SB} = 0$; (b) $V_{SB} = 5$ V.

Figure 41: CMOS inverter circuit.

Figure 42: CMOS inverter switching characteristics.

Figure 43: Device simulations flow diagram.

Figure 44: N-tub formation.

Figure 45: P-tub formation.

Figure 46: Active device area (Thinox) formation.

Figure 47: High pressure oxidation (HIPOX) for field oxide formation.

Figure 48: Threshold adjust ion implantation.

Figure 49: Polysilicon gate definition.

Figure 50: LDD formation.

Figure 51: Spacer formation.

Figure 52: Selective N^+ and P^+ S/D ion implantation.

Figure 53: Self aligned selective silicidation.

Figure 54: Dielectric I, Contact opening and Aluminum I metallization.

Figure 55: Dielectric II, Window II, Aluminum II and Passivation CAP.

Figure 56: Test structure for I-V and aging characterization.

Figure 57b: Schematic diagram of 89 cascaded NAND chain.

Figure 57a: Schematic diagram of 101 cascaded inverter chain.

Figure 58: Test structure for performance characterization.

Figure 59: INRB delay/stage as a function of gate oxide thickness.

Figure 60: ND2 delay/stage as a function of gate oxide thickness.

Figure 61: Inverter derating factor as a function of operating temperature.

Figure 62: C-V characteristics for 150Å gate oxide.

Figure 63: C-V characteristics for 130Å gate oxide.

Figure 64: C-V characteristics for 110Å gate oxide.

Figure 65: C-V characteristics for 90Å gate oxide.

Figure 66: C-V characteristics for 70Å gate oxide.

Figure 67: Breakdown voltage versus Tox.

Figure 68: Charge to breakdown versus Tox.

Figure 69: Snapback voltage versus Tox.

Figure 70a: I_{DS} vs V_{DS} characteristics for 150Å gate oxide thickness.

Figure 70b: Subthreshold characteristics for 150Å gate oxide thickness.

Figure 71a: I_{DS} vs V_{DS} characteristics for 130Å gate oxide thickness.

Figure 71b: Subthreshold characteristics for 130Å gate oxide thickness.

Figure 72a: I_{DS} vs V_{DS} characteristics for 110Å gate oxide thickness.

Figure 72b: Subthreshold characteristics for 110Å gate oxide thickness.

Figure 73a: I_{DS} vs V_{DS} characteristics for 90Å gate oxide thickness.

Figure 73b: Subthreshold characteristics for 90Å gate oxide thickness..

Figure 74a: I_{DS} vs V_{DS} characteristics for 70Å gate oxide thickness.

Figure 74b: Subthreshold characteristics for 70Å gate oxide thickness..

Figure 75: Substrate current as a function of gate voltage for variable Lgate and Tox.

Figure 76: Substrate current as a function of gate voltage for variable Lgate and Tox.

Figure 77: Substrate current as a function of gate voltage for variable Lgate and Tox.

Figure 78: Substrate current as a function of gate voltage for variable Lgate and Tox.

Figure 79: Substrate current as a function of gate voltage for variable Lgate and Tox.

Figure 80: 10% transconductance (g_m) degradation as a function of Isub for 150Å gate oxide.

Figure 81: 10% transconductance (g_m) degradation as a function of Isub for 130Å gate oxide.

Figure 82: 10% transconductance (g_m) degradation as a function of Isub for 110Å gate oxide.

Figure 83: 10% transconductance (g_m) degradation as a function of Isub for 90Å gate oxide.

Figure 84: 10% transconductance (g_m) degradation as a function of Isub for 70Å gate oxide.

ABSTRACT

The characterization of the effect of thin oxides on Metal Oxide Semiconductor performance has been completed. This encompassed gate oxide thickness from 150Å to 70Å which is the regime of increasing interest for sub-micron CMOS technologies. MOS device characteristics such as gain, current drive capability dependence on gate oxide thickness were also characterized. Device performance characterization was completed by using the ring oscillators with 101 and 89 stages to demonstrate the adverse impact of the gate capacitance on switching performance due to decreased gate oxide thickness, and a minima for gate oxide scaling in relation to switching performance was established. Process and Device simulators were used to fully investigate the various capacitance's associated with the ring-oscillator delay, which are contributing factors to the overall gate delay. Devices were characterized for hot carrier degradation and the lifetime due to increasing substrate currents were quantified as a function of gate oxide thickness. The aging characterization established that device degradation worsened as gate oxide was thinned.

CHAPTER 1

INTRODUCTION

Thermally grown layers of silicon dioxide has been an integral part of silicon device technology since their first applications in surface protection and selective masking and stabilizing silicon surfaces[1]. Silicon dioxide layers found additional applications in device isolation, impurity gettering, masking against impurities, junction passivation and insulation between metal layers[2]. Such applications have made silicon oxidation a vital processing step in both bipolar and MOS technologies. Its most important application, however, is that of the gate-dielectric material for MOS transistors where SiO_2 layers become an active component. As the VLSI era enters into the sub-micron regime, thin gate oxides must be grown routinely and reproducibly with high yield and long-term reliability.

Much work has been done to achieve uniform, thin oxides under controlled processing conditions. In order to controllably grow thin oxides, the growth rate must be reduced, thus allowing the process a reasonable time of growth. Various techniques such as dry oxidation, dry oxidation with HCL, trichloroethylene (TCE) or trichloroethane (TCA), reduced pressure oxidation, low temperature high pressure oxidation and rapid thermal oxidation have been investigated.

1.1 SCOPE OF THESIS

As gate oxide thickness is reduced to further meet the current drive capability of the ULSI MOS devices, there is a minima in oxide thickness, beyond which the increase in

gate capacitance due to decreased oxide thickness causes a decrease in the drive capability, which manifest itself as an increase in the ring oscillator delay time.

The work presented in this thesis will experimentally investigate the effects of thin gate oxides on MOS Device characteristics, and the impact on switching of digital systems. The characterization of thin oxides will include the use of high and low frequency C-V measurements to quantify the flatband voltage and recovery rate of the capacitance for the various oxide thicknesses. In addition to capacitance quantification, high field measurements will be used to investigate the breakdown voltage and the tunneling mechanisms of thin oxides. Electric field and temperature effects of Time-Dependent Dielectric Breakdown (TDDB) implications will be discussed. Hot Carrier Aging of MOS transistors will be used to examine the reliability implications of thin gate oxide dielectric for ULSI applications.

The gate oxide dielectric will be varied from 150Å to 70Å. Since the substrate doping in the MOS transistor channel had to be modified to achieve the proper threshold voltage, the implications of these changes to device I-V characteristics will be quantified. Furthermore, I will unify the correlation of thin oxide impact on device performance with the experimentally observed ring oscillator propagation delay per stage measurements.

This thesis chronology begins with the development of thin oxidation model, which will be followed by theoretical development of implications of thin gate oxide on MOSFET switching characteristics which encompassed the MOS capacitor and transistor The experimentation procedures, transistor modeling and advanced fabrication techniques are then discussed. Finally the results obtained are discussed and conclusions established to support the "minima" concept encountered in gate oxide thickness dependence on propagation delay.

1.2 REVIEW OF OXIDATION MODELS

Several models have already been proposed in an attempt to explain the initial fast regime of gate oxidation kinetics. These models fall into four major groups (i) space-charge effects where the enhancement is electrochemical in nature, (ii) oxide structure effects that provide additional oxidant transport, (iii) oxide stress effects influencing the diffusivity of oxidizing species in the oxide and the surface-reaction rate constant at the interface, and (iv) oxygen solubility considerations in the oxide where growth would be enhanced due to oxygen concentrations exceeding its solid solubility limit.

1.2.1 SPACE - CHARGE EFFECTS

A number of electrochemical mechanisms have been proposed to enhance the oxidation process in the thin oxide regime[4-8]. The transport of oxidant species by thermal diffusion has been proposed to be affected by an electric field across the oxide in the early stages of growth. Field-assisted oxidant diffusion was originally proposed by Cabrera et al[3] to account for metal-oxidation results. This phenomenon was proposed by Deal and Grove[3] to account for silicon-oxidation results in dry oxygen in the thin oxide regime. Theoretical analysis of the kinetics of metal oxidation emphasize only two types of rate-determining oxidation mechanisms[3]. In the first, the oxide thickness is small in comparison to the space charge regions within the oxide and the growth kinetics are strongly influenced by the voltage drop across the oxide film caused by contact-potential differences. In the second, the diffusion rate of either the oxidizing species or the metal across the oxide determines the oxidation kinetics; thus the growth is purely parabolic. The space-charge effect involves electrons readily available in the metal. These electrons penetrate the oxide either by tunneling or by thermionic emission into its conduction band and then fill the surface states provided by the adsorbed gas molecules at the oxide-gas interface. A steady -state condition is established in which the net electronic current is

close to zero, thereby equalizing the surface-state level and the Fermi level of the metal. This process results in a constant contact potential difference between the two interfaces and an internal electric field. This field is assumed to be strong enough to alter significantly the activation barrier for the diffusion of ions or other charged defects. As a result, field-dependent transport of ionic defects constitutes the rate-limiting step. The model of Cabrera et al [3] yielded an inverse-logarithmic growth law.

The rapid initial oxidation regime in dry oxygen and its absence in wet oxygen or steam have been interpreted by Deal and Grove [1] to be a result of space-charge effects. In their interpretation, the oxidizing species is ionic, and oxidation is enhanced until the oxide thickness becomes larger than the extent of the space-charge region within the oxide. The thickness is on the order of the extrinsic Debye length in the oxide

$$L_{Dox} = \sqrt{\frac{(kTk_{ox}\varepsilon_0)}{(2q^2C^*)}}$$

In dry oxygen, $C^*=5x10^{16}cm^{-3}$, k_{ox}=3.9 and therefore, at 1000°C $L_{Dox}=150$Å which correspond well with the oxide thickness beyond which the rapid initial-growth mechanism stops being effective. In contrast, for wet SiO_2 at 1000°C, using $C^*=3x10^{19}cm^{-3}$, $L_{Dox}=5$Å. Based on the linear pressure dependence of the parabolic rate constant B, Deal and Grove [1] concluded that no dissociation occurs at the oxide-gas interface, which implies that the diffusing species are undissociated molecules of oxygen or water for dry and wet SiO_2 oxidation, respectively. Coupled with the findings of Jorgensen [4], where an electric field was observed to affect the oxidation of silicon, they suggested that the diffusing species in dry oxygen is a singly ionized oxygen molecule O_2^-. The field created by the space-charge of the negatively charged oxygen ions should be directed toward the gas-oxide interface, as illustrated in fig. 1a.

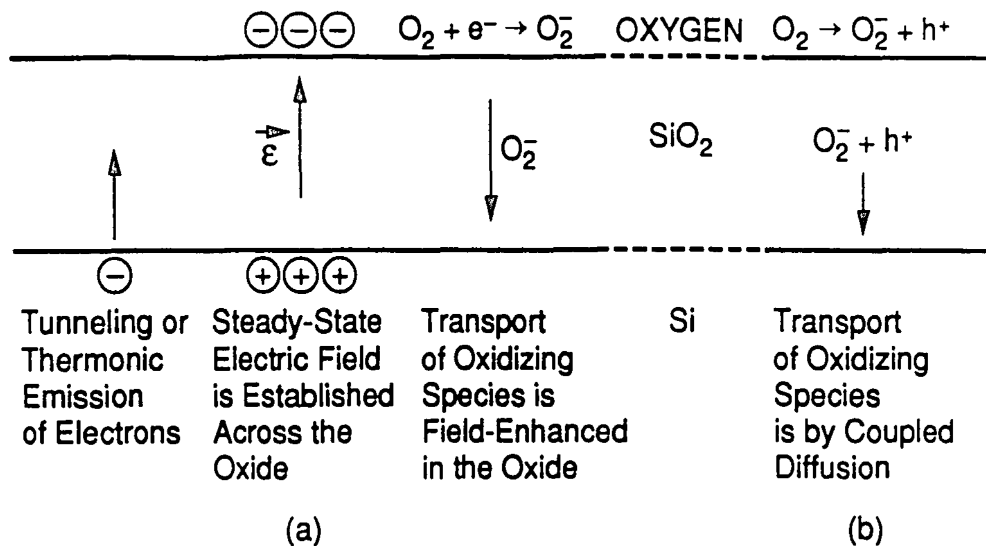


Figure 1: Enhanced transport of ionized oxidizing species during the initial stages of oxidation in dry oxygen by field enhanced diffusion (a) and by coupled diffusion (b). [18]

The transport of the oxidizing species could also be enhanced by a slightly different mechanism [5], illustrated in fig.1b. Here, it is assumed that as a molecule of oxygen from the gas is absorbed at the outer surface, and after entry into the oxide, it dissociates into a negatively charged oxygen molecule O_2^- and a hole. Holes, diffusing more rapidly than O_2^-, move toward the SiO_2 interface and drag the slower oxygen ions with them by the built-in electric field that results from the coupled motion of two charged species with different mobilities. This field aids the motion of the slower species.

Hamasaki [6] proposed that negatively charged oxidant species are prevented from reaching the SiO_2 interface by the oxide field near the interface caused by the positive oxide fixed charge Q_f generated by the oxidation reaction. In this description electrons reached the outer surface of the oxide by thermionic emission. The initial rapid growth is explained by oxides growing at a fast rate before a sufficient amount of Q_f has been generated by the oxidation reaction. Once formed, the oxide fixed charge establishes an

electric field that opposes the transport of negatively charged oxidizing species toward the interface. This description implies that the linear rate constant B/A is a measure of the surface reaction when the oxidizing species are under the influence of the Q_f induced retarding field. For the same description of the oxidation process, Lu and Cheng [7] further assumed a mathematical form for the distribution of the oxide fixed charge and derived the oxidation rate constants corresponding to the thin (< 300Å) and thick regimes.

In analyzing the less-than-linear pressure dependence of the linear rate constant B/A, Hu [8,9] proposed an intermediate oxygen-chemisorption step where chemi-sorbed molecular oxygen may directly oxidize silicon at a slow rate or dissociate slowly into atomic oxygen which then rapidly oxidizes silicon. The extra growth rate in the thin regime was proposed to be the result of a possible parallel flux of negatively ionized atomic oxygen (O^-) becoming more important for small thicknesses. Because of its smaller size, O^- ions are expected to diffuse substantially faster than oxygen molecules. The free energy of O^- at 1200K is 1.1eV [8]. The diffusion of ionized atomic oxygen is argued to be possible only if there is substantial electronic conduction by tunneling through the oxide [8]. As a result, this mechanism would stop for oxide thicknesses larger than 150Å. The above mentioned models based on space-charge effects predict an inverse-logarithmic relationship between oxide thickness and oxidation time.

1.2.2 OXIDE STRUCTURAL EFFECTS

A number of proposals have been made to relate the fast oxidation rate in the thin regime to the oxide structure. Revesz and Evans [9] proposed that SiO_2 films are non-crystalline and may contain structural microheterogeneities, especially micro channels, along which diffusing species may transport preferentially. This is illustrated in fig 2. It has been

crystalline and may contain structural microheterogeneities, especially micro channels, along which diffusing species may transport preferentially. This is illustrated in fig 2. It has been

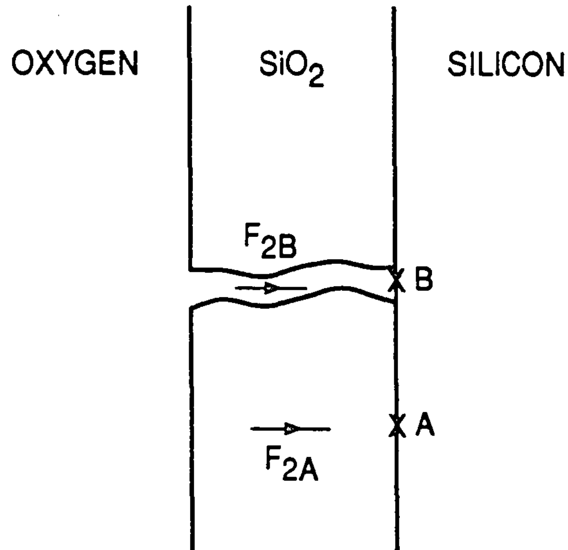


Figure 2a: Micropores in thin oxides affecting the transport oxygen species to the interface. [18]

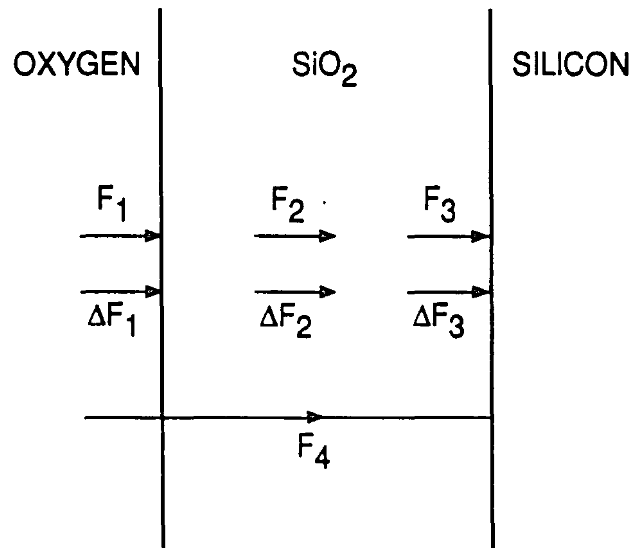


Figure 2b: Modeling oxidation-rate enhancement in the thin regime by adding additional processes in parallel to those described by the linear-parabolic model. [18]

demonstrated [10] that the room temperature gaseous permeability of thermally grown SiO_2 layers can be considerably higher than in fused silica, and this increase was attributed to microchannels whose diameters are less than 50Å.

The formation of channels has been proposed to be closely related to the mechanism of oxide growth in that both parabolic rate constant and the extent of the fast initial regime are higher in the <111> than in the <100> orientation of silicon, provided that the H_2O content in the oxidizing ambient is very low [11]. This behavior of SiO_2 films occurs when the channels are aligned perpendicularly to the silicon surface. The structure of thermally grown films, including the $Si-SiO_2$ interface, is determined by combination of ordering and disordering effects. Increased ordering enhances the diffusivity of O_2 molecules along structural channels. The presence of H_2O in the oxidizing ambient has a great disordering effect on structure of SiO_2 because it interacts with SiO_2 more strongly than does O_2 whose behavior is similar to that of inert gases. As illustrated in fig. 2 growth-rate enhancement occurs as a result of localized points such as point B where microchannels or structural defects meet the $Si-SiO_2$ interface. The oxidation rate at these points is enhanced as a result of localized high concentration of oxygen in direct contact with silicon rather than in faster transport of oxygen through the oxide which would not affect the oxidation reaction because oxide growth in the linear regime is rate limited by surface reaction. The lateral transport of high concentration of oxidant species at the $Si-SiO_2$ interface, as a result of micropores in the oxide, may be responsible for localized oxidation rate enhancement around micropores and microchannels in the thin regime.

1.2.3 STRESS IN SILICON DIOXIDE

Stress measurements in silicon dioxide have been made at room temperature [12,13]and at growth temperatures [14]. Stress in the oxide measured at room

temperature is a result of the mismatch between the thermal expansion coefficients of Si and SiO_2, and it develops as the oxidized wafer cools down from the oxidation temperature to room temperature. Intrinsic stresses resulting from the mismatch in the molar volume of the Si and SiO_2 exists at the oxide growth temperature. When present, the intrinsic stresses are large enough to cause plastic deformation of the silicon substrate at growth temperatures. The temperature dependence of the linear and parabolic rate constants above and below the viscous flow temperature was discussed elsewhere[2]. These results suggested that stress in the oxide affects the transport of oxygen through the oxide and its reaction with silicon at the $Si - SiO_2$ interface. A new formalism that explains the low and high temperature dependence of the linear rate constant on the intrinsic stress was proposed by Irene[15]. Intrinsic stress effects on the surface reaction rate constant are a possible source of oxidation-rate enhancement in the thin film regime.

In a study of silicon oxidation in dry oxygen in the thin film regime, Fargeix and co-workers[16] suggested that the diffusivity of the oxidizing species is not constant across the oxide layer as a result of the influence of stress in the growing film. They proposed that diffusion through the first hundreds of angstroms of the oxide layer is slowed down due to compressive stresses in the SiO_2 film. The increase in diffusivity far from the oxidizing interface is presumably related to the relaxation of the stresses. These stresses were assumed to be maximum at the $Si - SiO_2$ interface and decreased monotonically with the characteristic distance from the interface and towards the free surface of the oxide.

1.2.4 OXYGEN SOLUBILITY IN THE OXIDE

According to Henry's law, the equilibrium concentration C* of the oxidant in the oxide is related to the partial pressure Pg of the oxidant species in the gas (C*=HPg). The relationship holds only in the absence of dissociation or recombination of the oxidant

species at the oxide-gas interface and when their chemical potential in the oxide is independent of pressure. Increased solubility of O_2 in SiO_2 for very thin layers may be important in describing the fast initial dry oxidation of silicon[18] because it would reflect an increase in the constant of proportionality in Henry's law when the oxide thickness is less than the mean spacing of oxygen solute molecules dissolved in the oxide.

To obtain a more realistic description of oxides thinner than the average spacing between oxidant molecules, a two dimensional structure, wherein solute molecules are packed, was suggested by Derbenwick and Anderson[18]. The oxidation rate is directly proportional to the concentration of oxidizing species at the $Si - SiO_2$ interface for films with thinner than 270Å oxide, which implies a proportionally faster rate. A simple model to test this solubility concept was proposed[33] based on the assumption that the effective solubility of oxygen into the oxide decreases exponentially in thin oxides. This resulted in a growth law of the form[33]

$$X_{ox}(T_{ox}) - X_{ox}(0) - L_0 \ln \frac{[c^* + C_0 e^{-x_{ox}(0)/L_0}]}{[c^* + C_0 e^{-x_{ox}(0)/L_0}]} = \frac{(B)}{(A)} T_{ox}$$

where $X_{ox}(0)$ is the native-oxide thickness at the start of oxidation ($T_{ox} = 0$), C* the equilibrium bulk solubility, $(C^* + C_0)$ the equilibrium concentration at the surface, and L_0 a characteristic length related to the equilibrium spacing of the solute molecules.

1.3 SILICON OXIDATION IN THE THIN REGIME

In searching for the physical mechanisms responsible for the oxidation-rate enhancement in the thin regime for a surface-reaction rate-limited process, one must investigate the possible ways by which the two components of the surface reaction, namely, the concentration of oxygen species and the contribution of the silicon substrate, can be influenced. In other words one should investigate the mechanisms that would affect the flux of oxidant species F_3 consumed in the surface reaction at the $Si - SiO_2$ interface

This flux is expressed as $F_3 = k_s C_i$ where k_s is the surface-reaction rate constant, representing the contribution of the substrate and C_i is the concentration of oxidizing species at the interface. According to Deal-Grove[1], the oxidation is surface-reaction rate limited for thin oxides, and the concentration of oxidizing species is uniformly distributed in the oxide and is at solid-solubility limit.

The analysis of the rate enhancement in the thin regime at lower partial pressures of oxygen indicates that, when the oxygen supply to the surface reaction was varied by changing the oxygen partial pressure in the oxidizing ambient, both the linear surface reaction and the thin regime surface reaction are affected in the same manner. This result reinforces the fact that oxidation in this thin regime is surface-reaction rate limited, and that the rate enhancement in the thin regime would not be affected by any mechanism that would influence the transport of oxidizing species through the oxide, but by some mechanism that affects the concentration of O_2 at the $Si - SiO_2$ interface and or the way in which the silicon substrate contributes to the oxidation reaction. Further studies [18] have revealed that, a surface layer in the silicon substrate containing additional sites for oxidation is responsible for the oxidation rate enhancement in the thin film regime. The concentration of these sites has a profile which decays exponentially with a characteristic length of ~ 30Å. The analysis of SiO_2 growth rate in the thin regime under a variety of experimental conditions such as; substrate orientation, doping, and oxygen partial pressure by Massoud et al [18] can be expressed as

$$\frac{\partial X_{ox}}{\partial t} = \frac{B}{2X_{ox} + A} + C_1 e^{-(x_{ox}/L_1)} + C_2 e^{-(x_{ox}/L_2)}$$

The first term is the linear-parabolic term where B and B/A are the parabolic and linear rate constants, respectively, as defined by Deal and Grove[1]. The two exponential terms represent the rate enhancement in the thin regime, defined in terms of pre-exponential constants C_1 and C_2 and the characteristic lengths L_1 and L_2. The first decaying

exponential has a characteristic length on the order of 10Å, is non zero for the first 40-50Å of oxide growth, and vanishes for oxides thicker than 50Å. The second decaying exponential has a characteristic length on the order of 70Å and is present from the onset of oxidation to an oxide thickness of ~250Å, where it decays to zero and where the growth becomes pure linear-parabolic.

Oxidation-rate enhancement in the thin regime can, in summary be expressed by the exponentially decaying with thickness. Its thickness extent is independent of substrate orientation, doping level, and oxygen partial pressure in the oxidizing ambient. Its magnitude has the same pressure dependence as that of the linear rate constant and is only moderately affected by high phosphorus concentrations in the substrate.

1.4 THIN OXIDE EFFECT ON MOSFETs

The demand for increasing system performance and complexity has acted as a driving force to increase the complexity of monolithic integrated circuits. By shrinking the gate dielectric of these devices, increased drive current and speed are obtained. Each new generation of process introduces thinner dielectrics in order to more strongly couple the gate and channel potentials. As the dielectric is made thinner, gate current and capacitance increases and can limit the usefulness of the MOS transistor by modifying its voltage transfer characteristics in a circuit, or by eliminating its ability to act as an effective latch as employed in Dynamic Random Access Memory circuitry.

One of the main advantages of CMOS devices is that they can be configured in circuits having low standby power dissipation and are able to store charge on capacitors. As the gate leakage current increases, the time required between cycles to refresh the voltage values on a node decreases. The standby gate leakage current and power required to refresh the voltage on the capacitors increases the power dissipation of the chip and system.

Although it is well known that circuit performance can be enhanced by scaling down MOSFET dimensions, in the deep-sub micrometer regime the various design tradeoffs between reducing the channel length L_{eff}, the oxide thickness T_{ox}, and the power supply voltage V_{ds} is still not clear. Changing a particular device parameter such as decreasing oxide thickness to improve current drive can lead to an intolerable degradation in reliability. As a result of the high electric fields in the gate oxide due to device scaling, Fowler-Nordheim (F-N) tunneling current flows through the thin oxides[19], resulting in charge trapping and interface trap generation [20]-[25].

1.4.1 EFFECTS ON MOBILITY

Since the mobility in the inversion layer of MOS devices is a very important factor in determining the performance of devices, it has been studied extensively. In reference to Akizawa et al[26], from the flatband voltage shifts of high-frequency capacitance-voltage (C-V) characteristics of MOS capacitors, the effective interface charge at $Si - SiO_2$, N_{int}, which include both oxide charge and interface state charges, could be estimated. By measuring the drain characteristics of the MOSFETs, the effective mobility was derived from the DC drain conductance measurements and expressed by

$$U_{eff} = \frac{(I_{ds})}{(V_{gs} - V_t)} \frac{L}{W} \frac{T_{ox}}{\varepsilon_{ox}} \frac{(1)}{(V_{ds} - I_{ds}R_s)}$$

where V_{ds} and I_{ds} are the drain voltage and current, V_{gs} and V_t are the gate and threshold voltages, R_s is a parasitic resistance between the source and the drain.

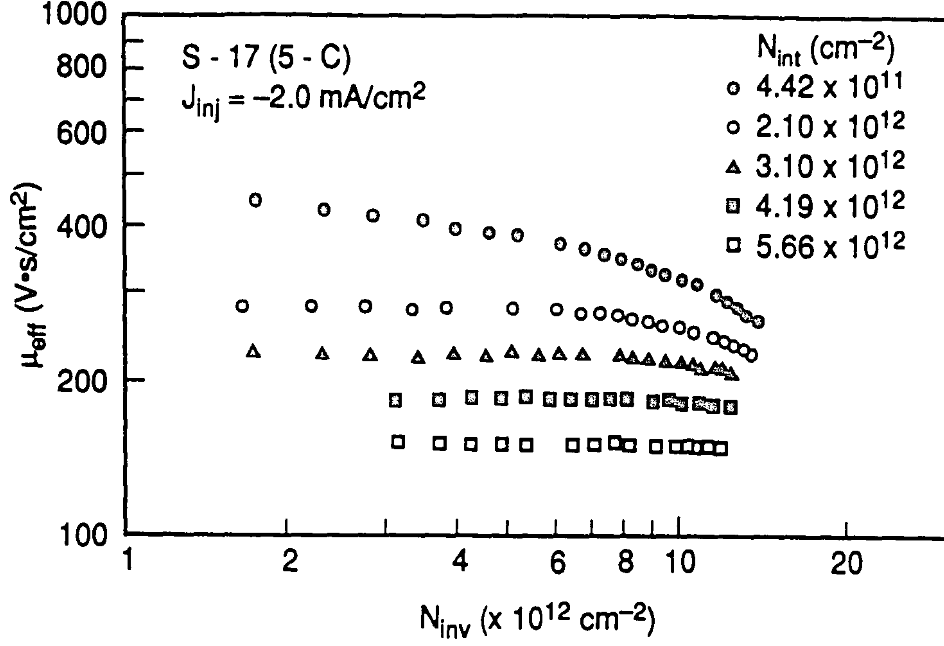


Figure 3: Interface states effect on device mobility. [18]

With the increase of N_{int},U_{eff} lowers over the entire region of N_{inv}, and the dependence of U_{eff} on N_{inv} becomes less with increasing N_{int}. Such behavior of U_{eff} on N_{inv} suggests that the increase of Coulomb scattering due to F-N stress influences the electron scattering predominantly. Since the current circuit design trends have reduced operating voltages less proportionally than the device geometries[27],the device performance improvement by scaling is limited still more fundamentally by the electric field induced mobility degradation.

1.4.2 EFFECT ON OXIDE CHARGE

Device degradation is a severe constraint for the long-term stability of modern VLSI circuits. The shrinking of device dimensions gives rise to high fields at certain bias conditions in the drain region of the MOSFET [28]. There is a general agreement that the observed drain current degradation is due to some localized charges caused by hot-carrier

injection in the high-field regions. Charges localized at the $Si - SiO_2$ interface act as Coulombic centers that scatter the carriers, thus reducing the channel mobility which is observed as a decrease in drain current after post-stress as shown in the figures below

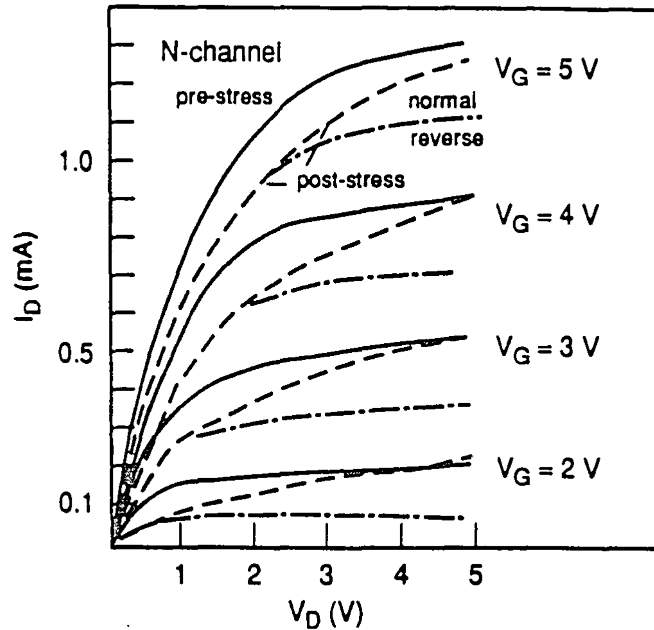


Figure 4a: Drain current versus drain voltage—experiment. Solid lines: pre-stress; dashed lines: post-stress normal; dashed-dotted lines: post-stress reverse. Stress conditions $V_G = 3$ V, $V_D = 8$ V, $I_{STRESS} = 5 \times 10^4$ s. [30]

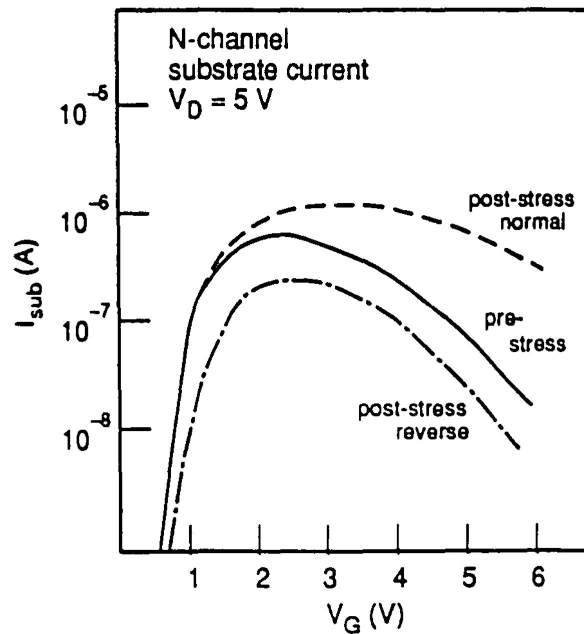


Figure 4b: Measured substrate current. Solid lines: pre-stress; dashed lines: post-stress normal; dashed-dotted lines: post-stress reverse. Stress conditions $V_G = 3$ V, $V_D = 8$ V, $I_{STRESS} = 5 \times 10^4$ s. [30]

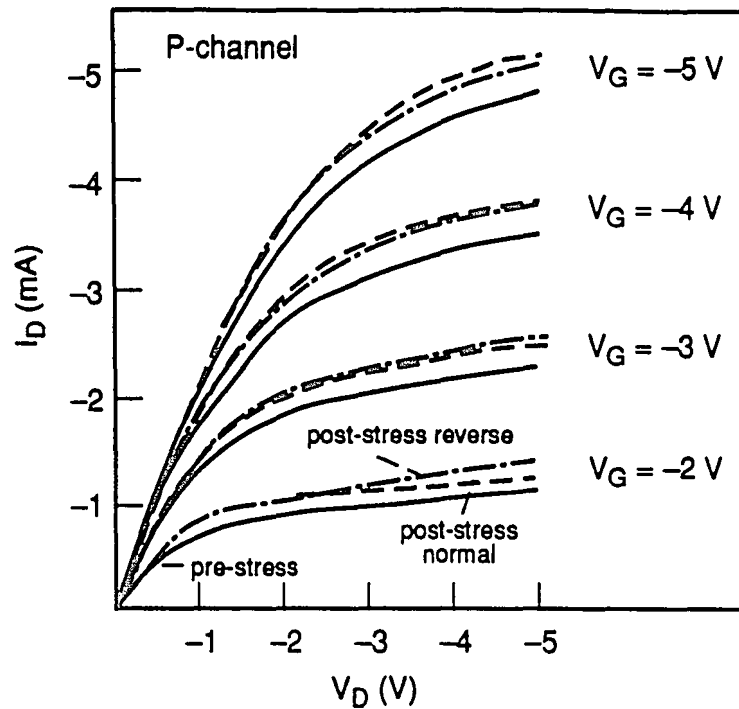


Figure 4c: Drain current versus drain voltage—experiment. Solid lines: pre-stress; dashed lines: post-stress normal; dashed-dotted lines: post-stress reverse. Stress conditions $V_G = -2$ V, $V_D = -8$ V, $I_{STRESS} = 1.5 \times 10^4$ s. [30]

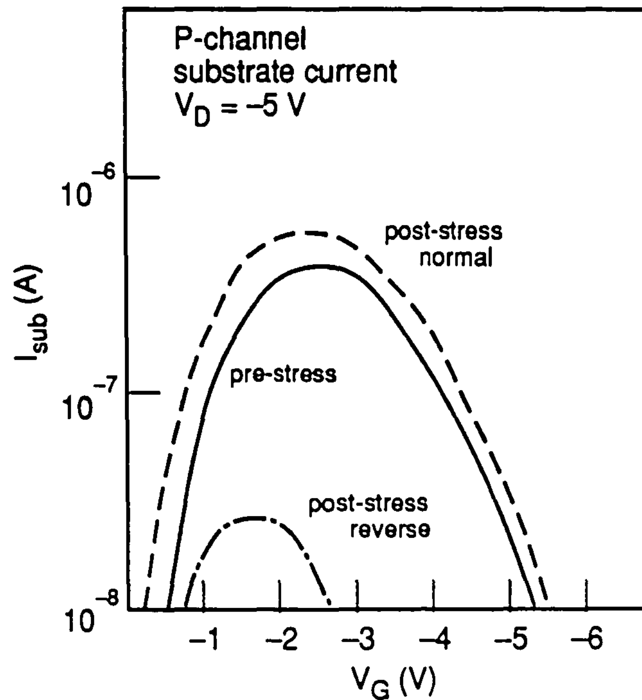


Figure 4d: Measured substrate current. Solid lines: pre-stress; dashed lines: post-stress normal; dashed-dotted lines: post-stress reverse. Stress conditions $V_G = -2$ V, $V_D = -8$ V, $I_{STRESS} = 1.5 \times 10^4$ s. [30]

The post-stress characterization shows a striking asymmetry between normal and reverse (interchanging source and drain) mode currents as in fig. 4a. Therefore one can conclude that the surface charges due to the stress are very inhomogeneously distributed over the channel. This is in good agreement with other studies of inhomogeneous charge distributions at the $Si - SiO_2$ interface[29]. The effect of degradation shifts the bulk current in opposite directions for normal and reverse mode operation, which gives direct access to the sign of the surface charge. Moreover , it should be noted that degradation affects the reverse-mode substrate current strongly over the whole range of gate bias as shown in figs. 4b&4d. In the p-channel devices the drain current and transconductance are increased as shown in fig. 4c. There are several ideas in the literature concerning the nature of stress-induced charges. These can be classified into three distinct categories: (i) Positive charges in fast or slow interface states, with the mobility degradation being responsible for the drain current degradation [30], (ii) Negative charges in slow interface states which is identical to negative charges trapped in the bulk oxide, as long as the potential effect is considered[31-34], (iii) Fast interface states that are occupied according to the position of the quasi-Fermi level in the Si band-gap[29,35-37].

1.4.3 BORON PENETRATION THROUGH THIN GATE OXIDES

PMOS devices which employ n+ doped polysilicon gates typically require a compensating p-type channel implant to lower the threshold voltage. Although this results in buried-channel structures with enhanced hole mobility, submicrometer channel-length devices typically suffer from excessive short channel behavior caused by sub-surface off-current conduction. As a result, P^+ doped polysilicon gates have been proposed for the fabrication of surface- channel PMOS devices which are scalable to deep submicrometer dimensions[38]. Submicrometer PMOS devices fabricated with P^+ polysilicon gates offer improved short-channel behavior similar to that of the NMOS transistor design. Several

disadvantages of using P^+ polysilicon gates in CMOS process include additional process complexity and a reduction in the low field hole mobility. In addition, the boron used to dope these P^+ poly gate electrodes can diffuse through the thin gate oxides and into the underlying silicon channel region, causing an instability in the PMOS threshold voltage (V_{tp}). Previous studies have shown that the V_{tp} shifts due to boron penetration through thin gate oxides become more severe with increasing post-implant anneal temperature[39] as shown in the fig. 5a below.

As shown in fig.5c the presence of fluorine further enhances the boron penetration problem resulting in PMOS devices with positive threshold voltage shifts [40,41] and an increase in the electron trapping rate in the thin gate oxide[40]. The amount of Vtp shift is enhanced for thinner gate oxdes as shown in fig.5b.

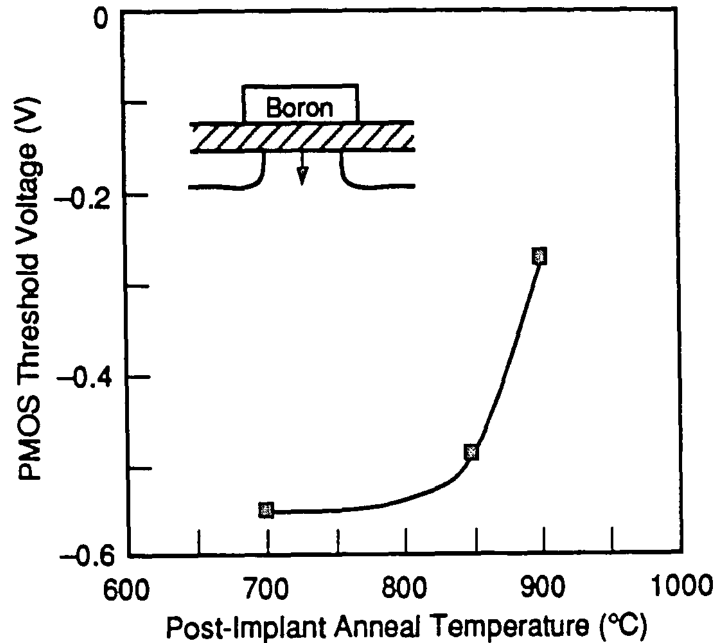


Figure 5a: Boron penetration through the gate oxide. [39]

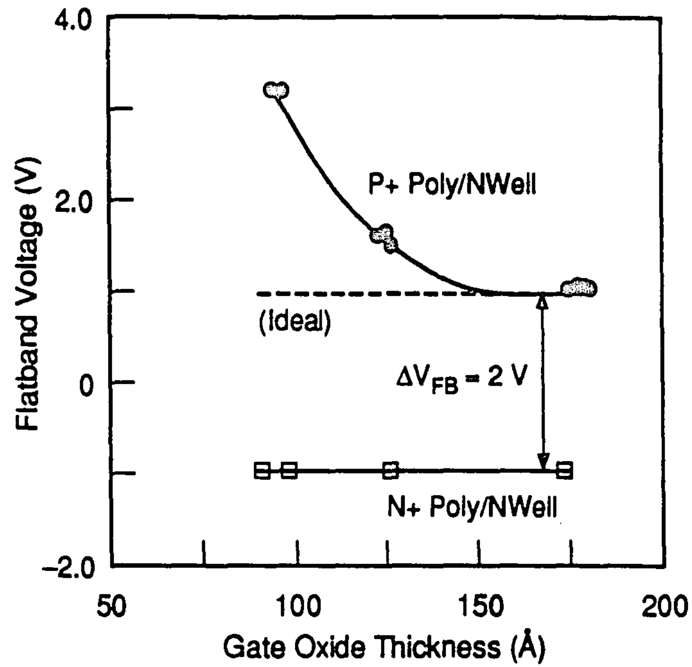


Figure 5b: V_{FB} as a function of gate oxide thickness for n-channel and BF_2-implanted p-channel capacitors. [39]

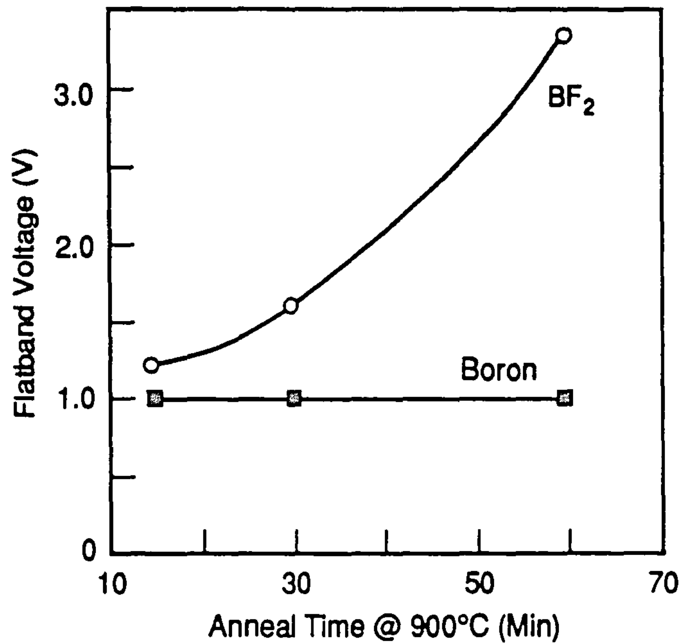


Figure 5c: Flat-band voltages for boron- and BF_2-implanted devices as a function of 900°C anneal time. [39]

CHAPTER 2

THEORY

2.1 THIN DIELECTRIC OXIDATION

Thermally grown silicon dioxide has been an integral part of silicon device technology, and applications have made silicon oxidation a vital processing step in VLSI technologies. Its most important application is the growth of gate-dielectric material for MOS transistors. Continued shrinkage of device dimensions in the sub-micron regime oxide layers of 150Å and below is being implemented in an attempt to meet the stringent device requirements. These thin gate oxides must be grown routinely and reproducibly with high yield and long-term reliability and must not be degraded by subsequent processing.

Earlier studies of silicon oxidation kinetics, revealed that the growth rates were higher than predicted by the linear-parabolic kinetics for layers below 250Å, also the oxide characteristics such as dielectric breakdown, interfacial optical properties and reliability differed[42]. To fully quantify the electrical effects there is an unquestionable need to understand the oxidation kinetics, the role of the oxidizing ambient and how they impact the intrinsic oxide quality and reliability yield strength.

2.1.1 BASIC OXIDATION MODEL

This model commonly known as the Grove Deal model is valid for oxide thicknesses above 300Å in a dry oxygen ambient. For oxidation to occur, the oxidizing species must: 1. Move from the oxygen-containing gas phase across the gas-oxide

interface with a flux F_1, 2. Move across the SiO_2 film already present toward the silicon with flux F_2 and 3. React with the silicon at the $Si - SiO_2$ interface with a flux F_3.

In steady state: $F_1 = F_2 = F_3$, and in equilibrium the concentration of oxidant in the oxide is related to the partial pressure of the oxidant by

$$C^* = kP_g$$

Assuming the flux across the oxide layer follows Fick's law at any point within the oxide layer

$$F_2 = -D\frac{dC}{dX}$$

From steady state assumption F_2 must be the same at all points in the $SiO_2 \Rightarrow \frac{dF_2}{dX} = 0$,

hence

$$F_2 = \frac{D(C_0 - C_i)}{X_0}$$

The flux representing the oxidation reaction occurring at the $Si - SiO_2$ interface is assumed to be proportional to the concentration of oxidant C_i

$$F_3 = k_s C_i$$

where k_s is the surface reaction constant. Equating $F_1 = F_2$, $F_2 = F_3$

$$C_i = \frac{C^*}{1 + \frac{k_s}{h} + \frac{k_s X_0}{D}} \text{ and } C_0 = \frac{(1 + \frac{k_s X_0}{D})C^*}{1 + \frac{k_s}{h} + \frac{k_s X_0}{D}}$$

The flux of the oxidant reaching the $Si - SiO_2$ interface is expressed as

$$N_I \frac{dX_0}{dt} = F_3 = \frac{k_s C^*}{1 + \frac{k_s}{h} + \frac{k_s X_0}{D}}$$

but $X_0 = X_i$ at t=0 hence the solution to the differential equation is

$$X_0^2 + AX_0 = B(t+\tau)$$

where $A = 2D(k_s^{-1} + h^{-1})$, $B = 2DC^* N_I^{-1}$ and $\tau \equiv \frac{X_i^2 + AX_i}{B}$. Thus solving for the quadratic relationship for X_0 as a function of t implies that

$$\frac{X_0}{A/2} = (1 + \frac{t+\tau}{A^2/4B})^{1/2} - 1$$

For long oxidation times : t >> $\frac{A^2}{4B}$ and t >> τ the above equation could be approximated as

$$\boxed{X_0^2 = Bt} \text{ --- Parabolic Law} \quad (1)$$

For short times t+ τ << $\frac{A^2}{4B}$ and again the above equation reduces to

$$\boxed{X_0 = \frac{B}{A}(t+\tau)} \text{ --- Linear Law} \quad (2)$$

The advanced model of oxidation proposed by Massoud et al [18] specifies that an oxidation rate higher than that predicted by the Grove Deal model results in additional processes in the thin regime. In a general case these processes can be represented by fluxes ΔF_1, ΔF_2 and ΔF_3 in parallel with F_1, F_2 and F_3 respectively, and F_4 in parallel with their total sequence as illustrated in fig.2.

As a result of the additional fluxes : F=$F_1 + \Delta F_1 = F_2 + \Delta F_2 = F_3 + \Delta F_3$ substituting the new F_1, F_2 and F_3 into the initial Grove Deal model yields

$$\frac{dX_{ox}}{dt} = \frac{B + (\frac{2D}{N_I h})\Delta F_1 + (\frac{2X_{ox}}{N_I})\Delta F_2 + (\frac{2D}{N_I k_s})\Delta F_3 + (\frac{2X_{ox} + A}{N_I})F_4}{2X_{ox} + A}$$

From the experiments of Massoud et al [18] , plots of oxidation-rate enhancement, expressed in the form of an additive term $\Delta R_e(X_{ox})$, from the dry oxidation of <100> lightly doped silicon oxidized in dry oxygen in the 800°C - 1000°C range as shown below

in fig. 6. It can be seen that the excess rate, at all temperatures, has an initial phase extending to 50Å and an intermediate phase extending from that point to the onset of

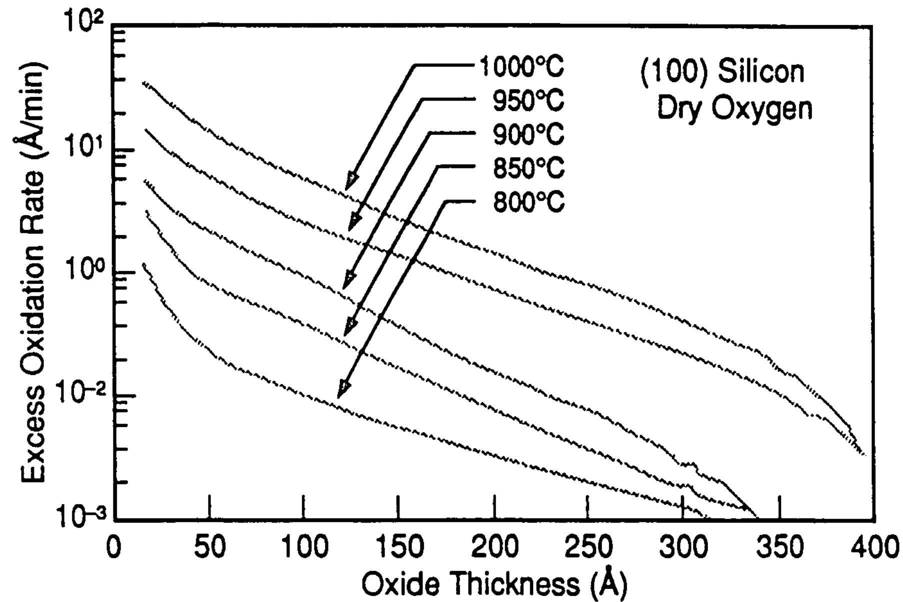


Figure 6: Oxide-thickness dependence of the excess in growth rate for the oxidation of lightly doped (100) silicon in dry oxygen in the 800°–1000°C range. [18]

linear-parabolic kinetics. This suggest that $\Delta R_e(X_{ox})$ can be expressed as two terms; the parallel straight-line dependence of $\Delta R_e(X_{ox})$ on thickness in the intermediate phase suggests that the fit could be accomplished with a term that exponentially decays with thickness. When $\Delta R_e(X_{ox})$ in the initial phase was calculated and plotted it could be fitted to a second term that decays similarly with thickness. The expression of the total growth rate is then expressed as

$$\frac{\partial X_{ox}}{\partial t} = \frac{B}{2X_{ox} + A} + C_1 e^{-(x_{ox}/L_1)} + C_2 e^{-(x_{ox}/L_2)}$$

The following observations were made by Massoud et al :

1. The characteristic length L_2 is approximately independent of temperature in the 800°C-1000°C range and has a value of 69Å for <100>.

2. The constant C_2 has a single activation energy in the Arrhenius-type plot

3. The characteristic length L_1 increases slowly from 7.7Å at 800°C to 12.4Å at 1000°C for <100>.

4. The dependence of C_1 on temperature has a break at ~900°C and is temperature independent below 900°C.

The lack of a well-behaved trend in the temperature dependence of C_1 may be attributed to the fact that it describes the growth rate of the first few layers of SiO_2 that has been shown to depend on the surface after cleaning and pre-oxidation procedures. It is therefore suspected that the first exponential contribution of the excess in oxidation rate is closely related to wafer cleaning and surface preparation. From their experimental data a good empirical expression for the oxidation rate in the thin gate oxide regime was proposed [18]

$$\boxed{\frac{\partial X_{ox}}{\partial t} = \frac{B}{2X_{ox} + A} + C_2 e^{-(x_{ox}/L_2)}} \qquad (3)$$

In summary , the oxidation-rate enhancement during the early stages of silicon oxidation in dry oxygen can be modeled by adding a term that exponentially decays with oxide thickness to the rate expression of the linear-parabolic model. The prefactor of the added term reflects the excess rate at which the reaction proceeds and a characteristic length that represents its extent in thickness. The activation energy of the excess rate was found to be a function of substrate doping density and oxygen partial pressure in the ambient.

2.2 EFFECTS OF FLUORINE

The most commonly encountered silicon oxidation ambients are, Fluorine, Nitrogen and Argon which could be used either as part of the oxidation process or for Post -Oxidation Anneal (POA).

Fluorine enriched gas ambients have been used for some time to increase the oxidation and nitridation rate of silicon. The fluorine could be introduced via fluorinated gas or ion implantation. Morita et al [43-46] examined the effect of adding low levels of NF_3 (< 1%) to dry oxygen ambient. They were able to increase the oxidation rate and observed other indications that the incorporated fluorine tended to neutralize positive charges at the $Si - SiO_2$ interface. The results of the XPS work showed that the fluorine atoms formed only Si-F bonds and no O-F bonds. Both the linear and parabolic rate constants of oxidation were enhanced by the addition of NF_3. The fluorine atoms enhanced the linear rate constant by acting as a catalyst at the interface. It was suggested that the increase in the coefficient was a result of the large number of Si-F bonds in the oxide modifying the microstructure of the oxide and enhancing the oxygen diffusivity.

Schmidt et al [47] performed a thorough study of the effects of fluorine on oxidation rates by adding 1,2-dichlorofluoroethane in various amounts to a dry O_2 ambient for temperatures ranging from 700°C to 800°C, for times 1 to 12 hours, and concentrations of $C_2H_3C_{12}F$ up to 0.11% by volume. Their experiments showed enhanced oxidation in both the linear and parabolic regimes with $C_2H_3C_{12}F$ concentrations as low as 0.011%. Kim et al [48,49] have examined the effect of fluorine-enriched oxidation on stacking faults and boron diffusion. In their work they reported a shrinkage of stacking fault size with the addition of NF_3 to their dry oxidation ambient. This shrinkage has been attributed to an increased vacancy concentration as a result of the fluorine.

2.2.1 INTERACTION WITH INTERFACE STATES

When fluorine bonds to a dangling weak Si bond, the associated trap state is moved outside the silicon band gap, thus rendering it ineffective. A model of continuously distributed trap states has been given by Sakurai et al [50]. They developed a method to calculate the electronic structures of crystalline and S^{-}-amorphous SiO_2 interface with and without microstructural defects. They modeled the interface as crystalline Si with <111> orientation and amorphous SiO_2 represented by a Bethe-lattice as shown in fig. 7a. The calculations of Sakurai et al suggest that both the perfect interface with a bonding angle of 120° to 180° and the interface with oxygen bonds do not have a state within the band gap. However the $Si_3 = Si$ dangling bond at the interface gives rise to a state at about the middle of the Si band gap as shown in fig. 7b. The O^{-} vacancy and Si-Si weak bond at the interface produce traps in the band gap whose energies vary in the lower half of the band gap by changing various bonding parameters. Fig. 7b demonstrates the dependence of the Si-Si weak bond level including the O^{-} vacancy level on the Si-Si bond length.

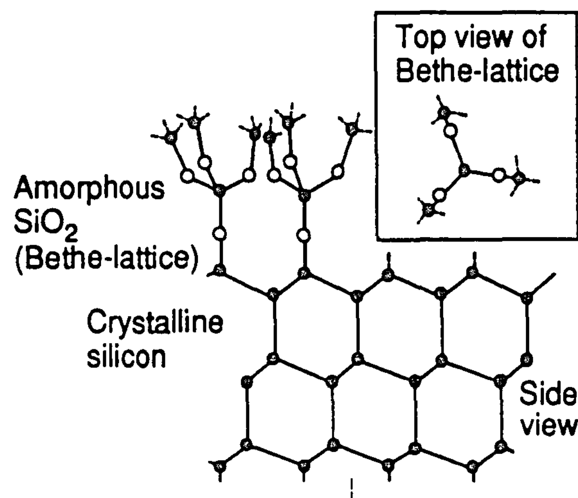


Figure 7a: Basic model constructed with amorphous SiO_2 represented by Bethe-lattice and Si substrate with (111) orientation. Open and closed circles denote O and Si atoms respectively. [50]

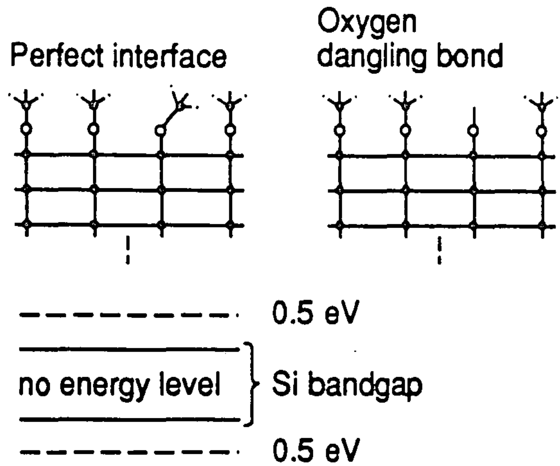


Figure 7b: Perfect interface and oxygen dangling bond at the Si-SiO_2 interface having no energy level in the range between 0.5 eV below the top of the valence band and 0.5 eV above the bottom of the conduction band of Si. Open and closed circles denote O and Si atoms respectively. [50]

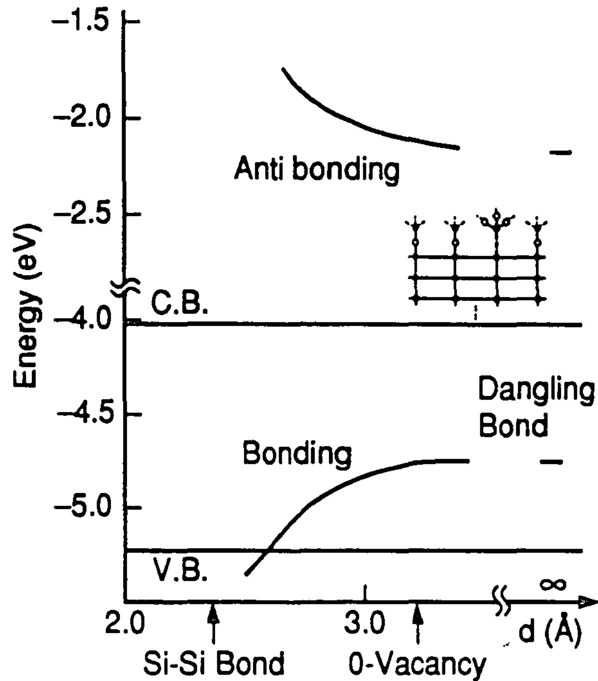


Figure 7c: Si dangling bond, Si-Si bond, and O vacancy level at the interface. These energy levels move in the lower half of the Si band gap with changing bond length d. Open and closed circles denote O and Si atoms respectively. [50]

The commonly observed U shaped distribution as shown in fig. 8a can be explained with a bonding length distribution. The rapid decrease of the distribution in fig 8a can be attributed to the normal bond lengths being shorter than 2.5Å and the longer bond lengths less likely. The gap states move out of the energy range between 0.5eV

below the top of the Si valence band and 0.5eV above the bottom of the Si conduction band when species H,OH,Cl of F is bonded to the Si atom at the interface as illustrated in fig. 8b. As a result a trap whose energy level is outside the band gap can not normally function as an interface state. Hence this explains the effectiveness of reducing the interface states by annealing semiconductor devices in H_2 HCl or F. Because of the stronger bonding energy of the Si-F(5.73eV) compared to Si-H bond(3.18eV) [51], fluorine bonds at the $Si - SiO_2$ interface should be more immune to hot electrons and show improved resistance to irradiation.

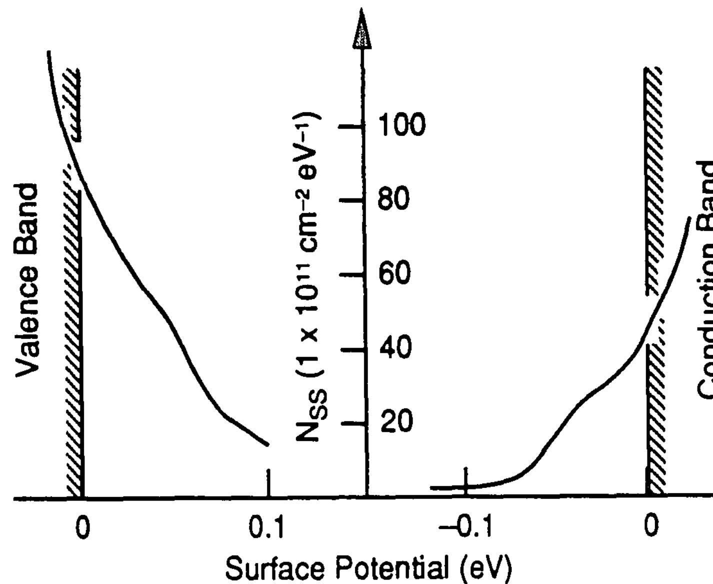


Figure 8a: Commonly observed U-shaped distribution of interface-trap state density in the forbidden gap of Si. [50]

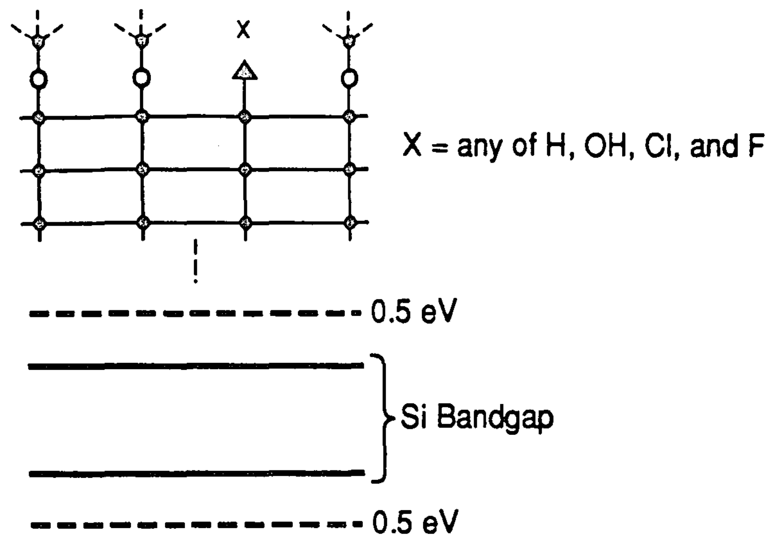


Figure 8b: Impurity at the Si-SiO_2 interface. If any of H, OH, Cl, and F is bonded to the Si dangling bond, no energy level exists in the energy range between 0.5 eV below the top of the valence band and the 0.5 eV above the bottom of the conduction band of Si. Open and closed circles denote O and Si atoms respectively. [50]

2.2.2 FLUORINE EFFECT ON FLATBAND VOLTAGE

Results of high frequency capacitance vs. voltage measurements are shown in fig. 9. As the fluorine dose increases the flatband voltage shift is more severe for the thinner oxide, due to the negative fixed charge at the $Si - SiO_2$ interface and a relaxation of the bonds.

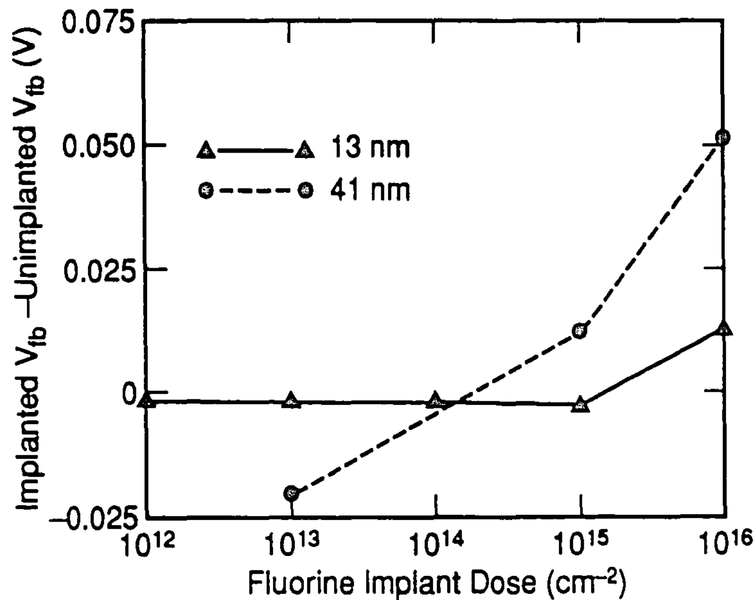


Figure 9: Change in flat-band voltage of fluorine enriched SiO_2 vs. the fluorine implant dose. [54]

2.2.3 Q_{BD}, V_{BD} AND AGING EFFECTS

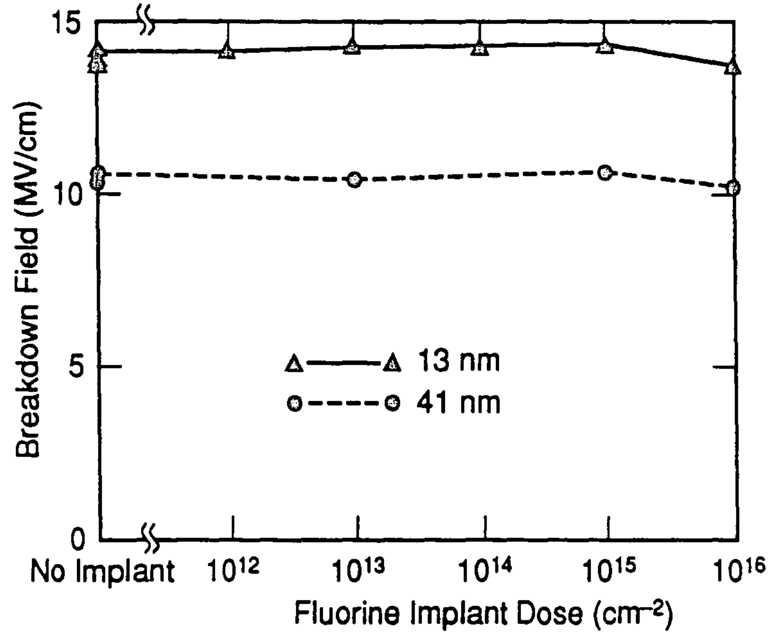


Figure 10a: Breakdown field of fluorine rich oxides as a function of implant dose. [54]

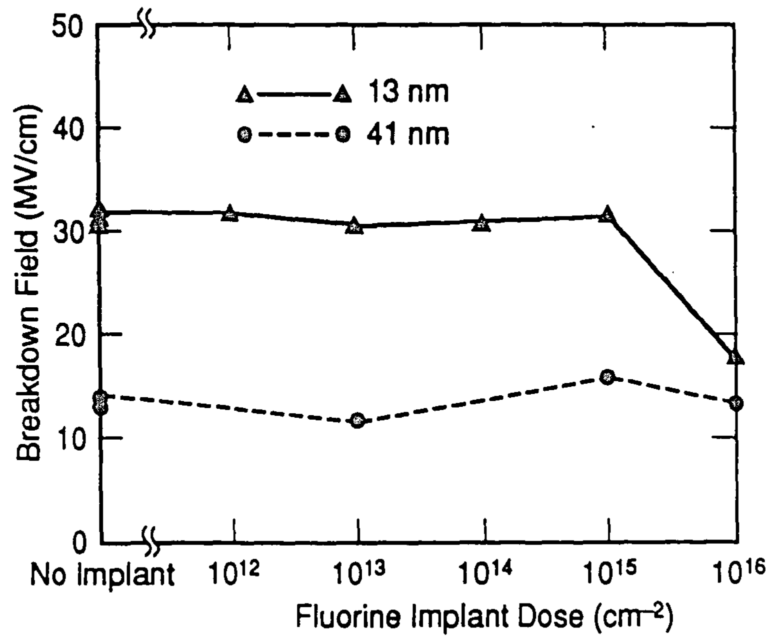


Figure 10b: Charge-to-breakdown of fluorine rich oxides as a function of implant dose. [54]

Fluorine has a small degradation in breakdown field at higher dose as shown in fig. 10a. The measurements of Q_{BD} are shown in fig.10b. Low to medium doses of fluorine are found to have little effect on the Q_{BD} of the MOS capacitors. However it is seen that higher doses of fluorine has a detrimental effect on Q_{BD} of the thinner oxides.

The mobility and concentration of ionic fluorine in oxides which have undergone high temperature processing has been examined to determine if the fluorine acts as a mobile ion in the oxide by Stagg et al[52]. They reported that the drift mobility of ions is SiO_2 dependent, to first order, on the ionic radius. The ionic radius of F^- and K^+ are both about 1.33Å[53]. Hence to a first order the two should have similar mobilities. Stagg et al showed that for a 10^{16} F^- implanted sample about 10^{10} mobile ions were detected. Hence fluorine which has been annealed in an oxide at high temperature and forms a stable bond does not act as a negative mobile ion.

Measurements of hot electron degradation as a function of fluorine dose is shown in fig. 11. This was reported by Wright [54]. The lifetime was defined as the time to 10% degradation in g_m. During the stressing the gate voltage was set to the peak value of substrate current for a given drain bias. The drawn gate lengths of the transistors were $1.25\mu m$. For high drain voltages, the fluorine enriched oxides showed enhanced degradation, but at 5V the lifetimes of the fluorinated oxides were about 1000 times greater than that of the pure unimplanted oxide.

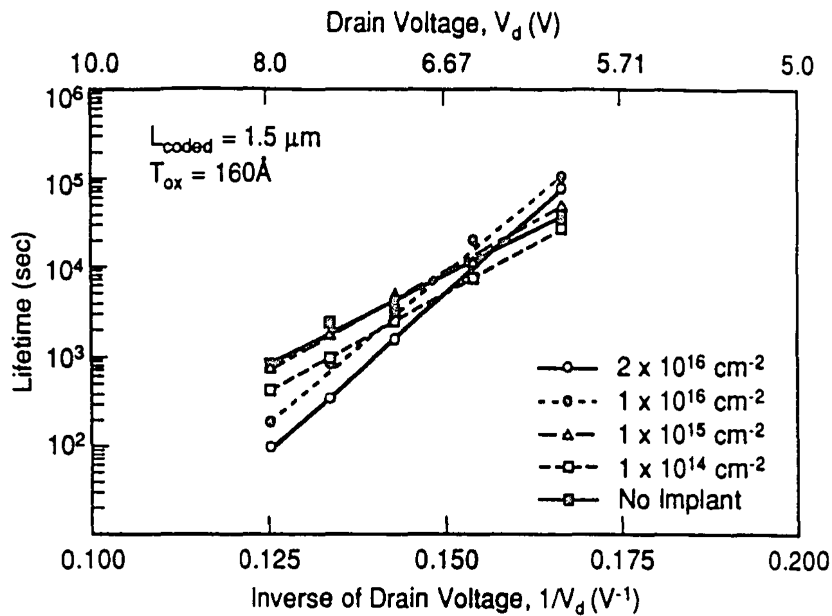


Figure 11: Transistor lifetime until 10% degradation in g_m vs. reciprocal of drain voltage. [54]

2.3 EFFECTS OF NITRIDED OXIDES ON DEVICES

Long-term reliability, electrical stability, and the poor diffusion barrier properties of sub 100Å thick oxides has created interest in alternative dielectrics to silicon dioxide. Of these nitroxides and reoxidized nitroxides appear to be the most promising for future gate dielectric applications. Nitroxides have shown good diffusion barriers against impurities [55] and superior hot electron immunity in comparison to thermal oxide[56]. But nitroxides have been reported to have lower charge to breakdown(Q_{BD})[57], increased trapping[58], and a higher fixed charge than pure oxide[59]. Although extensive nitridation of the film returns the fixed charge density to the original levels of the pure oxide, the mobility of FETs with these gate dielectrics never returns to the value of the starting oxide[60].

2.3.1 KINETICS OF NITRIDATION

The kinetics of nitridation have been extensively studied although many questions still remain. Parameters used to model the nitridation process are similar to those used by Hori et al[59]. For nitridation in an ammonia ambient, a nitriding species diffuses through

the gate dielectric from the oxide surface. The oxide then reacts with this species forming a nitride with a total amount controlled by the temperature and with the nitridation occurring preferentially at the $Si-SiO_2$ interface. Theories put forth to explain the increased nitrogen concentration at the $Si-SiO_2$ interface typically suggest that stress at the interface enhances the reaction rate. The increased surface nitridation rate relative to that of the bulk is typically thought of being a result of a higher concentration of the nitriding species. As a result of the nitrogen incorporation, fixed charge and interface states increase.

Hori et al[59] have developed a set of coupled partial differential equations to explain the nitridation process. The diffusion of the nitriding species is modeled by differential equation of diffusion as

$$\frac{\partial N_d(x,t)}{\partial t} = D_n \frac{\partial^2 N_d(x,t)}{\partial x^2}$$

where $N_d(x,t)$ is the volume density of nitrogen-related diffusing species and D_n is the diffusion coefficient. D_n decreases at the surface of the oxide for longer nitridation times but has been modeled as a constant. The differential equation for the nitrogen incorporation is given as

$$\boxed{\frac{\partial I(x,t)}{\partial t} = k_i N_d(x,t)\{I_{sat} - I(x,t)\}} \quad (4)$$

where $I(x,t)$ is the volume density of incorporated nitrogen, I_{sat} is the saturation value of incorporated nitrogen, and k_i is the rate constant of nitrogen incorporation. I_{sat} is a function of the nitridation temperature and increases with increasing nitridation temperature. The results of Pan and Paquette[61] indicate that the steady state value of nitrogen incorporation is determined by the temperature of the film, not the pressure of nitridation, and the surface boundary conditions must be known to solve the differential equation.

Defects generated in the oxide from the nitridation have been suggested as being responsible for the increase in fixed charge density and D_{it}. The density of the defect state is expressed by the differential equation

$$\boxed{\frac{\partial N_d(x,t)}{\partial t} = P\frac{\partial I(x,t)}{\partial t} - k_a N(x,t)} \tag{5}$$

where $N(x,t)$ is the volume density of the defect state, P is the probability of forming a defect as a result of nitrogen incorporation, and k_a is the rate constant of the annealing-type process.

2.3.2 NITRIDATION EFFECTS

Wright et al[62] examined the effect of nitridation conditions on dielectric properties. As shown in fig. 12, nitridation leads to a decrease in Q_{BD} compared with pure oxides. The peak is similar to the shift in flat-band voltage with nitridation[58] and is a result of positive charge in the dielectric reducing the Fowler-Nordheim electron tunneling distance. The peak Q_{BD} is higher for high temperature nitridation in comparison to that of low temperature. Hori et al[63] have demonstrated that the interface integrity of nitroxides increases for increasing interfacial nitrogen concentration. The ratio of hydrogen to nitrogen incorporated in the film has been found to decrease with increasing temperature[64].

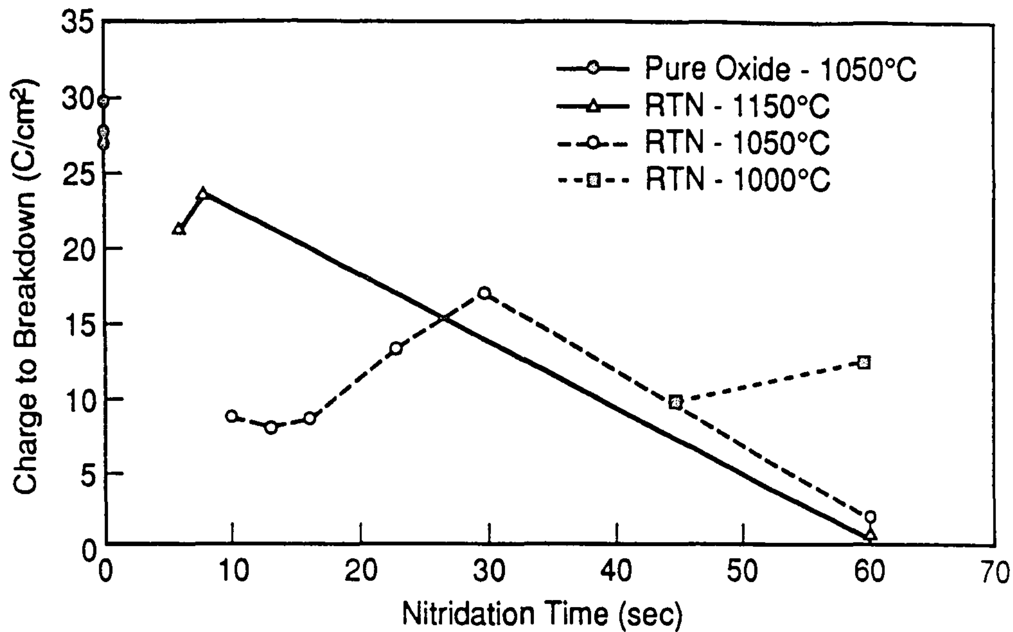


Figure 12: Charge to breakdown of nitroxide dielectrics. [54]

2.4 EFFECTS OF POST-OXIDATION ANNEAL (POA)

The effects of inserting an inert anneal in between the oxidation and nitridation on the dielectric properties have been examined. Anneals such as argon have been reported to improve the dielectric properties[65]. Fig. 13a shows the effect of these anneals on the breakdown field of MOS capacitors. Wright reported that there is no significant difference between argon and nitrogen for post-nitridation annealing. They also reported as shown in fig. 13b that the surface mobility of MOS devices decreases with nitidation time and begins to increase with reoxidation. Although continued reoxidation will further increase the mobility, additional oxide grown will decrease the g_m of the device. The resistance to hot electron stressing is shown in fig. 13c. Devices show improvement in lifetime with nitridation and further improvement with reoxidation anneals.

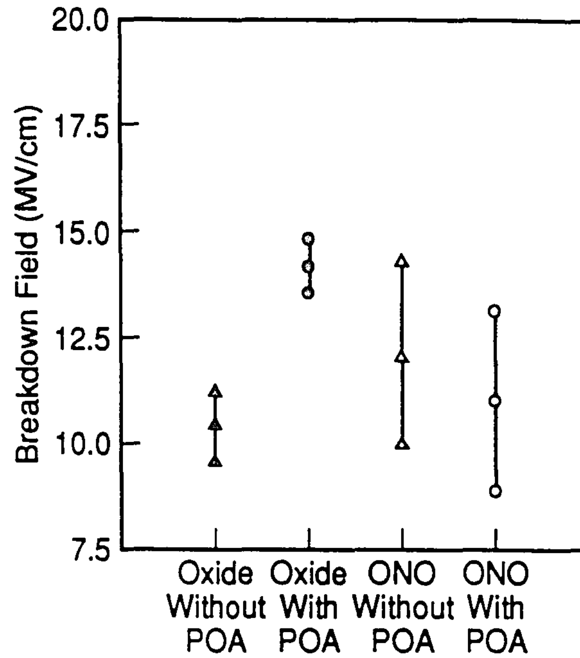


Figure 13a: Breakdown field for MOS capacitors with and without post-oxidation anneals. [54]

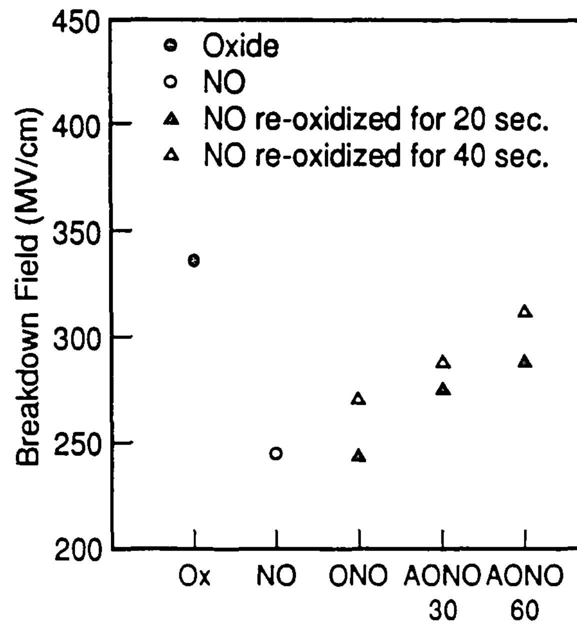


Figure 13b: The effect of gate dielectric formation conditions on surface mobility. **OX**: pure oxide grown at 1050°C for 35 seconds. **NO**: nitroxide formed at 1150°C for 24 seconds in a 25% NH_3 ambient. **ONO**: NO reoxidized at 1150°C for 20 or 40 seconds. **AONO 30**: ONO annealed an additional 30 seconds in argon at 1150°C. **AONO 60**: ONO annealed an additional 60 seconds in argon at 1150°C. [54]

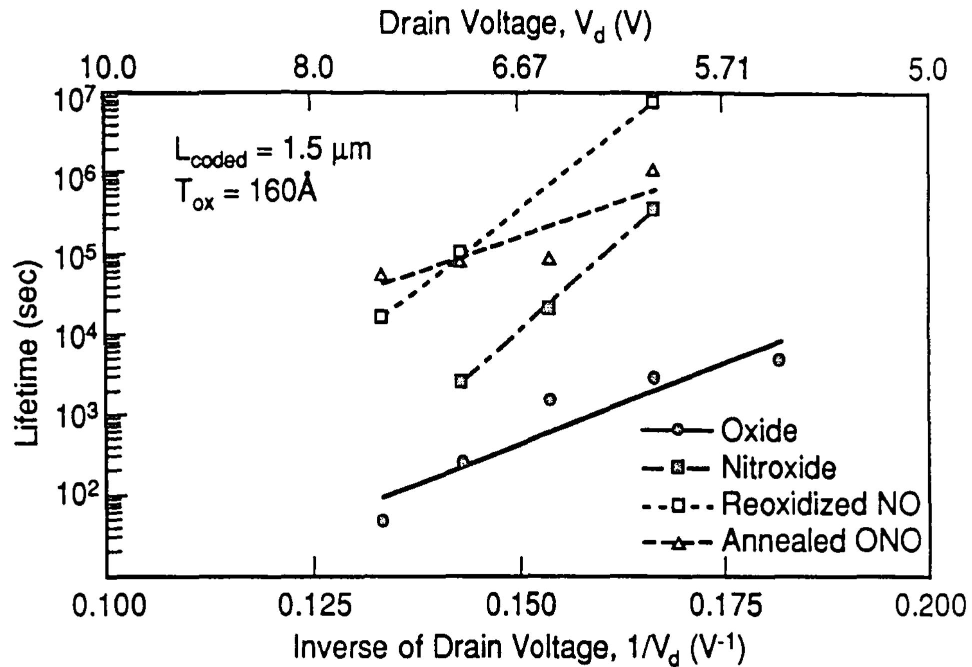


Figure 13c: Hot electron lifetime of annealed NO MISFETs. [54]

2.5 CHARACTERIZATION OF THIN OXIDES

To accurately predict the dielectric viability of thin oxides in ULSI devices, there is an ardent need to fully identify the phenomena of charge distribution and interactions at the $Si - SiO_2$ interface. The quality of the oxide is a function of oxidation and annealing conditions, oxide charges, surface crystallographic orientation, pre-oxidation surface preparation and a number of factors.

There are four basic types of charges that exist in the oxide or near the $Si - SiO_2$ interface namely: interface trap charge (Q_{it}), fixed oxide charge, (Q_f), mobile ionic charge (Q_m) and bulk oxide trapped charge (Q_{ot}).

2.5.1 INTERFACE TRAP CHARGE (Q_{it})

The interface trap charge refers to charge which is localized at the $Si - SiO_2$ interface on sites that can change their charge state by exchange of mobile carriers with the silicon. The charge state of the interface trap site changes with gate bias if the interface trap is moved past the Fermi level, causing its occupancy to change. These traps have energy levels distributed throughout the silicon bandgap with a U-shaped distribution across the bandgap. The density of these charges can be quantified by the extraction of D_{it}. The charge pumping technique as described in [66] can be used to accurately quantify the D_{it}. This can be achieved by pulsing the gate of a transistor with a triangular waveform, and plotting the recombined charge per cycle as a function of frequency on a semi-logarithmic plot. One obtains a straight line, the slope of which can be used to determine the interface state density by employing the equations below.

$$\boxed{Q_{ss} = \frac{Icp}{f} = 2qD_{it}A_G kt\{\ln(v_{th}n_i\sqrt{\sigma_p\sigma_n}) + \ln[\frac{|V_{fb} - V_t|}{|\Delta V_G|}\frac{1}{f}\sqrt{\alpha(1-\alpha)}]\}} \quad (6)$$

The slope of Q_{ss} with respect to log f is given by

$$\frac{dQ_{ss}}{d\log f} = \frac{2qktD_{it}}{\log e}A_G.$$

In as grown oxides Q_{it} depends on the oxidation temperature, oxygen partial pressure, and silicon substrate orientation, and is found to decrease with increasing temperature. The two most effective annealing techniques employed in the reduction of Q_{it} are low temperature post-metallization anneal and high temperature post-oxidation anneal. It is believed that the hydrogen diffuses to the $Si - SiO_2$ interface where it reacts chemically with traps thus rendering them electrically inactive.

2.5.2 FIXED OXIDE CHARGE (Q_f)

The fixed oxide charge is located within 35Å of the $Si - SiO_2$ interface in the so-called transition region between silicon and SiO_2. These charges do not change their state by exchange of mobile carriers with silicon as occurs in Q_{it}. The Q_f charge is predominantly positive and typically viewed as a sheet of charge. The lowest values of Q_f are obtained at high oxidation temperatures. A rapid rate of cooling is also known to reduce Q_f, but this quenching activity is impractical and may lead to warpage. For a given oxide thickness the total fixed charge Q_f is typically defined as

$$\boxed{Q_f = \int_0^{T_{ox}} \frac{x}{T_{ox}} \rho(x) dx} \tag{7}$$

where T_{ox} is the effective oxide thickness, and $\rho(x)$ is the spatial distribution function of the fixed charge in the oxide.

2.5.3 MOBILE IONIC CHARGE (Q_m)

The mobile ionic charge is commonly caused by the presence of ionized alkali metal atoms (Na^+, K^+). This charge is located either at the gate-SiO_2 interface or at the $Si - SiO_2$ interface. It is found to drift with applied gate bias even at room temperature, since the ions are extremely mobile in SiO_2. The degree of incorporation of the mobile ions in the gate oxide is dependent on the cleanliness of the fabrication process. The presence of these ions can in long term alter the threshold voltage of the MOS device. The mobile ionic charge density which ranges from 10^{10} to $10^{12} / cm^2$ can be measured either by C-V temperature biasing method or using the triangular voltage sweep (TVS) [67-69]

Methods of eliminating mobile ions includes growing the SiO_2 in HCl or TCA which neutralizes the alkali metals upon arrival at the interface. Also phosphosilicate glass used in device fabrication can act as a getter for mobile alkali metals.

2.5.4 OXIDE TRAPPED CHARGE (Q_{ot})

Bulk oxide traps can be located at the gate-SiO_2 interface, the $Si - SiO_2$ interface, as well as deep in the oxide. The traps are associated with defects in the SiO_2, such as impurities and broken bonds. Such traps are usually uncharged, but can become charged when electrons and holes are introduced into the oxide via mechanisms such as avalanche injection and ionizing radiation.

2.5.5 FOWLER-NORDHEIM TUNNELING MECHANISM

Thin gate oxide wearout is one of the major reliability concerns for MOS integrated circuits. The mechanism of oxide time-dependent breakdown has been attributed to charge trapping [70]. It has been demonstrated that there exist a close correlation between oxide breakdown and hole trapping by energetic electrons injected into oxide via Fowler-Nordheim (F-N) tunneling[70].

The general consensus in the literature is that the reliability of thin oxides improves with decreasing oxide thickness [70]. However all data presented so far has been obtained using the same techniques used for thick insulators which can lead to serious errors and overestimations. The coventional breakdown test assumes that oxide breakdown causes a large and sudden increase in electrical conductivity and V_{BD} is determined from a large jump in the measured current-voltage ramp.

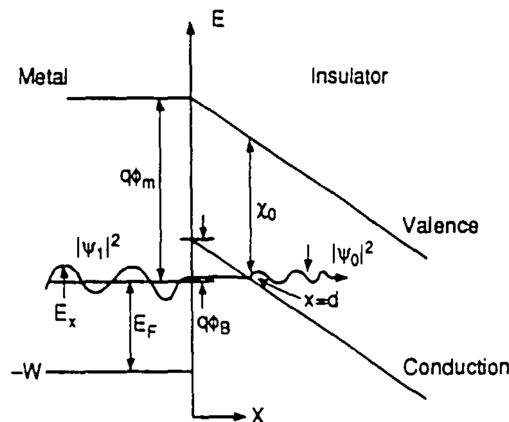


Figure 14: Fowler-Nordheim Tunneling Diagram

From quantum mechanics, the transmission probability is given by

$$T(x)=\left|\frac{\psi_0}{\psi}\right|^2=e^{-\int_0^x k(x)dx} \quad \text{and} \quad \text{E-V}=\varepsilon_X,\ \ \Phi_B=\Phi_M-\psi_0$$

where $k(x)$ is the vector associated with a particle in the barrier, V is the potential energy, E is the total energy. Assuming a single parabolic energy band model at X=0, V- ε_X,= ε_F, thus at any X in the barrier,

$$V-\varepsilon_X=W-q\varepsilon_{ox}-q\chi_0-\varepsilon_X$$

$$\text{as V-}\varepsilon_X\rightarrow 0\text{ , X}\rightarrow\text{d}\Rightarrow\ \ d=\frac{W-q\chi_0-\varepsilon_x}{q\varepsilon_0}$$

$$\Rightarrow 2\int_0^d k(x)dx=2\sqrt{\frac{2m^*}{\hbar^2}}\int_0^d (V-\varepsilon_x)^{\frac{1}{2}}dx=2\sqrt{\frac{2m^*}{\hbar^2}}\int_0^d\left[W-q\varepsilon_0-q\chi_0-\varepsilon_x\right]dx$$

It is known that

$$\boxed{J_x=\sum_{all-k}qV_xT(x)f} \tag{8}$$

where f is the Fermi-dirac electron distribution function defined as $f=\frac{1}{1+e^{(\varepsilon_x-\varepsilon_F)/kT}}$

with $\varepsilon_F=\varepsilon_x=\frac{\mathrm{h}^2k_x^{\ 2}}{2m^*}=\frac{\mathrm{h}^2k_F^{\ 2}}{2m^*}$. From the cold electron approximation when $k\le k_F\Rightarrow f(\varepsilon_X)=1$ and when $k>k_F\Rightarrow f(\varepsilon_X)=0$. Integrating the J_x over the first Brillouin zone yields

$$J_x=\frac{2}{(2\pi)^3}\oint d^3kq(\frac{\hbar k_x}{m^*})T$$

substituting $d^3k=dk_x\pi(k_F^{\ 2}-k_x^{\ 2})$, ϕ_B,ε_x into above equation implies that

$$J_x=\frac{4\pi qm^*}{h^3}\int_0^x d\eta\eta e^{-\frac{8\pi\sqrt{2m^*}(q\phi_B)^{\frac{3}{2}}}{3q\varepsilon_0 h}[1+\frac{\eta}{q\phi_B}]^{\frac{3}{2}}},$$

hence the Fowler-Nordheim tunneling current can be approximated by

$$J_{FN} = \frac{q^3 \varepsilon_0^{\ 2}}{16\pi^2 \hbar \phi_B} e^{-\frac{8\pi\sqrt{2m^*} q\phi_B^{\frac{3}{2}}}{3\hbar\varepsilon_0}} = A\varepsilon_0^{\ 2} e^{-\frac{\varepsilon_T}{\varepsilon_0}} \quad (9)$$

where ϕ_B is the barrier height, ε_T is the critical electric field for tunneling, ε_0 is the electric field across the oxide, $A = \frac{q^3}{16\pi^2 \mathrm{h} \phi_B}$ and $\varepsilon_T = \frac{8\pi\sqrt{2m^*} q\phi_B^{\frac{3}{2}}}{3\mathrm{h}}$.

Figure 15a shows the I-V characteristics before (solid) after stress (dot-dashed) for various oxide thickness, ranging from 51Å to 97Å stressed with voltage ramp of 5mV/s increasing until $|J_{FN}|$ tunneling current reaches $0.1 A/cm^2$, corresponding to an average oxide field of ~ 12MV/cm. It is seen that the low-field leakage phenomenon becomes

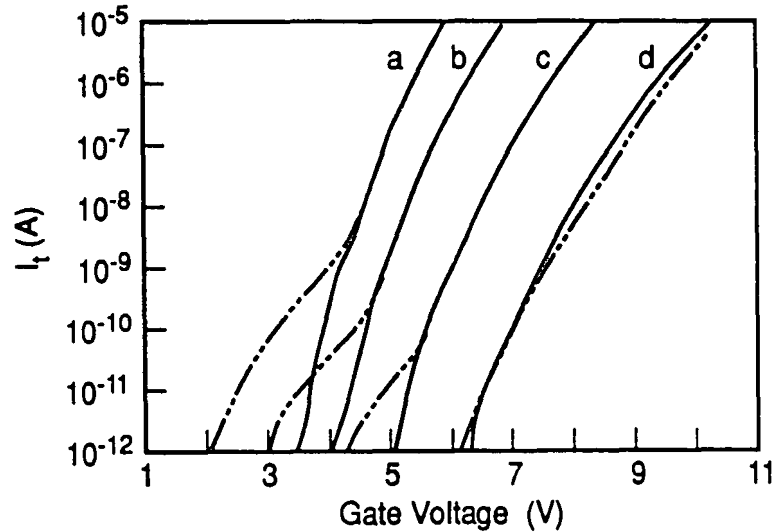


Figure 15a: Dependence of the leakage current on the oxide thickness. The oxide thickness was: 5.1 nm (curve a), 6.0 nm (curve b), 7.5 nm (curve c), and 9.7 nm (curve d). [70]

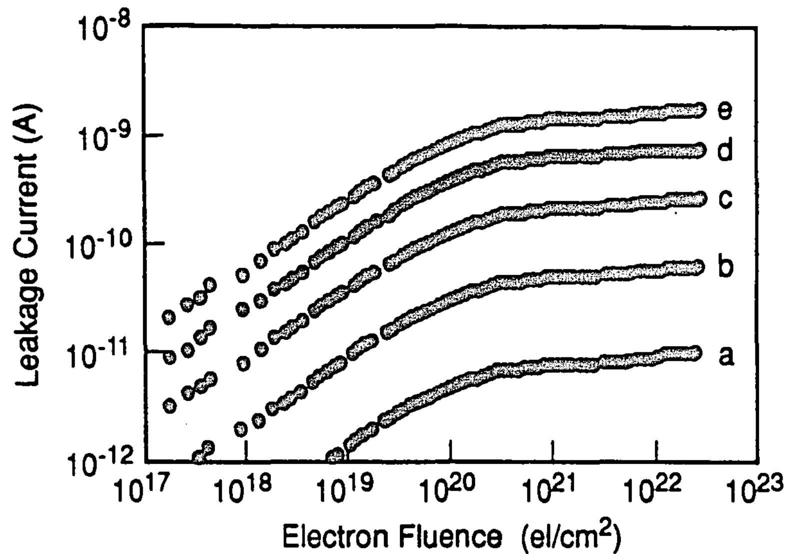


Figure 15b: Increase of the low-field leakage current as a function of electron fluence for a capacitor. The stress was 2.5 V (curve a), 3 V (curve b), 3.5 V (curve c), 4 V (curve d), and 4.5 V (curve e). [70]

increasingly significant in very thin oxides, while the predominant effect of electrical stress in thicker films is electron trapping as indicated by the positive voltage shift.

When the constant voltage stress is applied to thin oxides as done in TDDB, it is found that the stress induced leakage begins to occur at low levels of electron fluence and long before catastrophic failure as shown in fig.15b.

2.6 TIME DEPENDENT DIELECTRIC BREAKDOWN

Most conventional TDDB cannot detect the occurrence of stress-induced oxide leakage because measurements and breakdown determination concentrate in the high current and high-field region. Only repeated I-V measurements at low voltages will reveal the oxide leakage. Devices such as DRAMS, E^2PROMS and SRAMS, while sensitive to leakage currents at low fields will fail when the oxide becomes sufficiently leaky. Because oxide leakage always precedes destructive breakdown, it becomes the dominant failure mode of thin oxides. It is more relevant from device application viewpoint, to determine

the onset of a critical leakage rather than catastrophic breakdown. It is believed that thin oxides have localized weak spots that probably come from imperfections such as particles and surface roughness. High-fields stress destroys the integrity of thin oxides at these weak spots by changing the oxide physically. This change leads to a reduced tunneling barrier and causes a local enhancement of the tunneling current that appears as the stress-induced oxide leakage at low fields.

The reliability of thin oxides is one of the most important problems in MOS integrated circuits. Furthermore, since the gate oxide yield and the time-dependent reliability of an integrated circuit is defect related, statistical modeling is paramount for the prediction of yield and lifetime of devices. As reported by Jack C. Lee et al [70] the concept of modeling oxide defects as "effective oxide thinning" ΔXox and describing the defect density as $D(\Delta X_{OX})$ can be employed in the evaluation of thin oxides for the determination of stress time, stress voltage and stress temperature required for screening to meet a given failure rate.

2.6.1 INTRINSIC BREAKDOWN MODEL

There are numerous models regarding the breakdown mechanism, including the so-called hole-induced breakdown model [71,72], the electron trapping breakdown model [73], and the interface trap generation and resonant-tunneling induced breakdown models [74]. Experiments have shown that the lifetime of an oxide is determined by the time required for the hole fluence Q_p to reach some critical value

$$Q_p \,\alpha\, J_{FN}\mathrm{t} \Rightarrow Q_p \,\alpha\, e^{-G/E_{ox}}$$

where $J_{FN} \,\alpha\, e^{-B/E_{ox}}$ is a more compact form of the Fowler -Nordheim current density previously developed. The time required to accumulate Q_p (t_{BD}) has E- field dependence of $t_{BD} \,\alpha\, e^{G/E_{ox}}$ from graph of fig. 16 which illustrates a plot of t_{BD} vs. $1/E_{ox}$. Furthermore

the electric field acceleration factor defined as the tangential slope of the log (t_{BD}) vs. E_{ox} curve, can be expressed as

$$\beta \equiv \frac{-d(\log(t_{BD}))}{dE_{ox}} = \frac{G}{2.3E^2_{ox}}$$

which implies that B cannot be a constant but proportional to $1/E_{ox}$.

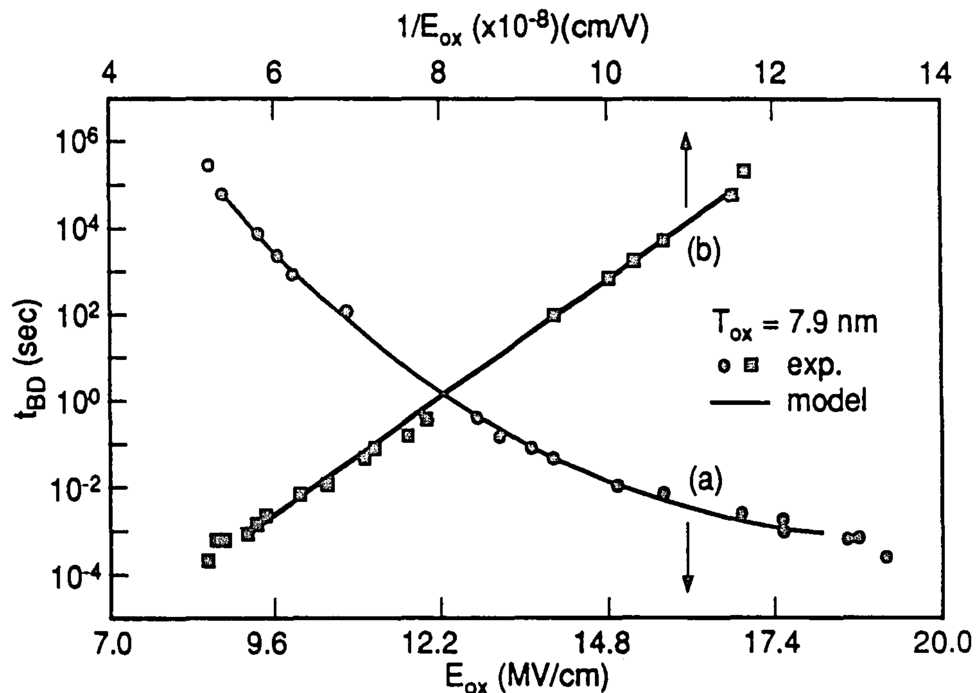


Figure 16: log (t_{BD}) versus $1/E_{ox}$ for a 79Å gate oxide. [70]

2.6.2 DEFECT RELATED BREAKDOWN MODEL

Discussion so far assumes small area oxide samples in which intrinsic breakdown dominates. However in VLSI circuits oxide yield and reliability are determined by defect-related breakdown. Defects that result in localized weak spots in the oxide could lead to effects such as localized high field, higher trap generation rate, or higher current density at these localized areas.

Using the effective oxide thinning by an amount ΔXox as shown in fig.17a and modifying the intrinsic breakdown model[76]

$$t_{BD} = \tau_0 e^{G(X_{ox} - \Delta X_{ox})/V_{ox}} = \tau_0 e^{GX_{eff}/V_{ox}} \quad (10)$$

where $X_{eff} = X_{ox} - \Delta X_{ox}$ G ~ 350MV, Xox is the oxide thickness, ΔX_{ox} is the amount of oxide thinning at the localized defective spot, τ_0 is determined by the intrinsic breakdown time under an applied voltage V_{ox}. Fig. 17b shows a plot of the dependence of the time to breakdown of a given oxide sample under an oxide field E_{ox} and the severity of defect $\Delta X_{ox}/X_{ox}$. τ_0 is 1x 10^{-11} s and is fairly constant for oxide thickness such that X_{ox}-X_t>> t where X_t is tunneling distance and λ ~ 15Å for electron in SiO_2.

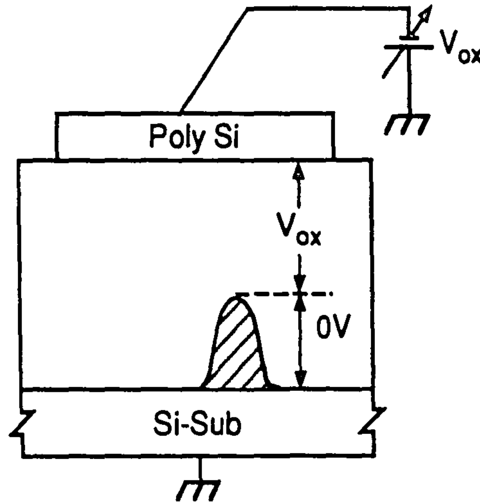


Figure 17a: Oxide defects modeling as effective oxide thinning by ΔX_{ox}. [70]

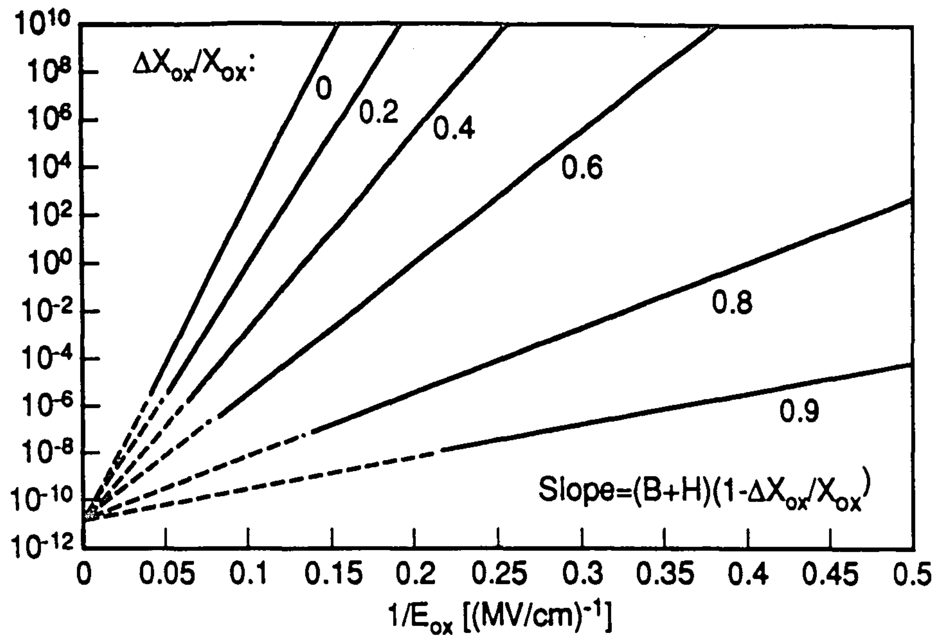


Figure 17b: An illustration of the dependence t_{BD} on the severity of defects ΔX_{ox} as a function of the oxide field E_{ox}. [70]

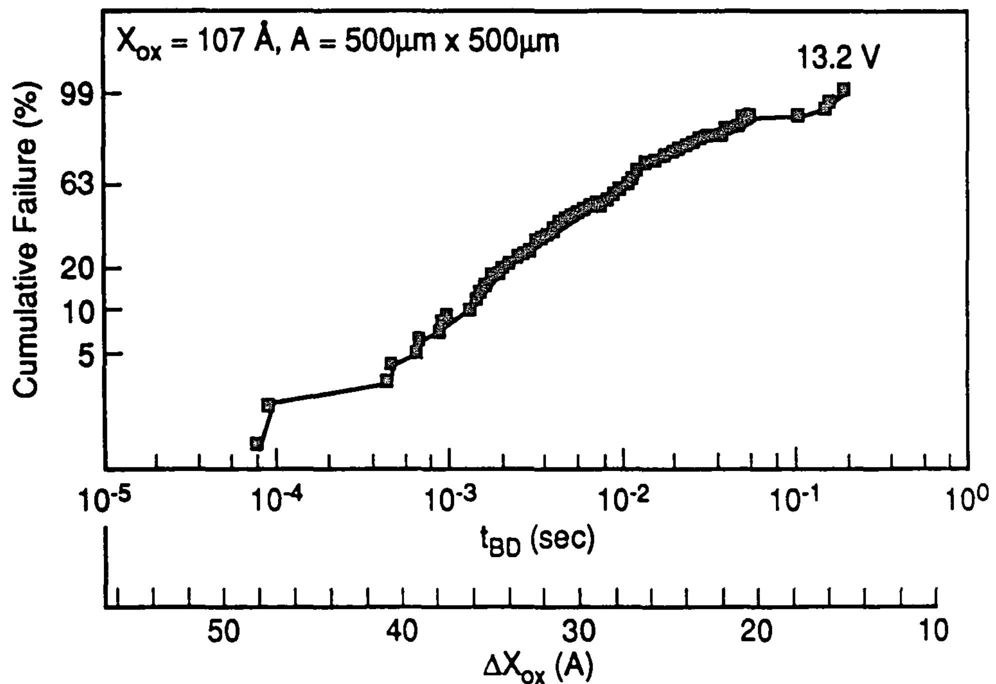


Figure 18a: Time-dependent dielectric breakdown (TDDB) data for an oxide of 107Å under 13.2 V plotted in cumulative percent failure versus t_{BD} or equivalently versus ΔX_{ox}. [70]

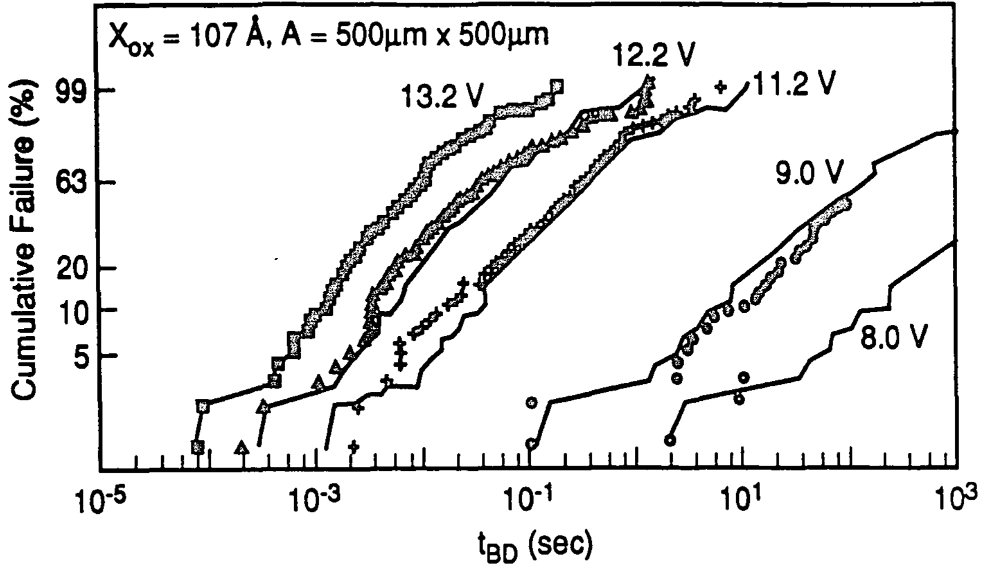


Figure 18b: TDDB data for an oxide of 107Å under various stress voltages. Good agreement is obtained with the ΔX_{ox} model (solid curves) through the equation $t_{BD2} = t_{BD1}(\tau 0/t_{BD1})^{(1-V_{ox1}/V_{ox2})}$. [70]

Fig. 18a shows the time-dependent dielectric breakdown (TDDB) data for $0.25mm^2$ MOS capacitor of 107Å under a stress bias of 13.2V. Since t_{BD} can be related directly to DXox through (5) one can construe cumulative failure versus $D(\Delta X_{OX})$. Fig. 18b shows TDDB data for the same sample set under various oxide fields. The predicted theoritical curves we generated by [77]

$$\boxed{t_{BD2} = t_{BD1}(\tau_0 / t_{BD1})^{(1-V_{ox1}/V_{ox2})}} \tag{11}$$

And as seen in fig. 18b the $D(\Delta X_{OX})$ model prediction agrees quite well with the experimental results. Using the DXox model it can be shown from (5) that the field acceleration factor is a function of the oxide defect[77], hence

$$\beta = \frac{G}{2.3E^2_{ox}}(1 - \frac{\Delta X_{ox}}{X_{ox}}) = \frac{G}{2.3E^2_{ox}}\frac{X_{eff}}{X_{ox}}$$

2.6.3 TDDB(t_{BD}) PREDICTION

To project the lifetime of the oxide, the cumulative defect density as a function of effective oxide thinning is first deduced from TDDB data as shown in fig. 18b. For a given oxide area, the percentage failure below a certain t_{BD} , $P(t'_{BD} < t_{BD})$ is equal to the probability of finding a defect with an effective oxide thinning larger than $D(\Delta X_{OX})$, $P(\Delta X'_{ox} > \Delta X_{ox})$ which implies that

$$P(t'_{BD} < t_{BD}) = P(\Delta X'_{ox} > \Delta X_{ox})$$

$$P(\Delta X'_{ox} > \Delta X_{ox}) = 1 - \text{P(no defect with effective thinning} > D(\Delta X_{OX}))$$

Assuming defects are randomly distributed, using a poisson distribution to estimate probability [28], then

$$\text{P(no defect with effective thinning} > D(\Delta X_{OX})) = e^{-AD(\Delta X_{ox})}$$

where A= oxide area, D(DXox) is the area density of all defects with an effective thinning larger than DXox. Combining above equations yields

$$P(t'_{BD} < t_{BD}) = 1 - e^{-AD(\Delta X_{ox})}$$

$$\boxed{P(t'_{BD} < t_{BD}) = 1 - \frac{1}{(1 + ADS)^{1/s}}} \qquad (12)$$

where s is a measure of the degree of clustering.

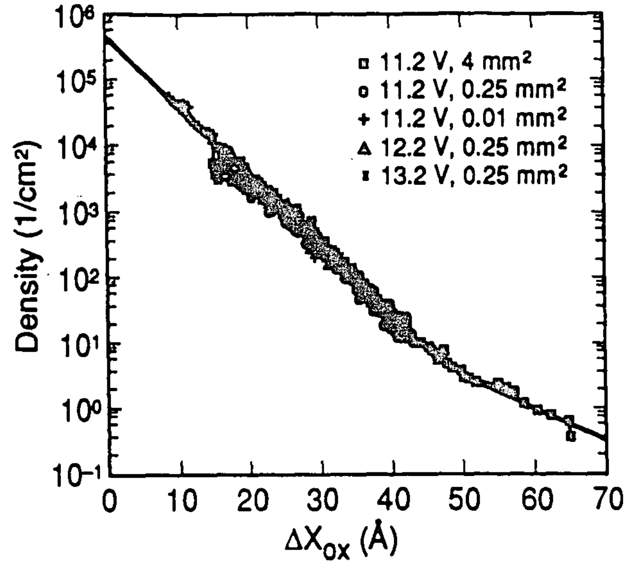


Figure 19: Defect density as a function of ΔX_{ox}, $D(\Delta X_{ox})$ using Gamma distribution for the data in Figure 18a. Once $D(\Delta X_{ox})$ is derived for a given process, many reliability parameter such as yield, failure rate, and optimum screen condition can be determined. [70]

Fig. 19 shows an example of the defect density $D(\Delta X_{ox},)$ deduced from the measured t_{BD} distribution data of fig. 18b using equations above with s=0.6. The decreasing trend of the defect density with ΔX_{ox} explains why larger capacitors are more likely to incorporate defects with larger ΔX_{ox} and exhibit shorter t_{BD}'s.

2.6.4 RELIABILITY PREDICTION

With the $D(\Delta X_{ox})$ curve, one can predict the TDDB curves and failure rates for any oxide area and stress voltage. The data in fig.19 can be curve fitted to equation

$$\boxed{D(\Delta X_{ox}) = a_1 e^{b_1 \Delta X_{ox}} + a_2 e^{b_2 \Delta X_{ox}}} \tag{13}$$

for data in fig. 9 a_1=13.1, a_2=6.3, b_1=-0.26 b_2=-0.11

Using this $D(\Delta X_{ox})$ equation, the cumulative percentage of failure F which is a function of the oxide area, the oxide field and time can be expressed as

$$F(A, V_{ox}, t_{BD}) = 1 - \frac{1}{[1 + A(a_1 e^{b_1 \Delta X_{ox}} + a_2 e^{b_2 \Delta X_{ox}})S]^{1/s}}$$

where DXox is related to V_{ox}, through t_{BD} equation above . As shown in fig. 20a&b, the solid curves, which are the cumulative failure curves calculated using $D(\Delta X_{ox})$ agree well with the actual TDDB data for different stress voltages and oxide areas.

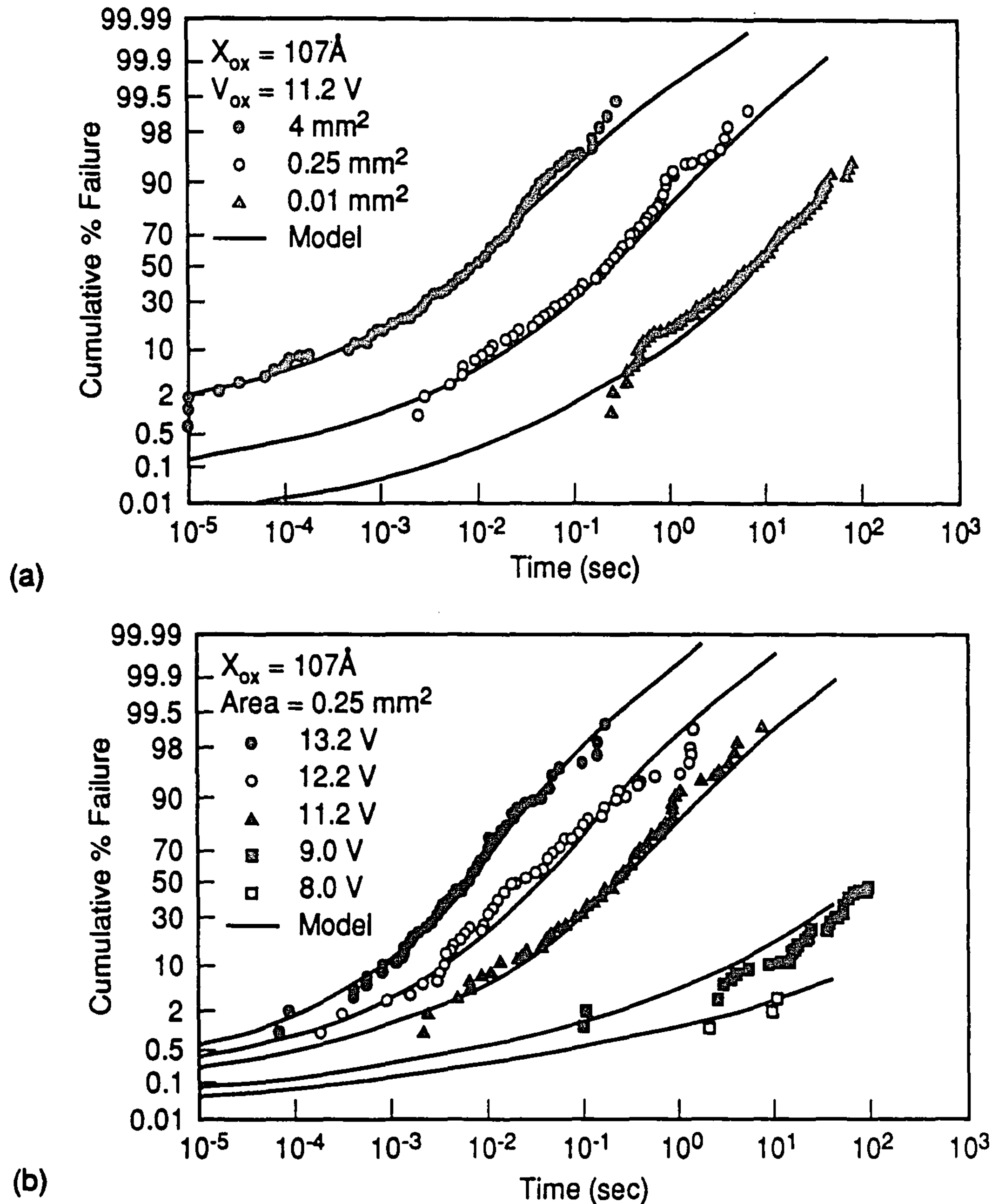


Figure 20: Cumulative percentage failure versus time for (a) different areas and (b) different stress voltages of 107Å oxide. Solid curves are derived from the defect density and ΔX_{ox} models. [70]

2.6.5 FAILURE RATE

The failure rate is an important reliability parameter. One common specification for reliability is to have a failure rate of less than 0.01% per 1000 device hours (i.e. 100 FIT's) [78] before shipping to the customers. The failure rate $\lambda(t)$ as defined as the rate of increase in the cumulative percentage divided by the percentage of remaining good devices.

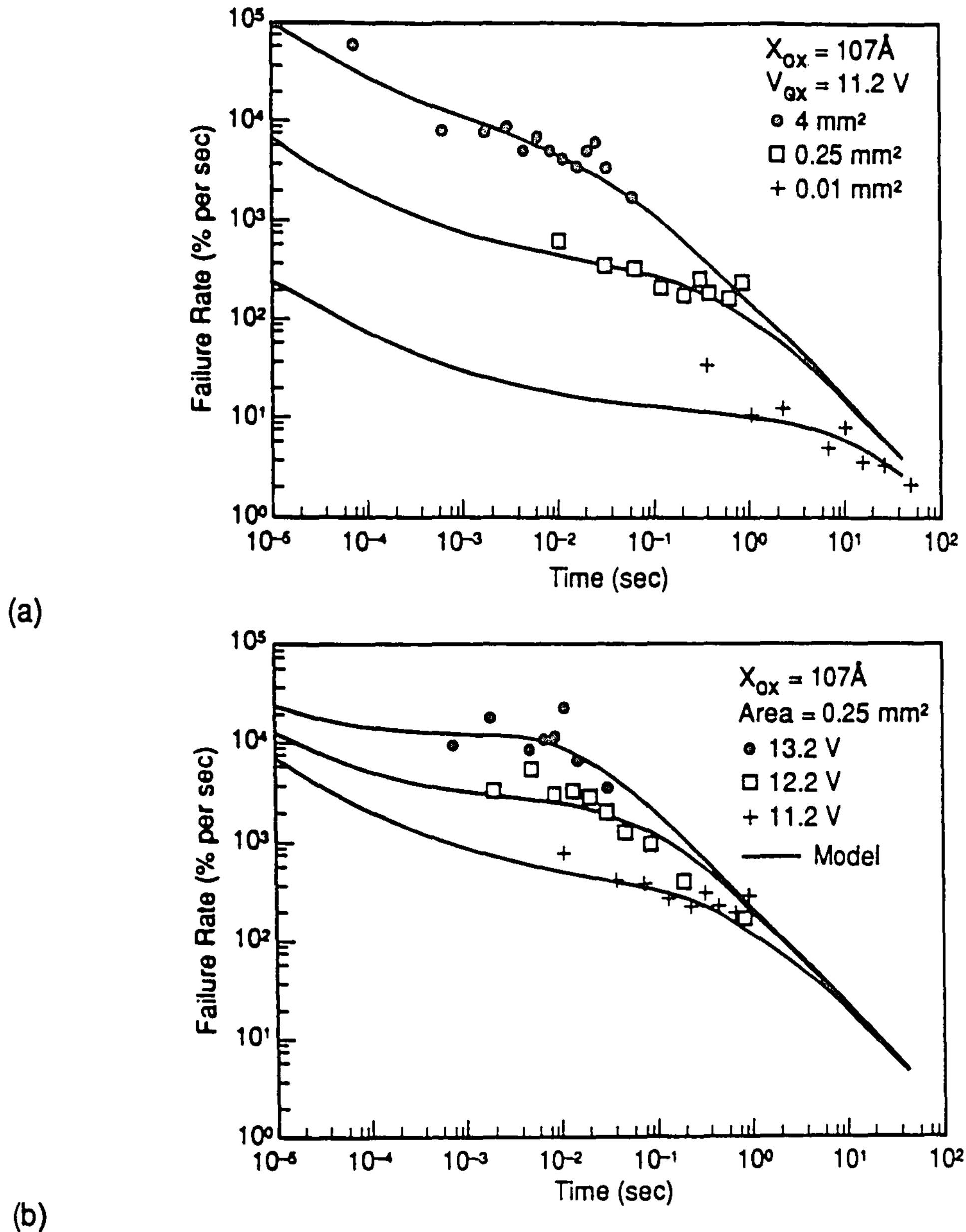


Figure 21: Failure rate versus time for (a) different areas and (b) different stress voltages of 107Å oxide. Solid curves are derived from the defect density and ΔX_{ox} models. [70]

$$\lambda(t) \equiv \frac{1}{1-P}\frac{dP}{dt}$$

combining previous equations yields

$$\boxed{\lambda(t) = \frac{A}{1+ADS}[(a_1 e^{b_1 \Delta X_{ox}} + a_2 e^{b_2 \Delta X_{ox}})](\frac{V_{ox}}{tG})} \qquad (14)$$

Fig. 21 shows comparison between calculated failure rates and experimental data, which shows reasonable agreement. Hence with the above equations one can predict the TDDB and failure rate behavior for different oxide areas and different applied voltages without additional measurements. The statistical model has been applied to screening which is an effective technique for eliminating early failure devices. Fig. 22 shows the theoritical failure rate curves at 5V predicted from model for data shown in figs 18-20. For example a product with a total oxide area $0.01 cm^2$ would not meet the failure rate specification of 0.01% per 1000 hours until after 100years. Only then would a sufficient number of weaker devices have died to meet the failure goal. As the upper axis of fig. 22 shows one can accelerate the failure by applying higher screening voltage of 8V for 1min, or 11V for 20ms, which is oxide thickness dependent.

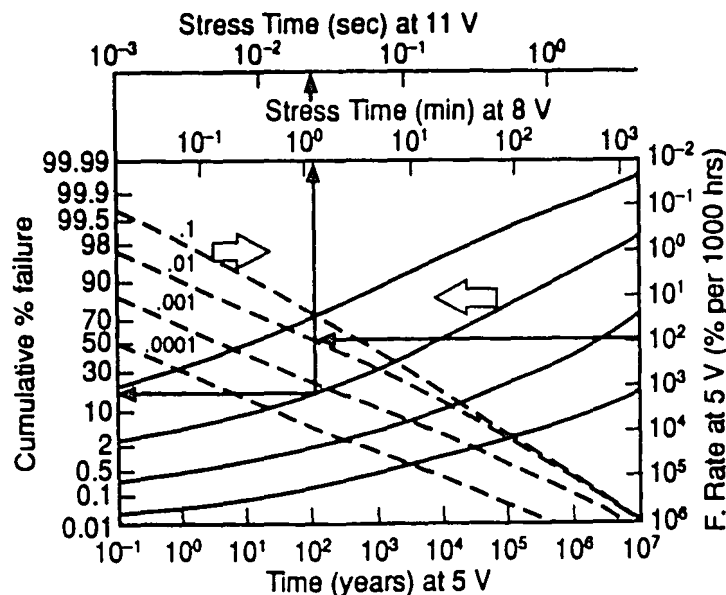


Figure 22: Theoretical curves of failure rate and cumulative percent failure for different areas at 5 V operation. These curves are used for screening and screen yield optimization. [70]

2.7 THE MOS CAPACITOR

The simplest device in the CMOS technology is the MOS capacitor, consisting of Metal gate typically made of polysilicon, oxide and a semiconductor as shown in fig. 23a. The electron energy band diagrams of MOS capacitors using both N and P substrates are also shown in fig. 23b,c&d

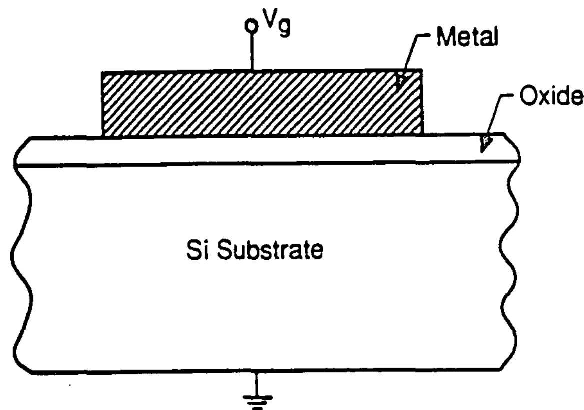


Figure 23a: An MOS (Metal Oxide Silicon) capacitor. [86]

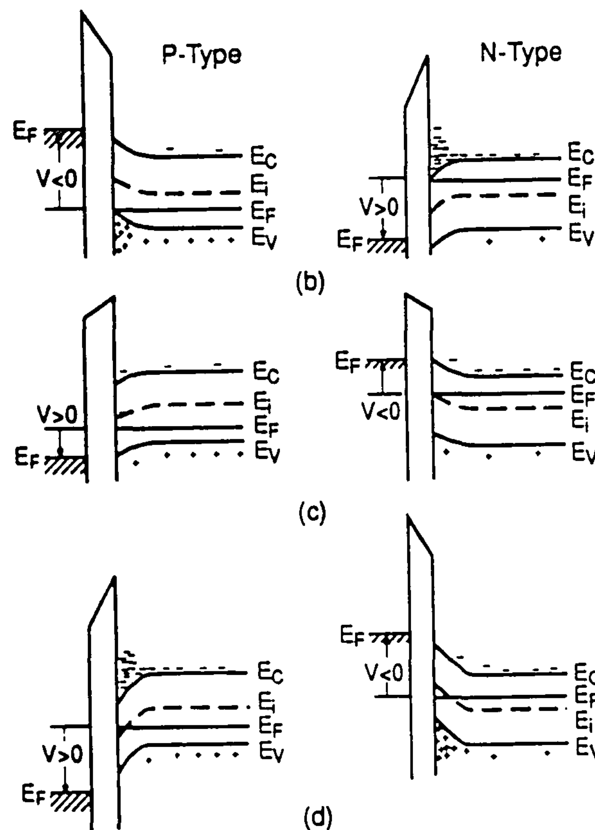


Figure 23: Energy band diagrams for ideal MOS capacitors for: (b) accumulation; (c) depletion; and (d) inversion [86]

E_C and E_V are the conduction and valence bands edges, E_i is the intrinsic level which is as the center of the energy bandgap. For the n-type semiconductor the Fermi level is closer to the conduction band whiles in a p-type material it is closer to the valence band.

The MOS capacitor structure has been subject to extensive studies over many years and a detailed study and related history can be found elsewhere [79]. Thus my discussion will encompass the characteristics of the MOS capacitor relevant to the gradual development of the MOS transistor theory.

2.8 MODES OF OPERATION

2.8.1 FLATBAND

At zero applied gate voltage condition the Fermi levels in the bulk silicon and the gate material line up . But this condition is achieved only when a given gate voltage has been applied namely the flatband voltage. The charges in the oxide and at the $Si - SiO_2$ interface acts in a manner to reduce the work function difference between the gate and bulk semiconductor hence a reduction in the flatband voltage required to align the Fermi levels. Let $Q = Q_m + Q_{ot}^{eff} + Q_{it} + Q_f$ be the total charge of the MOS system then the flatband voltage is given as

$$\boxed{V_{FB} = \phi_{MS} - \frac{Q}{C_{ox}}} \qquad (15)$$

2.8.2 ACCUMULATION

As V_{gb} decreases below V_{FB} i.e. more negative, the negative charge in V_{gb} will cause a negative change in Q'_g which must be balanced by a positive change in Q'_c. Thus holes will accumulate at the surface as shown in fig. 23b giving rise to the conditions:

$$V_{gb} < V_{FB},\ Q'_c > 0 \text{ and } (\frac{E_g}{2q} - \phi_F) \leq \phi_s \leq 0,\ \text{where } \phi_s = \frac{E_{ib} - E_{is}}{q}$$

2.8.3 DEPLETION AND INVERSION

Assume the case in which V_{gb} increases above V_{FB}, the total charge on the gate Q_g will become more positive than the value at flatband. The positive Q_g must be balanced by negative change in Q_c' to satisfy charge neutrality. Hence $V_{gb} > V_{FB}$, $Q_c' < 0$, $0 < \phi_s < \phi_f$

If V_{gb} is not much higher than V_{FB}, the positive potential at the surface with respect to the bulk will simply drive holes away from the surface leaving it depleted as illustrated in fig. 23c.

As V_{gb} is increased further, more acceptor atoms are uncovered and ϕ_s becomes sufficiently positive to attract a significant number of free electrons to the surface from electron-hole generation in the depletion region. Eventually at a significantly high V_{gb} the density of electrons will exceed that of holes at the surface leading to the condition of "surface inversion" as shown in fig 23d.

2.9 ELECTRIC FIELD EFFECT AT SEMICONDUCTOR SURFACE

Assuming that the MOS system is in thermal equilibrium, then the electrostatic potential in the system ϕ, can be related to the total space charge density ρ by Poisson's equation

$$\frac{d^2\phi}{dx^2} = \frac{-\rho}{\varepsilon_s} = \frac{dE_x}{dx}.$$

Since the total charge in the material is the sum of fully and partially ionized species is given by

$$\rho = q(N_D^+ - N_A^- + p - n).$$

But deep in the bulk of the silicon the electrostatic potential and electric field are approximately equal to zero thus:

$$N_D^+ - N_A^- = n - p$$

At the surface the electrostatic potential ϕ_s increases with increasing gate voltage can be seen as a direct measure of the amount of band-bending occurring at the surface of the silicon. Assuming complete ionization of carriers implies: $N_D^+ = N_D$, $N_A^- = N_A$ and thus

$$N_D - N_A = 2n_i \sinh(\frac{q\phi_F}{kT}).$$

Near the surface the carrier concentrations are given as: $n = n_i e^{-\frac{q(\phi_F-\phi)}{kT}}$ and $p = n_i e^{\frac{q(\phi_F-\phi)}{kT}}$, hence the Poisson's equation could be written as

$$\boxed{\frac{d^2\phi}{dx^2} = -\frac{2qn_i}{\varepsilon_s}\left\{\sinh(\frac{-q\phi_F}{kT}) - \sinh[\frac{q(\phi-\phi_F)}{kT}]\right\}}. \qquad (16)$$

Letting $U = \frac{q\phi}{kT}$, $U_F = \frac{q\phi_F}{kT}$, $U_s = \frac{q\phi_s}{kT}$ and $\sinh(U_F) = -\sinh(U_F)$

$\Rightarrow \frac{d^2U}{dx^2} = \frac{1}{L_D^2}[\sinh(U-U_F) + \sinh(U_F)]$ where $L_D = \left(\frac{\varepsilon_s kT}{2n_i q^2}\right)^{\frac{1}{2}}$ is the intrinsic Debye length.

Integrating from deep in bulk of the substrate to the surface of the silicon:

$$\Rightarrow \int_0^{\frac{\partial u}{\partial x}} \left(\frac{\partial u}{\partial x}\right) d\left(\frac{\partial u}{\partial x}\right) = \frac{1}{L_D^2}\int_0^u [\sinh(U-U_F) + \sinh(U_F)]du$$

which implies that the electric field at the surface is given by:

$$\boxed{E_x = \frac{kT}{q}\left\{\frac{2}{L_D^2}[\cosh(U-U_F) - \cosh(U_F) + U\sinh(U_F)]\right\}^{\frac{1}{2}}} \qquad (17)$$

By Gauss' law the total charge density in silicon is given by: $Q_T = -\varepsilon_s E_s = Q_{inv} + Q_{SD}$ where Q_{inv} is the charge density contained in the inversion layer and Q_{SD} is the charge density in the surface depletion layer. Using the previous equations Q_T can be formulated as

$$Q_T = -2qn_i L_D \{2\cosh(U-U_F) - \cosh(U_F) + U_s \sinh(U_F)\}^{\frac{1}{2}}.$$

If the inversion layer has not been formed yet, then the surface is intrinsic. As a result $\phi_s = \phi_F$ and Q_{inv}=0 then:

$$Q_{SD} = -2qn_iL_D\{2[1-\cosh(U_F)+U_F\sinh(U_F)]\}^{\frac{1}{2}}$$

If $Q_{inv} \neq 0$ then:

$$Q_n = Q_{inv} = -qn_iL_D\int_{u_F}^{u_s}\frac{e^{(U-U_F)}dU}{\{2[\cosh(U-U_F)-\cosh(U_F)+U\sinh(U_F)]\}^{\frac{1}{2}}}$$

A similar expression could be developed for Q_p (positive) charge density associated with a p-type surface inversion layer in an n-type substrate.

Grove et al[80] have shown that, for high values of surface potential, practically all of the charge density associated with Q_T lies within the inversion layer . The physical implication is that once an inversion layer has been formed any additional surface charge resulting from increased gate voltage will lie within the inversion layer and hence increase surface conductance.

At strong inversion: $\phi_s \cong 2\phi_F$

$$\Rightarrow Q_{SD\max} = Q_T \cong -2qn_iL_D\left[\frac{4q\phi_F}{kT}\sinh(\frac{q\phi_F}{kT})\right]^{\frac{1}{2}},$$

under typical conditions such that $\frac{q\phi_F}{kT} > 1 \Rightarrow \sinh(\frac{q\phi_F}{kT}) \cong \frac{1}{2}e^{\frac{q\phi_F}{kT}}$ thus approximating

$$\boxed{Q_{SD\max} \cong [4q\varepsilon_sN_A\phi_F]^{\frac{1}{2}} = -qN_AX_d} \tag{18}$$

where $X_d = \frac{2n_iL_D}{N_A}\{2[\cosh(U_s-U_F)-\cosh(U_F)+U_s\sinh(U_F)]\}^{\frac{1}{2}}$

At strong inversion conditions the depletion width reaches a maximum and can be approximated as

$$X_{d\max} \cong \left[\frac{4k_s\varepsilon_s\phi_F}{qN_A}\right]^{\frac{1}{2}} \tag{19}$$

and the inversion layer charge can also be approximated by:

$$Q_n = qN_I \cong \sqrt{2}L_D n_i e^{\frac{q\phi_s}{2kT}} \tag{20}$$

2.10 CAPACITANCE CHARACTERIZATION

As discussed earlier the width of the surface depletion region for an MOS structure in equilibrium will remain virtually constant after the formation of a surface inversion layer even at higher gate voltages. However if a condition of non-equilibrium exists where the charge density in the inversion layer is unable to follow a high frequency A.C. variation of the applied gate voltage, it follows that the capacitance under inversion conditions will be a function of frequency.

2.10.1 DEPENDENCE OF GATE TO SUBSTRATE CAPACITANCE ON GATE VOLTAGE

The applied voltage is related to the surface potential ϕ_s, the voltage across the insulator V_{ox} and the gate-semiconductor work function ϕ_{MS} by

$$V_G = V_{ox} + \phi_{MS} + \phi_S$$

assuming charge neutrality and recalling V_{FB} equation implies

$$V_G - V_{FB} = \phi_s - \frac{Q_T}{C_{ox}}$$

but the gate to substrate capacitance is given by

$$C = \frac{dQ_G}{dV_G} = \left[\frac{dV_{ox}}{dQ_G} + \frac{d\phi_s}{dQ_G}\right]^{-1} = \left[-\frac{dQ_T}{d\phi_s} - \frac{dQ_{ss}}{d\phi_s}\right]$$

where $C_{ox} = \frac{dQ_G}{dV_{ox}}$, $C_s = -\frac{dQ_T}{d\phi_s}$ and $C_{ss} = -\frac{dQ_{ss}}{d\phi_s}$ are the oxide , surface space charge and interface charge capacitance respectively. Combining these expressions and assuming that $C_{ss} = 0$ the gate to-substrate capacitance could be written as

$$\boxed{\frac{1}{C_{LF}} = \left[\frac{1}{C_{ox}} + \left(\frac{1}{C_s + C_{ss}}\right)\right] = \frac{1}{C_{ox}} + \frac{1}{C_s}} \qquad (21)$$

2.11 FREQUENCY DEPENDENCE OF CAPACITANCE

Under inversion conditions, the minority carriers in the substrate must provide the charge required to terminate the gate field. If the response time of the surface inversion layer is denoted by τ_{inv}, then the inversion layer can follow a signal of frequency f applied to the gate if $\omega f \leq 1/\tau_{inv}$, however for frequencies such that $\omega f \geq 1/\tau_{inv}$, the inversion layer will not follow the gate signal and a non-equilibrium condition exist. Hence while the C-V curve associated with an MOS capacitor will be independent of frequency under accumulation and depletion conditions for all frequencies of practical interest this will not be true when a surface inversion layer is present.

2.11.1 C_{GS} AT LOW FREQUENCY

For low frequencies such that $\omega f << 1/\tau_{inv}$ the charge in the inversion layer can follow the gate signal and a condition of thermal equilibrium will exist. This follows that the capacitance associated with the surface space charge region C_s can be obtained by differentiating the Q_T equation presented earlier.

Hence at low frequency

$$\boxed{C_s = \frac{\varepsilon_s q(p_s - n_s + N_D - N_A)}{Q_T}} \qquad (22)$$

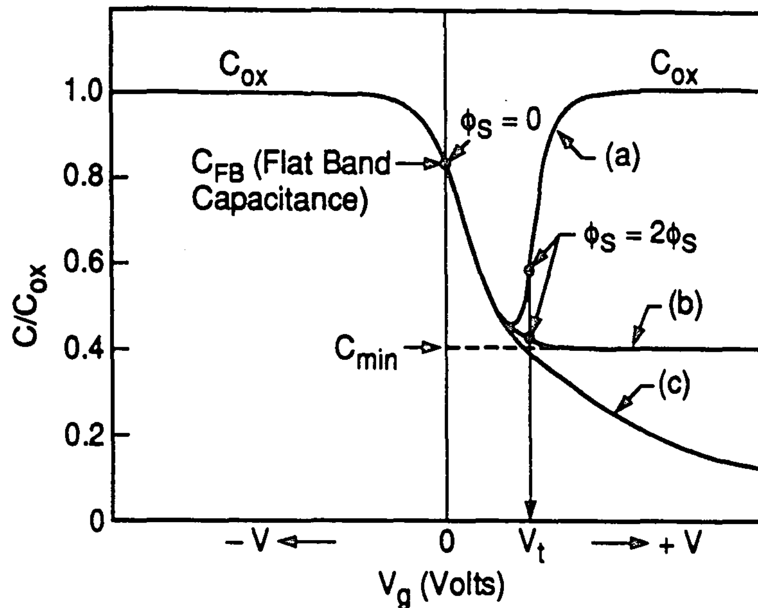


Figure 24: MOS capacitance-voltage curves. (a) low frequency; (b) high frequency; and (c) deep depletion. [86]

Under accumulation conditions the capacitance is approximately equal to C_{ox}. When the surface depletion region forms, the space charge capacitance adds in series to the dielectric capacitance; consequently, the total gate-to-substrate capacitance decreases.

At inversion the charge in the layer follows a low frequency gate signal and the charge in the surface depletion region will remain constant at its maximum value. Hence for low frequency gate signal $\frac{dQ_{SD}}{dV_G} \to 0$ as a result any small-signal variations in the gate field will now be accommodated by fluctuations in the charge stored in the surface inversion layer and gate-substrate capacitance will rise again and approach C_{ox}.

C_{GS} AT LOW FREQUENCY

For frequencies such that f>> $1/\tau_{inv}$, the charge in the surface inversion layer does not follow the gate signal thus resulting in non-equilibrium conditions. Q_{inv} remains

constant for a given D.C. bias $\frac{dQ_{inv}}{dV_G} \rightarrow 0$. The high frequency gate-substrate capacitance will be equal to C_{ox} in series with the minimum capacitance associated with the maximum width of the depletion region. The capacitance of the surface space charge region when the depletion region is at its maximum width will be denoted by $C_{SD\min}$ which can be expressed as

$$C_{SD\min} = -\frac{dQ_{SD\max}}{d\phi_s} = qN_A\left[\frac{\varepsilon_s}{2qN_A\phi_s}\right]^{\frac{1}{2}},$$

thus yielding a high frequency capacitance as

$$\boxed{\frac{1}{C_{HF}} = \left[\frac{1}{C_{ox}} + \frac{1}{C_{SD\min}}\right]} \tag{23}$$

2.12 THE MOS TRANSISTOR

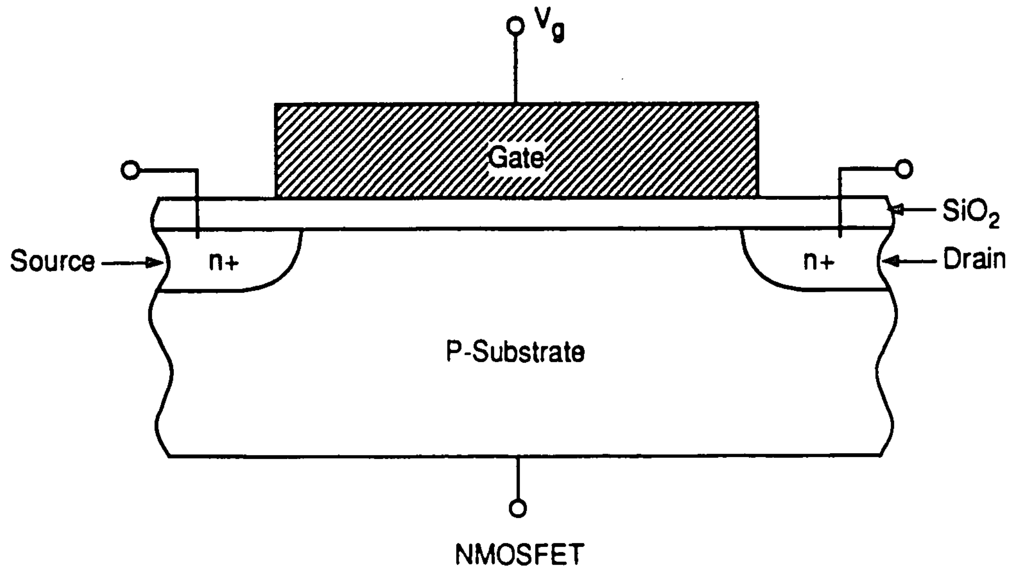


Figure 25: An MOS Field Effect Transistor (FET).

The MOS transistor is fabricated by attaching two p-n junctions at both sides of the MOS capacitor. As seen in fig 25 , it hence becomes a four terminal device: gate, source, drain and substrate or body. The gate is typically made of polysilicon material whiles the two n^+ regions for an NMOS are isolated by the substrate and no current flows unless the surface of the substrate is inverted by an adequate gate voltage. At this point an inversion layer is formed as described earlier, and so that positive voltage on the drain can attract electrons to flow from the source to drain.

2.12.1 BAND STRUCTURE AT DRAIN EDGE

At zero drain bias, only a built-in potential V_{bi} exists between the p-n junction to counter balance the diffusion force; hence we have p-n junction equilibrium and the Fermi level is constant along the channel. When $V_{gs} > V_T$, the band bends $2\phi_B$ i.e. $\phi_s = 2\phi_B$, thus making the conduction band E_C closer to E_F as shown in fig. 26a.

Now if a positive V_{ds} is applied at the n^+ of the p-n junction, the reverse bias will create a larger potential barrier ($\phi_B + V_{ds}$) across the p-n junction, thus lowering the quasi-Fermi level for electrons (E_{Fn}) by V_{ds} and placing the system in non-equilibrium. Therefore the band bending is not enough to bring E_C near E_{Fn} to cause inversion as seen in fig 26b. As a result higher V_{gs} must be applied to bend the bands by ($2\phi_B + V_{ds}$) and lower the E_C under the gate closer to E_{Fn} as shown in fig. 26c. Hence the gate voltage for inversion to occur in a MOSFET channel is a function of the applied drain voltage.

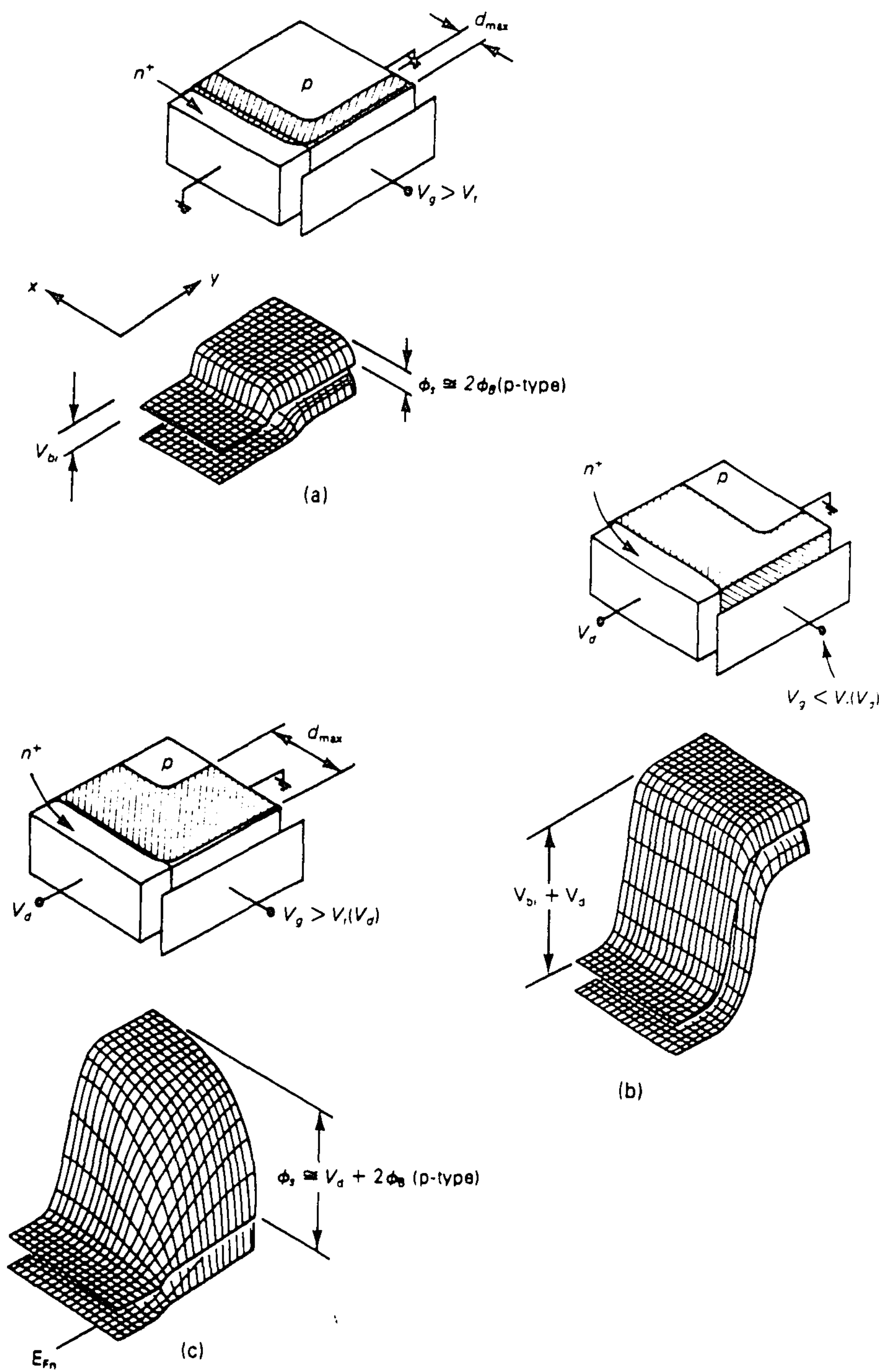


Figure 26: Energy band diagrams near the drain of NMOSFET with different gate and drain bias conditions: (a) $V_g > V_t$, $V_d = 0$; (b) $V_g < V_t(V_d)$, $V_d > 0$; and (c)$V_g > V_t(V_d)$, $V_d > 0$. [5]

2.12.2 CHANNEL CROSS-SECTION FOR NMOSFET

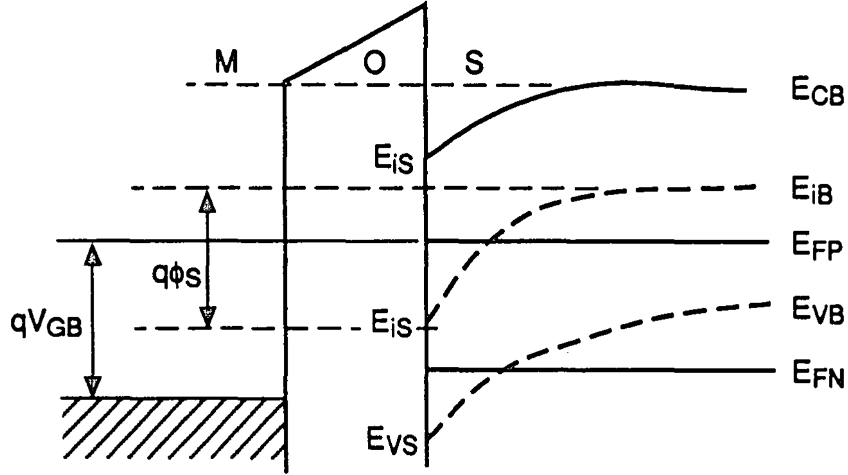


Figure 27: Cross-section of N-channel band structure.

$q\phi_n = E_{iB} - E_{Fn}$, $qV_{GB} = E_{Fp} - E_{Fn}$, $V_{gb} = V_{ss} + V_{sb}$, $\phi_s = \dfrac{E_{iB} - E_{is}}{q} = \varphi_s - V_{SB}$ and also

$\varphi_s = \varphi_{so} + V(y)$ hence

$$E_{Fp} - E_{Fn} = qV_{SB} + qV(y) = q(\phi_n - \phi_p)$$

which implies that $n_p = n_i e^{\frac{q}{kT}(\phi(x)-\phi_n)}$ and $p_p = n_i e^{\frac{q}{kT}(\phi_p - \phi(x))}$.

At the surface x=0 and $\phi(x) = \phi_s$, $\phi_p = \phi_n$ hence the condition of thermal equilibrium can be written as

$$n_p p_p = n_i^2 e^{\frac{q}{kT}(\phi_p - \phi_n)}$$

The electron current density is then given as

$$J_n = qn_p \mu_n E + qD_n \frac{\partial n_p}{\partial y} \text{ where } D_n = \mu_n \frac{kT}{q}$$

But also $\dfrac{\partial n_p}{\partial y} = \dfrac{\partial n_p}{\partial \phi}\dfrac{\partial \phi}{\partial y} + \dfrac{\partial n_p}{\partial \phi_n}\dfrac{\partial \phi_n}{\partial y}$, $J_n = -q\mu_n n_p \dfrac{\partial Q_n}{\partial y}$ and $Q_n = -q\int_0^{x_d} n_p(x,y)dx$

knowing that $Q_s = Q_n + Q_B$ and $V_{gb} = V_{ox} + \phi_{gs} + \phi_s$ then

$$Q_n = -C_{ox}\left[V_{gs} - V_{FB} - \varphi_{so} - V(y) - \frac{Q_B(y)}{C_{ox}}\right]$$

where $Q_B(y) = -\lambda C_{ox}\sqrt{2\phi_F + V_{SB} + V(y)}$ and

$$\boxed{\lambda = \frac{\left[2k_s\varepsilon_0 qN_B\right]^{\frac{1}{2}}}{C_{ox}}} \text{ ---- the Body Effect parameter.} \quad (24)$$

2.12.3 LONG CHANNEL NMOSFET MODEL

The assumptions made for the long channel devices is that mobility is constant, λ =0 implying zero bulk charge Q_B=0, for long device (L >> X_d) the gate length is much longer than the depletion width and that it is operated in strong inversion. For such a device

$$\int_0^L I_{ds}dy = \int_0^L -W\overline{\mu}_n Q_n(y)\frac{\partial V}{\partial y}dy = \int_0^{V_{ds}} -W\overline{\mu}_n Q[V(y)]dV$$

which implies that

$$I_{ds} = \overline{\mu}_n \frac{W}{L}C_{ox}\left[(V_{GS} - V_{FB} - 2\phi_F)V_{ds} - \frac{Vds^2}{2}\right]$$

including bulk charge (i.e. λ≠0) implies that

$$\int_0^L I_{ds}dy = \int_0^{V_{ds}} W\ddot{\mu}_n C_{ox}\left[V_{GS} - V_{FB} - 2\phi_F - V(y) - \frac{Q_B(y)}{C_{ox}}\right]dV$$

where $$Q_B(y) = -\lambda C_{ox}(2\phi_F + V_{SB} + V(y))^{\frac{1}{2}}$$

$$\Rightarrow I_{ds} = \beta_0 \left[(V_{GS} - V_{FB} - 2\phi_F) V_{ds} - \frac{V_{ds}^2}{2} \right] - \frac{2}{3} \lambda \left[(V_{ds} + V_{SB} + 2\phi_F)^{\frac{3}{2}} - (V_{SB} + 2\phi_F)^{\frac{3}{2}} \right]$$

which can be written in a compact form as

$$\boxed{I_{ds} = \beta_0 \left[(V_{GS} - V_T) V_{ds} - \frac{1}{2} V_{ds} (1 + \delta) \right]} \tag{25}$$

where $\delta = \frac{\lambda}{2} (2\phi_F + V_{SB})^{\frac{1}{2}}$ and $V_{T0} = V_{FB} + 2\phi_F + \lambda (2\phi_F)^{\frac{1}{2}}$ i.e. V_T at $V_{SB} = 0$

$$\boxed{V_{TL} = V_{T0} + \lambda \left[(2\phi_F + V_{SB})^{\frac{1}{2}} - (2\phi_F)^{\frac{1}{2}} \right]} \tag{26}$$

MOBILITY CONSIDERATIONS

Recognizing the fact that the channel carriers encounter collisions resulting in scattering which influence the carriers mobility thus

$$\frac{1}{\mu_n} = \frac{1}{\mu_0} + \frac{1}{\mu_s} + \frac{1}{\mu_c},$$

after mathematical manipulations the mobility in the drain current equation can be better modeled as

$$\mu_{eff} = \bar{\mu}_n = \frac{\mu_0}{1 + \frac{E_{eff}}{E_{crit}}} = \frac{\mu_0}{1 + \theta_s \left[V_{GS} - V_T + 2\lambda (V_{SB} + 2\phi_F)^{\frac{1}{2}} \right]}$$

Hence for the Long channel NMOSFET the drain currents could be approximated by:

$$\boxed{I_{ds} \cong \mu_n \frac{W}{L} C_{ox} (V_{GS} - V_T) V_{ds} = \beta_0 (V_{GS} - V_T) V_{ds}} \text{ --------LINEAR REGION} \tag{27}$$

$$\boxed{I_{ds} \cong \mu_n \frac{W}{L} C_{ox} (V_{GS} - V_T)^2 = \frac{\beta_0}{2} (V_{GS} - V_T)^2} \text{ ----SATURATION REGION} \tag{28}$$

2.12.4 SHORT CHANNEL NMOSFET MODEL

Figure 28: Graphical representation of "Charge Sharing Phenomena".

When the MOS transistor is scaled downward to achieve shorter gate lengths to increase performance and packing density in VLSI applications, two dimensional effects near the edges of the smaller transistors become significant in the static and dynamic characteristics of the short channel MOSFETs. The effect of decreasing L causes more depletion of the region under the inversion layer. This deeper depletion region is accompanied by a larger surface potential which makes the channel more attractive for electrons or holes. Thus the device can conduct more current than what would be predicted by the long channel theory derived earlier.

In the short channel device using the "charge sharing phenomena", the charges in the depletion regions I&III in fig. 28 are generated by the field lines from the source and drain thus controlling those charges. Hence the charge directly controlled by the gate

voltage is that in the region II. As a result the bulk charge $Q_{B2}(V)$ depleted by the gate voltage is greatly reduced.

The long channel drain current derived earlier could be written in a compact form for typically employed in device modeling as

$$I_{ds} = \frac{\beta_0[f-\lambda g]}{1+\frac{\theta_c}{V_{ds}}[f+\lambda g]+\theta_c V_{ds}}$$

where $f = (V_{GS} - V_{FB} - 2\phi_F)V_{ds} - \frac{V_{ds}^2}{2}$ and $\lambda g = \frac{1}{C_{ox}}\int_0^{V_{ds}} Q_B(V)dV$

SHORT CHANNEL EFFECT

$$\lambda g^* = \frac{1}{C_{ox}}\int_0^{V_1} Q_{B1}(V)dV + \frac{1}{C_{ox}}\int_{V_1}^{V_2} Q_{B2}(V)dV + \frac{1}{C_{ox}}\int_{V_2}^{V_{ds}} Q_{B1}(V)dV$$

Let the depletion widths at the source be X_S and the drain be X_D and assuming that $N_d >> N_a$ then

$$X_s = \alpha(\phi_B + V_{SB})^{\frac{1}{2}} \text{ and } X_d = \alpha(\phi_B + V_{SB} + V_{ds})^{\frac{1}{2}}$$

where $\alpha = \sqrt{\frac{2k_s\varepsilon_0}{qN_A}}$ and $\phi_B = \frac{kT}{q}\ln\frac{N_A N_D}{n_i^2}$

AT SOURCE END: $(X_s + X_j)^2 = (X_j + y_1)^2 + x_1^2$

AT DRAIN END: $(X_D + X_j)^2 = (L - u_2 + X_j)^2 + x_2^2$

The bulk charge in each region due to the charge sharing phenomena can be written as

$$Q_{B1}(V) = \lambda C_{ox}(\sqrt{V_{SB} + \varphi_{S0} + V_1})\frac{y}{y_1}$$

$$Q_{B2}(V) = \lambda C_{ox}(\sqrt{V_{SB} + \varphi_{S0} + V(y)})$$

$$Q_{B3}(V) = \lambda C_{ox}[\ \sqrt{V_{SB} + \varphi_{S0} + V_2}\ ]\frac{L-y}{L-y_2}$$

after manipulation and solving, the change in threshold voltage due to the short channel effect is given as

$$\boxed{\Delta V_{TSC} = -\lambda f_0 \sqrt{V_{SB} + 2\phi_F}} \qquad (29)$$

$$\text{where } f_0 = \frac{X_j}{L}\left[\left(1 + \frac{2X_s}{X_j}\right)^{\frac{1}{2}} - 1\right]$$

2.12.5 NARROW WIDTH EFFECT

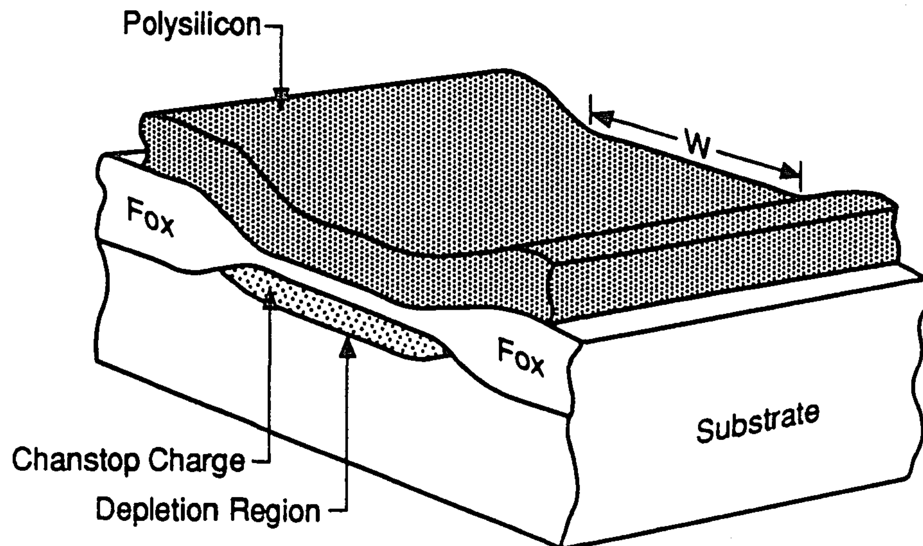


Figure 29: Realistic picture for a device with bird's beak effect on active area.

In the MOS fabrication process the field oxidation scheme is used to isolate non-communicating devices and also for the minimization of cross-talk. The thinning of the field oxide as you approach the active device region gives rise to the characteristic "bird's

beak" shape, which consequently causes the effective width of the device to differ from the mask coded width. As seen in the above figure the depletion region is not limited to just the area directly below the thin oxide. This is because some of the electric field lines emanating from the gate charges terminate on ionized acceptor atoms on the sides, which constitute what is called the "fringing field". Assuming that the side parts of the depletion region have quarter circle cross-section [82] then

$$Q_B(y) \cong WX_d(y) + \frac{2\pi X_d^2(y)}{4} = W_{eff}(y)X_d(y)$$

$$\Rightarrow W_{eff} \cong W\left[1 + \frac{\pi X_d(y)}{2W}\right]$$

but

$$X_d(y) = \sqrt{V_{SB} + V_1 + \varphi_{S0}}$$

$$Q_{B1W} = \left[1 + \frac{\pi\alpha}{2W}(V_{SB} + V_1 + \varphi_{S0})^{\frac{1}{2}} \frac{y}{y_1}\right]$$

$$Q_{B2W} = \left[1 + \frac{\pi\alpha}{2W}(V_{SB} + V_1 + \varphi_{S0})^{\frac{1}{2}}\right]$$

$$Q_{B3W} = \left[1 + \frac{\pi\alpha}{2W}(V_{SB} + V_2 + \varphi_{S0})^{\frac{1}{2}} \frac{L-y}{L-y_2}\right]$$

Hence evaluating in the same manner as the short channel, the narrow width contribution to the threshold voltage is expressed as

$$\boxed{\Delta V_{TNW} = \frac{\lambda\pi\alpha f_0}{W}(2\phi_F + V_{SB})} \qquad (30)$$

2.12.6 DRAIN-INDUCED BARRIER LOWERING

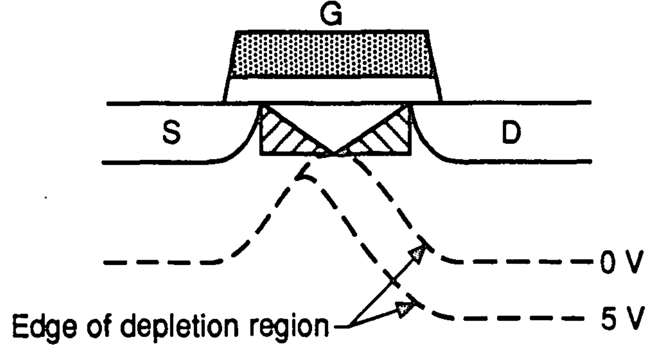


Figure 30: Shape of the depletion region in a short-channel MOSFET at zero and high drain bias. The grey areas indicate the charge shared by the gate and the junctions at zero drain bias.

The previous derivations for short channel MOSFETs had been obtained at low V_{ds} values. But if V_{ds} is increased, the depletion region width around the drain will widen. This in turn will cause a decrease in bulk charge Q_B and thus further decreasing V_T. Hence for short-channel devices, V_T becomes a decreasing function of V_{ds}.

Many attempts have been made to model this effect in short channel MOSFETS. The method of "voltage-doping transformation" which assumes that the potential distribution along the channel varies gradually with distance [83,84], has proven to be the most accurate to date. In this approach

$$\psi(x,y) = \psi_1(x) + \psi_2(y).$$

Using the boundary conditions of $\psi_1(0) = \phi_B + V_s$ and $\psi_1(0) = \phi_B + V_d$ where ϕ_B is the built-in potential, it can be shown that

$$\frac{\partial^2 \psi}{\partial x^2} = 2\frac{V_{ds}*}{L^2}\text{, and if } \phi_c = \psi_1(y) \text{ then}$$

$$V_{ds}* = (V_d - V_s) + 2(\phi_B + V_s - \phi_C) + 2[(\phi_B + V_s + \phi_C)(V_d + \phi_B - \phi_C)]^{\frac{1}{2}}$$

thus substituting the above equations into 2-D Poisson's equation yields

$$\frac{\partial^2 \phi}{\partial y^2} = \frac{q}{\varepsilon_s}\left[N(y) - \frac{2\varepsilon_s V_{DS}{}^*}{qL^2}\right]$$

This equation eventually reduces the 2-D Poisson's equation to a 1-D form . Physically this means that the influence of the lateral drain-source field on the potential is equivalent to and can be replaced by a reduction of the doping concentration according to previous equation. The contribution to threshold voltage decrease due to DIBL phenomena has thus been approximated by

$$\boxed{\Delta V_{T(DIBL)} = -\frac{4k_s\varepsilon_0}{qN_BL^2}(2\phi_F + V_{SB})} \tag{31}$$

Hence the threshold voltage of sub-micron MOSFET considering all the 2-D behaviors can be accurately modeled as

$$\boxed{V_{TH} = V_{TH}(long) - \Delta V_{TS}(shortchannel) + \Delta V_{TW}(narrowwidth) - \Delta V_{TD}(DIBL)} \tag{32}$$

2.12.7 SUBTHRESHOLD REGIME

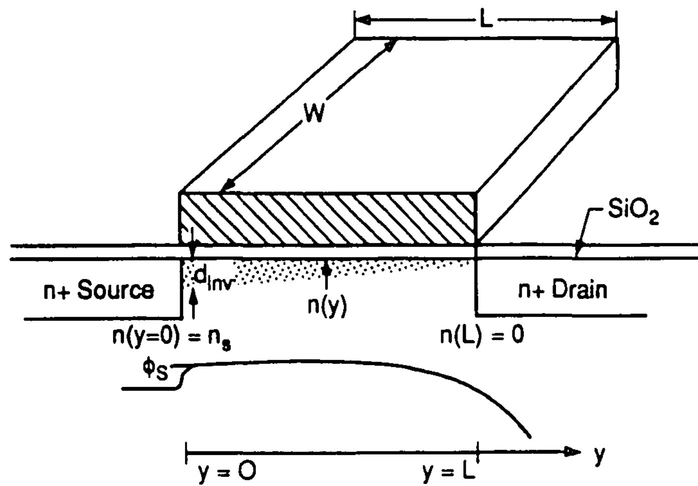


Figure 31: A cross-sectional schematic of a transistor illustrating subthreshold conduction.

The previous discussions of the MOSFET characteristics addressed the conditions when the applied gate voltage exceeded the threshold voltage of the device and the device completely turned on. But another mode of operation is the weak inversion mode in which the Si surface has changed to n-type, but is not strongly inverted, meaning minority carrier concentration is still lower than the bulk doping concentration. Unlike in the linear region in which the onset of strong inversion occurs and the minority charge density increases with V_{gs}, in the subthreshold regime, the minority carrier concentration is too low to change the band bending significantly, thereby resulting in a low electric field along the channel. Both the lower carrier concentration and low field region makes the E-field assisted drift current much less than the diffusion current, hence the subthreshold current is dominated by the diffusion component given as[85]

$$I_{ds} = AqD_n \frac{dn(y)}{dy} \cong qW \frac{dV}{dy} \int_0^{x_d} \mu_n(x,y) n(x,y) dx$$

$$Q_s = Q_n + Q_B = k_0 \varepsilon_0 \frac{d\phi}{dx} = -\lambda C_{ox} \sqrt{\frac{kT}{q}} F(\phi_s, V, V_{SB}, \phi_F)$$

where $F \cong \sqrt{\frac{q\phi_s}{kT} - 1 + e^{\frac{q}{kT}[\phi_s - (2\phi_F + V + V_{SB})]}}$ and also $Q_n = -q \int_0^{x_d} n(\phi) \frac{\partial x}{\partial \phi} d\phi$ which implies

$$Q_s(\phi_s, V) = Q_n + Q_B = -C_D(\phi_s) \frac{kT}{q} e^{\frac{q}{kT}[\phi_s - (2\phi_F + V + V_{SB})]} - \lambda C_{ox} \sqrt{\phi_s - \frac{kT}{q}}.$$

$V_{GB} = V_{GS} + V_{SB} = V_{FB} + \phi_s - \frac{(Q_s + Q_{it})}{C_{ox}}$ which implies that

$$\phi_s = \frac{V_{GS} - V_{TH}}{n} + \frac{3}{2}\phi_F + V_{SB} + (1 - \frac{m}{n})V$$

where $n = 1 + \frac{C_D + C_{it}}{C_{ox}}$, $m = 1 + \frac{C_D}{C_{ox}}$ and $n - m = \frac{C_{it}}{C_{ox}}$. Substituting the above equations into the current equation and after manipulations yields

$$I_{DS} \cong I_{D0} e^{\frac{q}{kT}[V_{GS}-V_{TH}]}\left[1-e^{-\frac{qm}{nkT}V_{DS}}\right] \quad (33)$$

where $I_{D0} = \beta_0 \left(\frac{C_D}{C_{ox}}\right)\left(\frac{kT}{q}\right)^2 \left(\frac{n}{m}\right) e^{-\frac{q\phi_F}{2kT}}$ and $\phi_F + V_{SB} < \phi_s < 2\phi_F + V_{SB}$.

SUBTHRESHOLD SLOPE

Because ϕ_s is roughly proportional to V_{gs} in the subthreshold region[86], I_{ds} is also exponentially dependent on V_{gs}. The subthreshold slope is then defined as

$$S_t = \left|\frac{\partial \log I_{DS}}{\partial V_{GS}}\right|^{-1} = 2.3\frac{kT}{q}\left[1+\frac{C_D+C_{it}}{C_{ox}}\right]\left(\frac{MV}{dec}\right)$$

A large S_t measuring a gradual slope of $\log I_{ds}$ vs. V_{ds} results in a significant amount at the off state. Also an advantage of thinner oxide is to reduce S_t thus increasing Ids. Also the gradual slope S_t can be used to monitor the incidence of the Drain-Induced Barrier Lowering phenomena since increased V_{ds} causes an increase in the subthreshold current conduction due to barrier lowering effects.

2.13 HOT CARRIER EFFECTS ON RELIABILITY

It has long been recognized that hot-electron/hole-induced device degradation can pose a limit on device scaling. This problem is more serious for an N-channel MOSFET than P-channel due to the higher impact ionization rate of electrons as compared to holes.

Device lifetime may be predicted reliably using the substrate current I_{sub}, which if unchecked can overload the substrate-bias generator, cause substrate potential fluctuations or electron injection into the substrate and induce snap-back breakdown and CMOS latchup [87,88]. Studies[89-91] have produced convincing evidence that the generation of interface traps (surface states) is the dominant cause of MOSFET degradation.

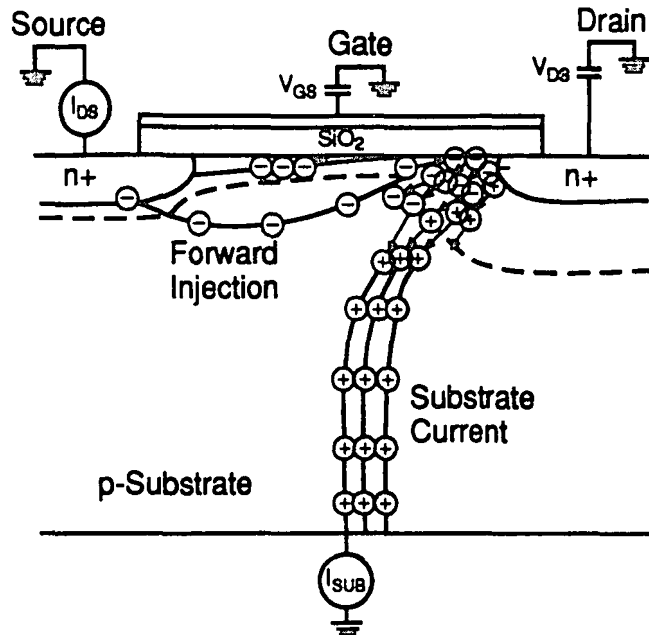


Figure 32a: Schematic of hot-electron effects in an n-channel MOSFET.

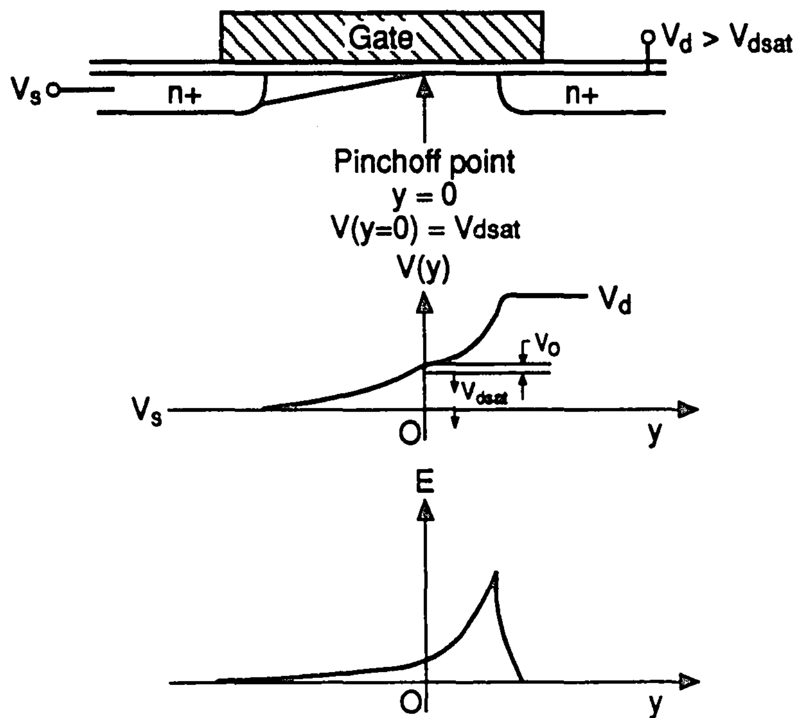


Figure 32b: Schematics of channel potential and horizontal electric field of a MOSFET biased at saturation region.

2.13.1 SUBSTRATE CURRENT

For an N-channel MOSFET the electrons in the channel near the drain end experience a very large field which can cause impact ionization resulting in the generation

of electron-hole pairs via the process of weak avalanche multiplication. The generated electrons are attracted to the drain adding to the channel current, the holes are collected by the substrate contact resulting in the "substrate current" as shown in the fig. 32a above.

Analytically the substrate current can be expressed as [92]

$$I_{SUB} = (M-1)I_{ds}$$

I_{ds} is the channel current and M is the avalanche multiplication factor given by

$$M = \frac{1}{1-\int \alpha dx}$$

where α is the impact ionization coefficient. For low level avalanche multiplication M≅1 hence the substrate current equation reduces to

$$I_{SUB} = I_{ds}\int_0^L \alpha dx.$$

As explained in reference [91] α is a strong function of peak E-field at the drain edge given by $\alpha(E) = Ae^{-\frac{B}{E(x)}}$ where A and B are constants obtained by fitting data to equations above. Hence can be expressed as

$$\boxed{I_{SUB} = \frac{I_{ds}AE_{max}e^{-\frac{B}{E_{max}}}}{B[dE(x)/dx]}} \qquad (34)$$

where $E_{max} = \dfrac{V_{ds}-V_{dsat}}{l+L_{\bar{n}}}$, $l = 0.2T_{ox}^{\frac{1}{3}}X_j^{\frac{1}{2}}$ and $L_{\bar{n}}$ is the length of the LDD region.

As shown in my experimental graphs I_{sub} has a maximum at low gate voltages ($\cong 0.4V_{ds}$) where a significant number of carriers are available. The device lifetime dependence on I_{sub} can be expressed in a power law relationship as[94]

$$\boxed{\tau_{LIFE} = a(I_{SUB})^{-b}}. \qquad (35)$$

2.13.2 SYMPTOMS OF DEGRADATION

The hot carriers, when injected into the oxide, can be trapped and also result in the generation of interface states. MOSFET degradation shows up in at least three quantities; a shift in the threshold voltage[93], a shift in the subthreshold current swing ΔS_t [9] and a reduction in the transconductance in the linear or saturation region Δg_m [10] as shown in the figures below.

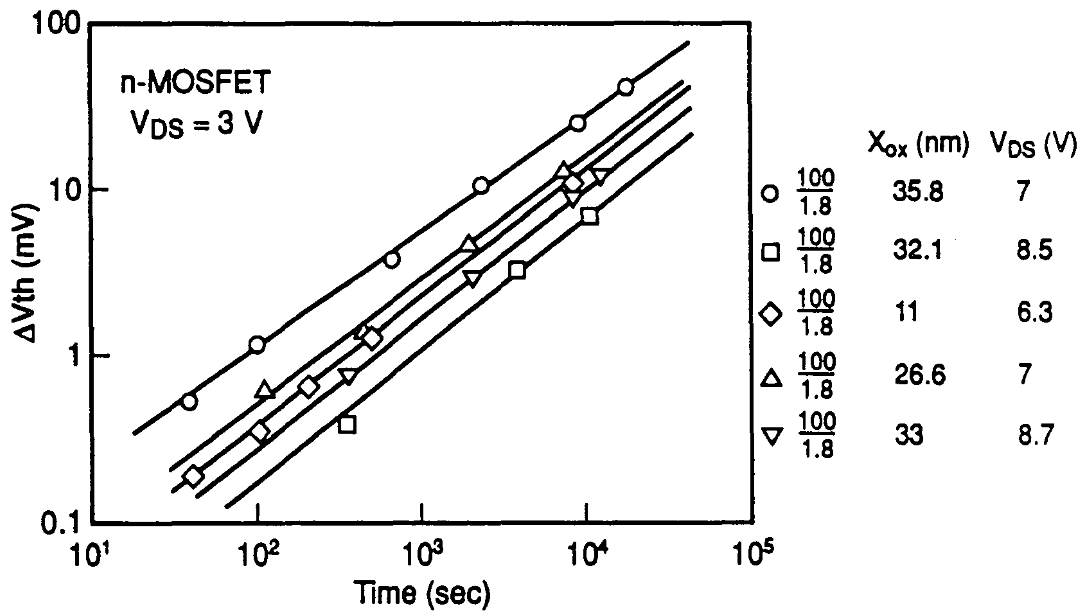


Figure 33a: ΔV_{th} increases as t^n with n ~ 0.65 in this figure. [70]

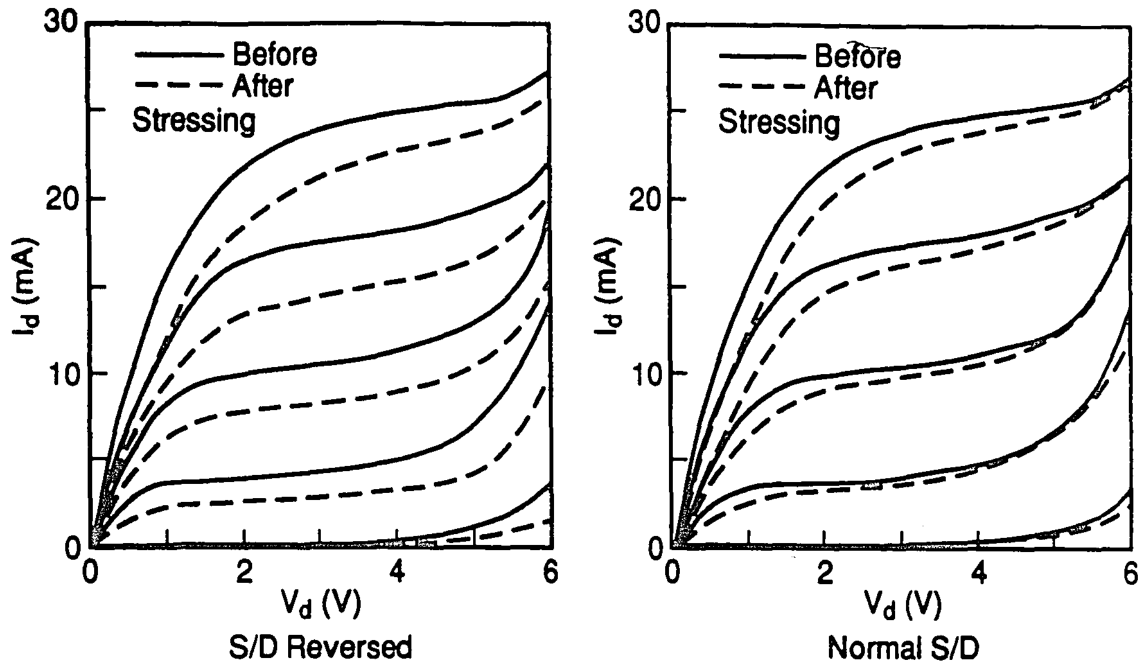


Figure 33b: Degradation characteristics after stressing. [106]

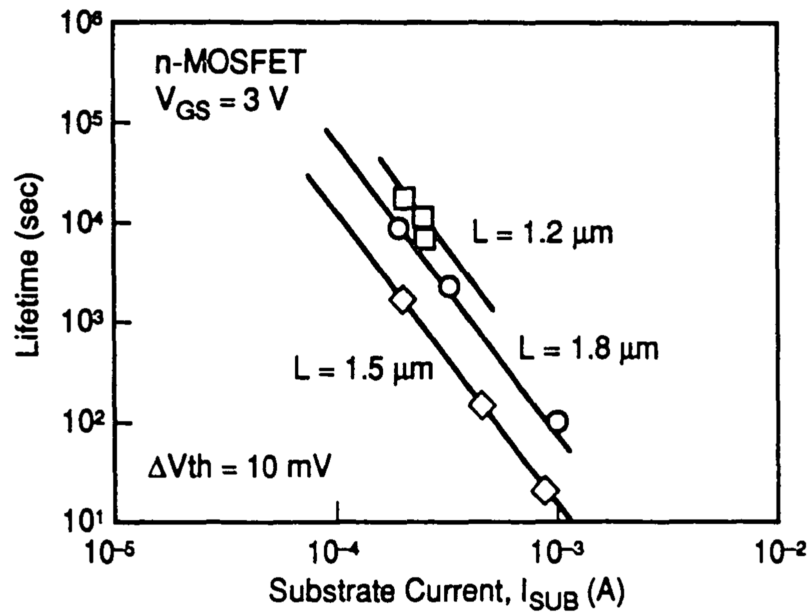


Figure 33c: Lifetime as a function of substrate current, lifetime defined as 10 mV shift in threshold voltage (V_{th}). [106]

The device lifetime power law relationship mentioned earlier when plotted on a log-log scale as shown above yields a straight line. Thus showing that I_{SUB} is a good indicator of the peak E-field and also that the lifetime is a strong function of device channel length.

Controlling hot carriers in today's technologies involve the implementation of Lightly Doped Drain (LDD) engineering in the attempt to limit the extent of the high field region. As shown in the fig. 34 below the peak field at the drain edge which is the dominant culprit behind generation and injection of hot electrons, could be moved deeper into the bulk and further away from the drain edge and thus separating it from the path of the channel current and effectively minimizing impact ionization.

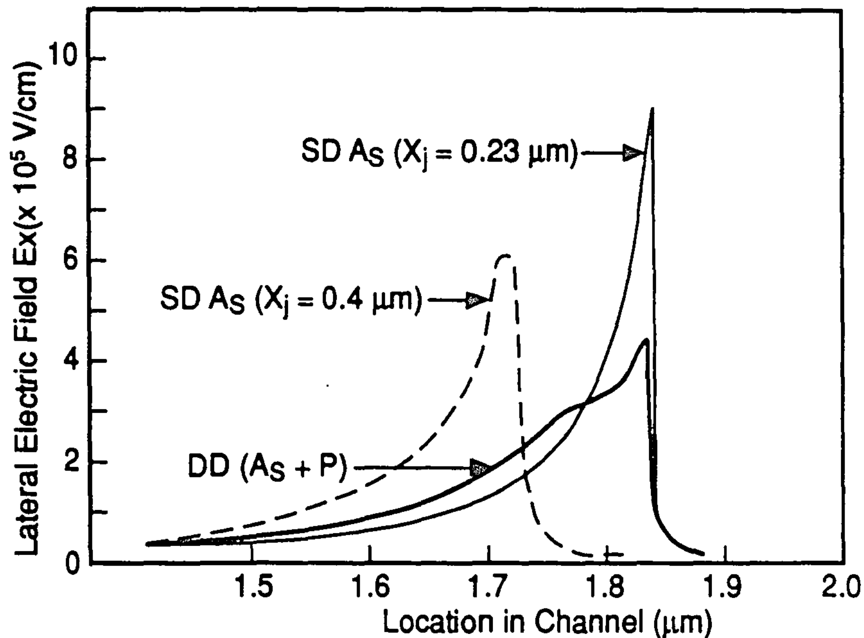


Figure 34: Lateral electric field distributions associated with LDD structures. [107]

2.14 GATE OXIDE THICKNESS LIMITATIONS FOR ULSI

Although thinner gate oxides have allowed a certain increase in performance for MOS circuits, studies have shown that below 50Å the oxide begins to have noticeable

amounts of direct tunneling current. This can limit the transfer characteristics of the MOS transistor by modifying its voltage transfer characteristics and hence eliminating its ability to act as an effective latch in MOS storage devices.

2.14.1 STATIC LOGIC LIMITATIONS

Static CMOS logic has large noise margins, which are nearly equal to the power supply voltage, and has a very low standby power dissipation. The voltage transfer characteristics of a CMOS inverter are shown in fig. 35a. The critical voltages V_{OH} and V_{OL} are the equilibrium high and

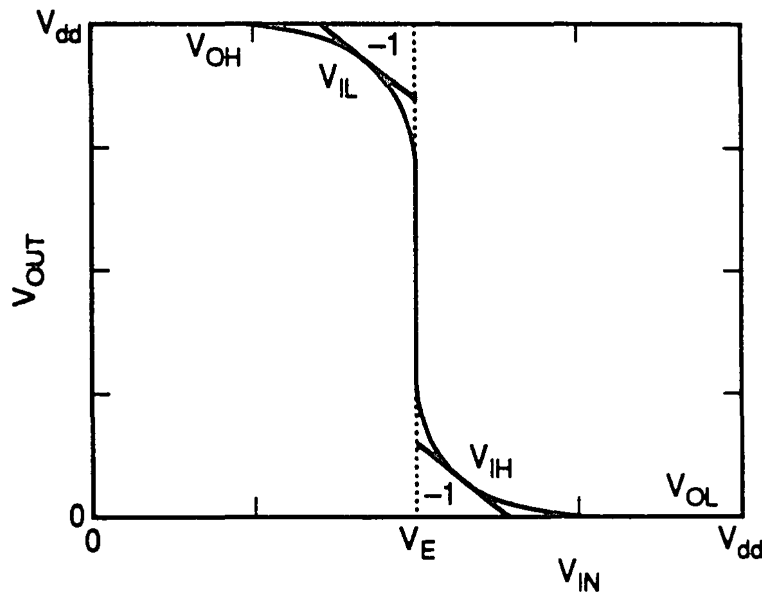


Figure 35a: Transfer voltage characteristics of a CMOS inverter with the critical voltages V_{OH}, V_{IL}, V_{IH}, and V_{OL} labeled where V_E is the only point at which the NMOS and PMOS transistors are both saturated and the output voltage is equal to $V_{dd}/2$. [62]

low voltages for a chain of inverters without noise. As explained in [96], V_{IL} and V_{IH} are the points at which $\frac{\partial V_{out}}{\partial V_{in}} = -1$. If there is sufficient input noise to increase a low input voltage value V_{IL} or to decrease a high input voltage value below V_{IH}, then the noise will

disturb the proper logic operation of the inverter. The noise margins of the CMOS inverter are then defined to be and shown in fig. 35b as

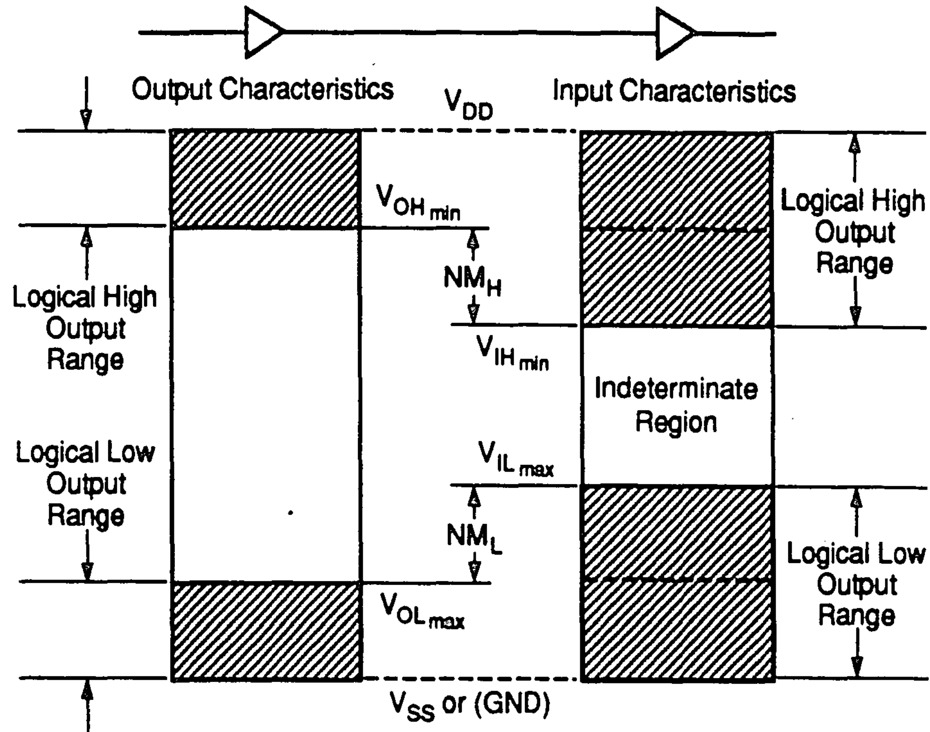


Figure 35b: Noise margin definitions.

$$NM_L = \left|V_{IL_{max}} - V_{OL_{max}}\right| \cong \frac{3V_{DD} - 3\left|V_{tp}\right| + 5V_{tn}}{8}$$

$$NM_H = \left|V_{OH_{min}} - V_{IH_{min}}\right| \cong \frac{3V_{DD} + 5\left|V_{tp}\right| - 3V_{tn}}{8}$$

where NM_L and NM_H represents the minimum and maximum allowable amount of noise to the input of a gate. Any voltage noise above this can change the logic state of the inverter. If either NM_L or NM_H for a gate are reduced to ~ $0.1V_{DD}$, then the gate may be susceptible to switching noise present at the inputs. Modeling the MOSFETs [96] without gate leakage current, the critical voltages are determined to be: $V_{OH} = V_{DD}$, $V_{OL} = 0$.

Hence

$$V_{IH} = \frac{(V_{DD} + V_{TP})(1 + 2\sqrt{1+3K}) + 3KV_{TN}}{1 + 3K + 2\sqrt{1+3K}}$$

$$V_{IL} = \frac{3(V_{DD} + V_{TP}) + 3KV_{TN}(1 + 2\sqrt{1+3/K})}{K + 3 + 2\sqrt{1+3/K}}$$

where V_{DD}, V_{TP} and V_{TN} are power supply voltage, and threshold voltages of PMOS and NMOSFETs respectively. K is the ratio of the transconductance of the NMOS to the PMOSFET.

When the circuit is subject to gate leakage current, there is a decrease in the noise margins of the inverter. If the input resistance of the FETs remain much greater than the output resistance, then the new critical voltages could be defined as

$$V_{OH} = V_{DD} - \frac{I_L(V_{DD} + V_{TP})}{2I_{PMOS|0}} \text{ and } V_{OL} = V_{DD} - \frac{I_L(V_{DD} - V_{TN})}{2I_{NMOS|V_{DD}}}$$

This implies that

$$V_{IH} = \frac{(V_{DD} + V_{TP})(1 + 2\sqrt{1+3K}\left\{1 + \frac{3K}{1+3K}\frac{I_L}{2I_{PMOS|V_{IH}}}\right\}) + 3KV_{TN}}{1 + 3K + 2\sqrt{1+3K}\left\{1 + \frac{3K}{1+3K}\frac{I_L}{2I_{PMOS|V_{IH}}}\right\}}$$

$$V_{IL} = \frac{3(V_{dd} + V_{TP}) + 3KV_{TN}(1 + 2\sqrt{1+3/K})\left\{1 + \frac{3}{3+K}\frac{I_L}{2I_{NMOS|V_{IL}}}\right\}}{K + 3 + 2\sqrt{1+3/K}\left\{1 + \frac{3}{3+K}\frac{I_L}{2I_{NMOS|V_{IL}}}\right\}}$$

where I_{PMOS} and I_{NMOS} are the drain-to-source currents through the PMOS and NMOS transistors at the given input voltages.

The degradation in the critical voltages and noise margins of the inverters as a function of leakage current I_L is shown for a symmetrical inverter in fig 36. The leakage current have been normalized to I_E which is the current through either of the transistors in inverter at the point at which V_{out} is equal to $V_{DD}/2$.

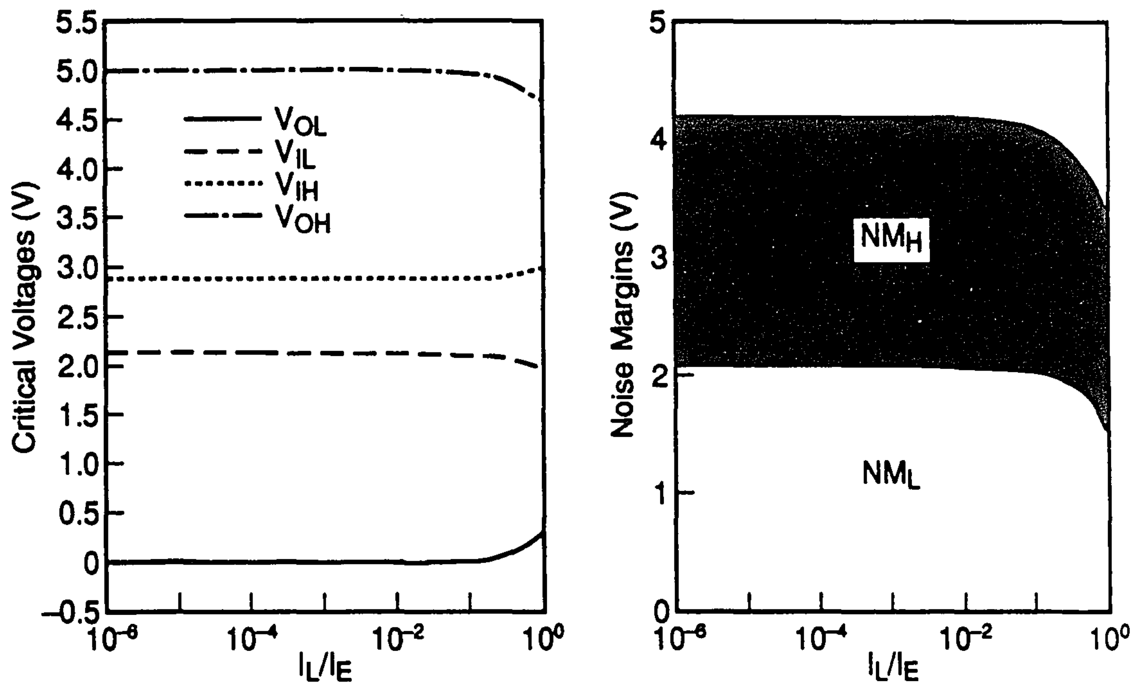


Figure 36: Dependence of Noise Margin on Leakage Current I_L/I_E. [62]

2.14.2 GATE OXIDE TUNNELING

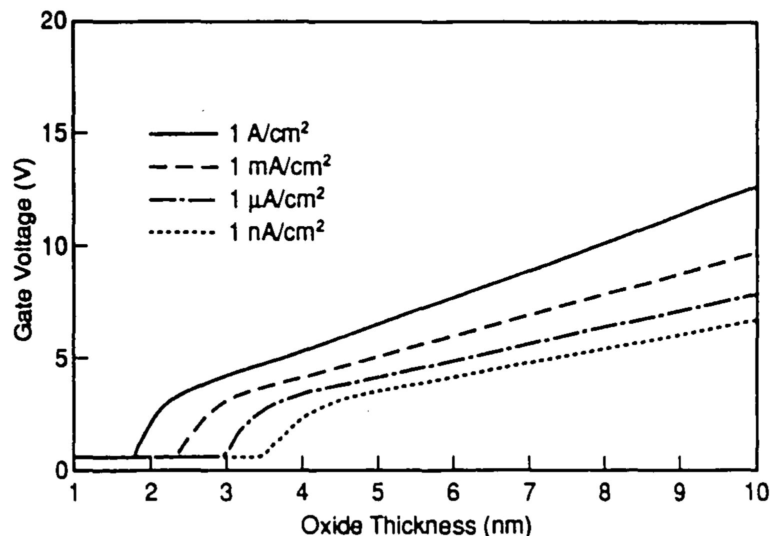


Figure 37: Gate voltage for given current densities and oxide thicknesses. [62]

The most predominant mechanism of MOS leakage current associated with thin dielectrics is tunneling through the gate dielectric .Direct tunneling current limits operation for both the *ON* and *OFF* regions of the MOS transistor .

Direct tunneling and Fowler-Nordheim tunneling have been examined by numerous authors [97-100]. Maserjain [97] has examined the tunneling current through thin MOS structures for different oxide thickness and bias voltages. For voltage across the gate oxide less than the barrier height for the injected electrode, direct tunneling is the dominant mechanism of conduction in the thin oxides. The barrier height for a silicon electrode injecting electrons into the conduction band of the gate oxide is 2.90eV[98]. Hence the current density for direct tunneling is

$$J = B(V_G)T_{ox}^{-2}e^{-2\bar{k}(V_G)T_{ox}}$$

where V_g is the gate voltage and $k(V_g)$ is the average value of the attenuation of the electron waves in the oxide[97] and $B(V_g)$ models the effect of the indirect bandgap of silicon.

When the barrier height is above 3.4eV the dominant current conduction mechanism is Fowler-Nordheim tunneling, whose current density could be approximated as given in previous equations. Hence from the above figure one can deduce the maximum power supply voltage by the oxide thickness and the maximum permissible leakage current through the gate dielectric. The fig. 38 below shows the impact of gate leakage current or direct tunneling current on the transfer characteristics of a CMOS inverter and also on the switching speed.

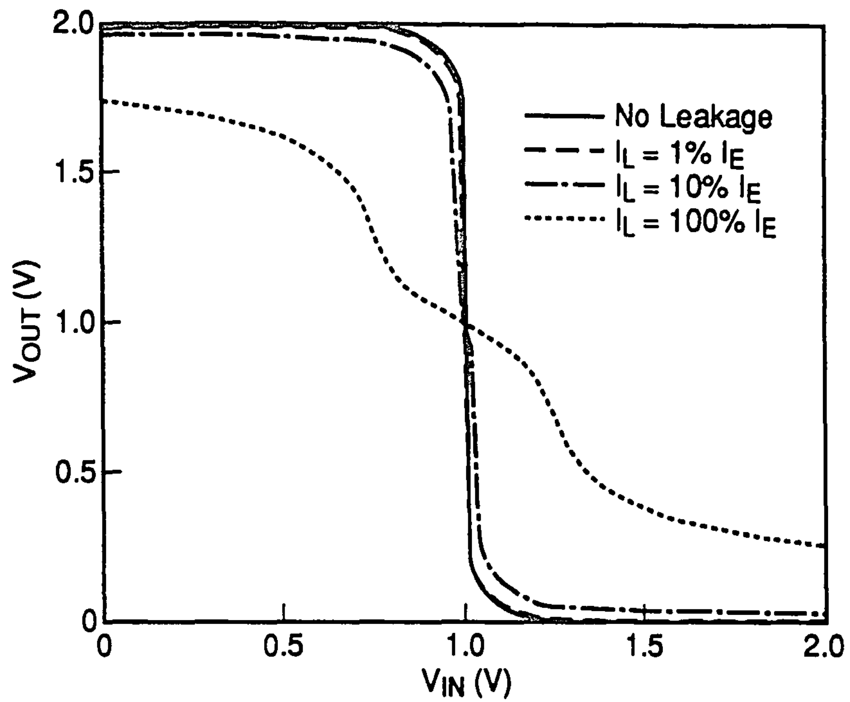


Figure 38a: Effect of leakage current on the voltage-transfer characteristics of an unloaded inverter. [62]

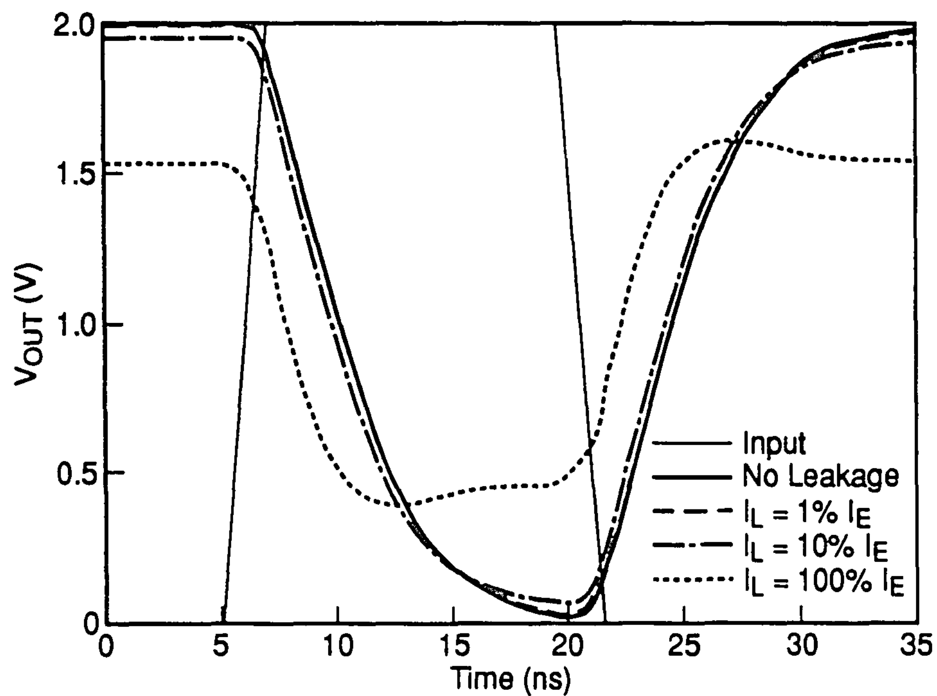


Figure 38b: Simulated effect of leakage current on the switching speed of an inverter. The gate has a fan-out of 3 and each gate is driving a 1 pF line capacitance. [62]

2.15 MOS CAPACITANCE EFFECTS ON SWITCHING CHARACTERISTICS

The effect of thin oxides on device performance manifests itself as an increase in the device capacitance as the oxide thickness is reduced. In recent generations, device scaling of MOSFETS has been accomplished without scaling the supply voltage, and thus the drive capabilities and parasitic capacitances of MOSFETS were satisfactorily scaled. However in 0.5um generation and below the V_{DD} must be scaled in accordance with the requirements of MOSFET reliability and the total power consumption which is quite severe in ASICs especially microprocessors with large cache memories.

2.15.1 DEVICE CAPACITANCE

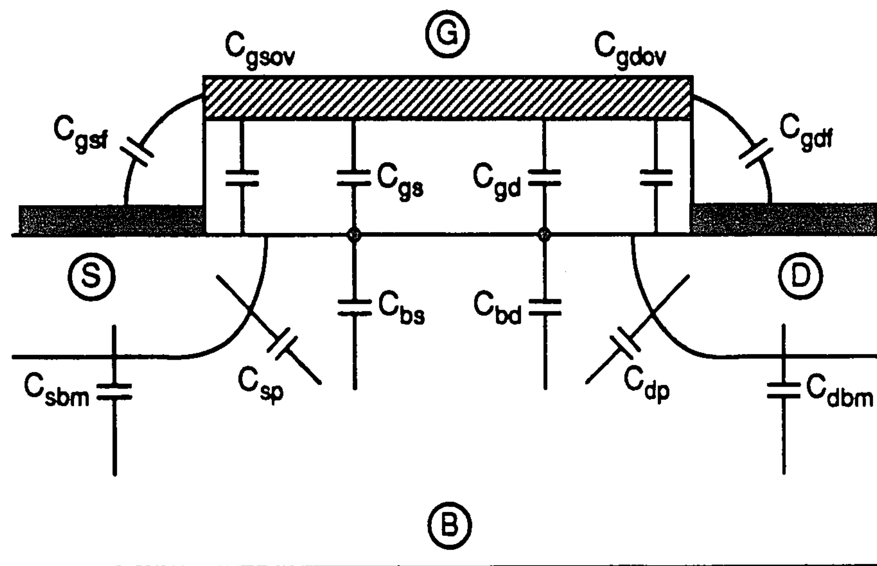


Figure 39: Definitions of device and parasitic capacitances.

The performance of MOS circuits is limited by the various capacitances associated with the interconnect wiring and the devices themselves. An MOS device in a given operation mode manifest itself as a capacitive load to its driver. Recalling from previous MOS treatment

$$Q_n = W\int_0^L Q'_n(y)dy$$

where Q'_n is the charge density in the channel.

A. LINEAR REGION

In this region of operation

$$Q'_n = -C'_{ox}(V_{GS} - V_{TS}) = -C'_{ox}X_S \text{ and } Q'_n = -C'_{ox}(V_{GD} - V_{TD}) = -C'_{ox}X_D$$

$$\text{hence } Q_n = -\frac{2}{3}C'_{ox}WL\frac{X_S{}^2 + X_S X_D + X_D{}^2}{X_S + X_D}$$

since $Q_n = Q_n(V_S, V_D)$, there are two capacitors associated with the device with C_S accounting for the charge due to V_S and C_D accounting for the charge due to V_D.

$$C_S = C_{gs} + C_{bs}$$

$$\Rightarrow C_S \cong \frac{\partial Q_n}{\partial V_S} = \frac{\partial Q_n}{\partial X_S}\frac{dX_S}{dV_S} = -\frac{\partial Q_G}{\partial V_S} - \frac{\partial Q_B}{\partial V_S}$$

differentiating 1 with respect to x yields

$$\frac{\partial Q_n}{\partial X_S} = -\frac{2}{3}C'_{ox}WL\frac{X_S{}^2 + 2X_S X_D}{(X_S + X_D)^2} \text{ but } \frac{dX_S}{dV_S} = -\left(1 + \frac{C'_S}{C'_{ox}}\right)$$

$$\Rightarrow C_S = \frac{2}{3}C'_{ox}WL\left(1 + \frac{C'_S}{C'_{ox}}\right)\frac{X_S{}^2 + 2X_S X_D}{(X_S + X_D)^2}$$

hence

$$\boxed{C_{gs} \cong \frac{2}{3}C'_{ox}WL\frac{X_S{}^2 + 2X_S X_D}{(X_S + X_D)^2} = \frac{2}{3}C_{ox}\frac{1 + 2\alpha}{(1 + \alpha)^2}} \quad (36)$$

and

$$\boxed{C_{bs} \cong \frac{2}{3}C'_S WL\frac{X_S{}^2 + 2X_S X_D}{(X_S + X_D)^2} = \frac{2}{3}\delta_1 C_{ox}\frac{1 + 2\alpha}{(1 + \alpha)^2} = \delta_1 C_{gs}} \quad (37)$$

similarly at the drain end:

$$C_D = C_{GD} + C_{BD} = \frac{\partial Q_n}{\partial V_D} = -\frac{\partial Q_G}{\partial V_D} - \frac{\partial Q_B}{\partial V_D}$$

hence

$$\boxed{C_{gd} \cong \frac{2}{3} C'_{ox} WL \frac{X_D{}^2 + 2X_S X_D}{(X_S + X_D)^2} = \frac{2}{3} C_{ox} \frac{\alpha^2 + 2\alpha}{(1+\alpha)^2}} \tag{38}$$

$$\boxed{C_{bd} \cong \frac{2}{3} C'_S WL \frac{X_D{}^2 + 2X_S X_D}{(X_S + X_D)^2} = \frac{2}{3} \delta_1 C_{ox} \frac{\alpha^2 + 2\alpha}{(1+\alpha)^2} = \delta_1 C_{gd}} \tag{39}$$

where

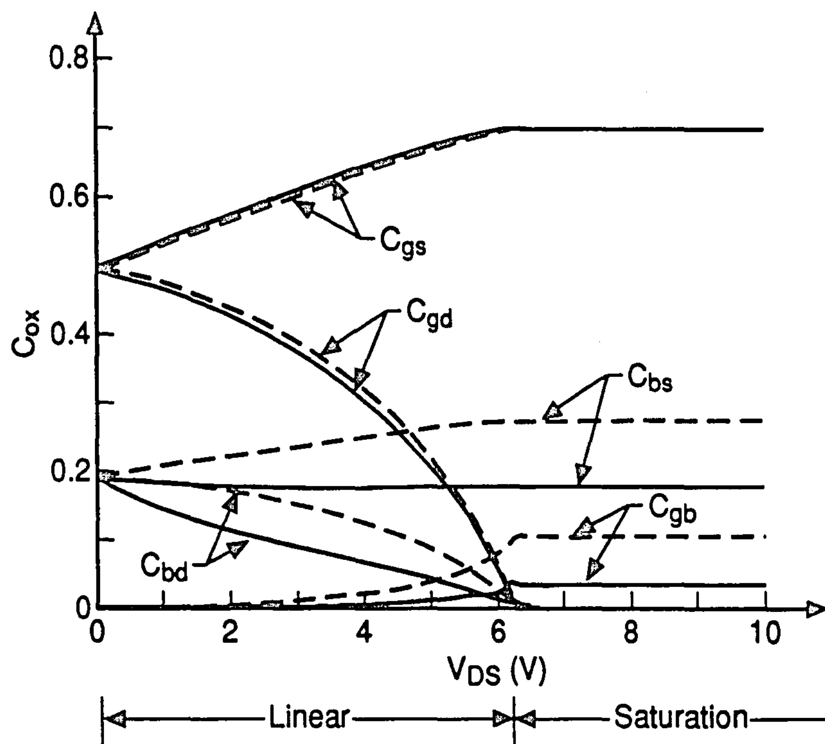
$$C_{ox} = C'_{ox} WL \,,\; \delta_1 = \frac{\lambda}{2\sqrt{(\phi_B + V_{SB})}} = \frac{dV_T}{dV_{SB}} \,, \alpha = \begin{cases} 1 + \delta_1 - V_{gs} + V_T \rightarrow \textit{linear} \\ 0 \rightarrow \textit{saturation} \end{cases}$$ and

$$C'_S WL = \delta_1 C_{ox}$$

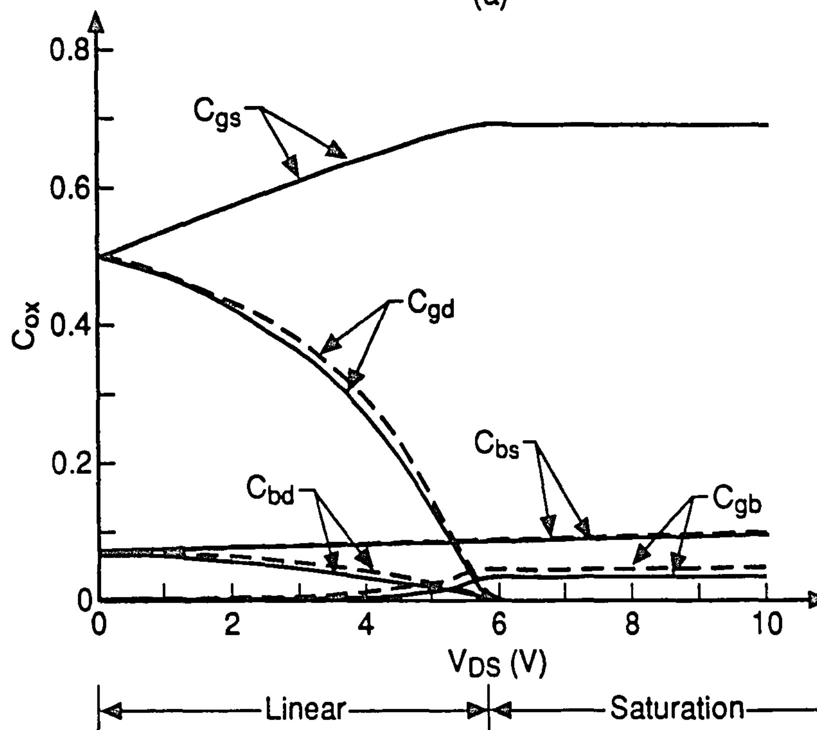
B. SATURATION REGION

In the saturation region for the first order approximation, I_{ds} is independent of V_{ds} and C_D hence X_D=0 which implies that

$$C_S = C_{gs} + C_{bs} \text{ and } C_{gs} = \frac{2}{3} C_{ox} \,,\; C_{bs} = \frac{2}{3} \delta_1 C_{ox}$$



(a)



(b)

Figure 40: Small-signal capacitances vs V_{DS} for a MOSFET in linear and saturation regions: (a) $V_{SB} = 0$; (b) $V_{SB} = 5$ V. [104]

From the above graph it can be seen that at $V_D = V_S = 0$ and $C_{gs} = C_{gd}$. As V_{ds} is increased, C_{gs} increases and C_{gd} decreases. As device enters into saturation C_{gd} decreases to zero. Note that the total device capacitance associated with the gate and channel $C_g = C_S + C_D$ and one can see that from the above figure that C_{gs} dominates when device is in saturation.

2.15.2 PARASITIC CAPACITANCE

In addition to the previously discussed device capacitances are the "parasitic capacitance" surrounding the device. As shown in the fig. 39 , C_{sbm} extends from the bottom of the source diffusion to substrate and C_{sp} extends from the perimeter of the source diffusion to the channel area and the substrate. C_{gsov} due to the overlap between the gate and the source as a result of LDD under-diffusion and spacer oxide overlap. And also due to the finite thickness of the gate material C_{gf} fringes from edges of the gate to adjacent conductors. All these capacitances that can affect device speed in MOS circuits are given as [101]:

$$\boxed{C_{bm} = \frac{C_{bm0}}{(\phi_B + V_R)^{\frac{1}{2}}}} \tag{40}$$

$$\boxed{C_p = \frac{C_{p0}}{(\phi_B + V_R)^{\frac{1}{2}}}} \tag{41}$$

using Schwartz-Chrstoffe transformation [102,103]

$$\boxed{C_{gf} = \gamma_c \varepsilon_{ox} W_G} \tag{42}$$

where $\gamma_c = \dfrac{\left[2 - \ln 4 + \ln a + \ln\left(\dfrac{u}{a}\right)\right]}{\pi}$ with $a = 2K\left(K^2 - 1\right)^{\frac{1}{2}} + 2K^2 - 1$ and

$K = 1 + \dfrac{T_G}{T_{ox}}$ with T_G=gate thickness and u determined trancedentally.

Furthermore the gate overlap capacitance is given as

$$\boxed{C_{ov} = \frac{\varepsilon_{ox} \Delta L_G W_G}{T_{ox}}} \tag{43}$$

2.16 INVERTER DELAY CHARACTERIZATION

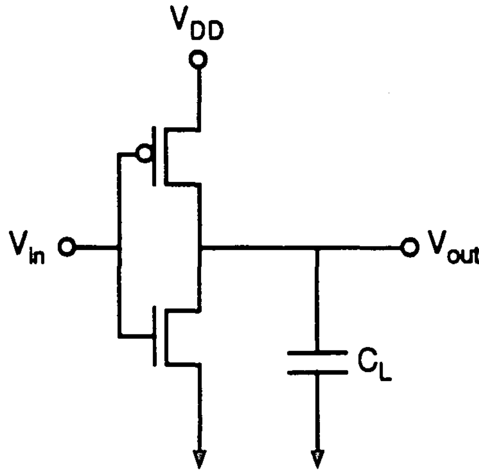


Figure 41: CMOS inverter circuit

The device capacitances shown in fig. 39 are logically divided into three categories namely: the input capacitance contributing to the gate input capacitance, C_{gb} and C_{gs} output capacitance contributing to the output capacitance C_{db} and the feedback capacitance C_{gd}. In this section, a newly developed versatile model will be presented that

describes the switching characteristics for a CMOS device with considerations for carrier velocity saturation effects under high field conditions which are not considered in the conventional models. The switching speed for a CMOS gate is limited by the time required to charge or discharge load capacitance C_L by the MOSFET drain current. The delay time consists of the following three terms:

Fall time τ_f = time for a waveform to fall 90%, to 10% of V_{DD}

Rise time τ_r = time for a waveform to rise 10%, to 90% of V_{DD}

2.16.1 FALL TIME DETERMINATION

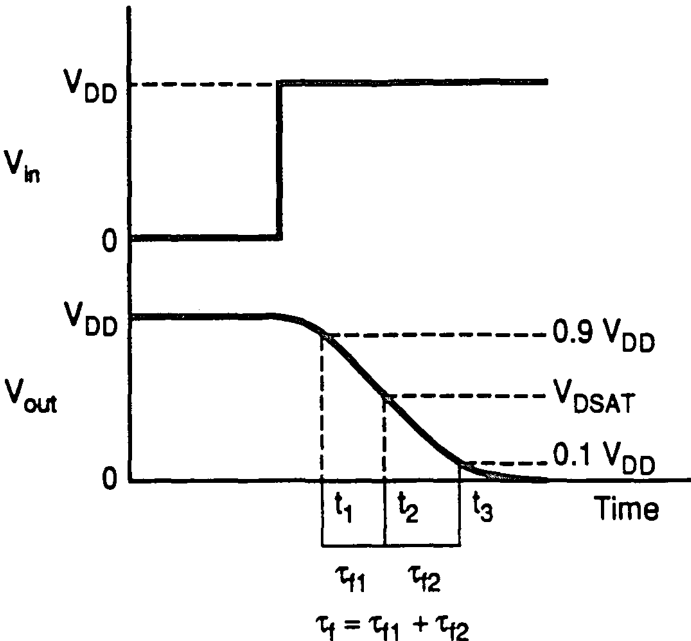


Figure 42: CMOS inverter switching characteristics.

Fig. 42 above shows the familiar CMOS inverter with capacitive load C_L which represents the capacitance of the next gate C_G and the output capacitance of the previous stage. Since the N-channel transistor pull-down action controls the fall-time, then it can be

shown that [104] when the input is driven by a step waveform as shown in fig. 42, the τ_f

V_{DD} operation consist of:

τ_{f1} - Period during which capacitor voltage V_{out} drops from $0.9V_{DD}$ to pinch-off voltage V_{dsat}

τ_{f2} - Period during which V_{out} drops from V_{dsat} to $0.1V_{DD}$.

The transfer characteristics could be expressed as

$$C_L \frac{dV_{out}}{dt} + I_{DS} = 0 \qquad \text{with } V_{out} \geq V_{DD} - V_{TN}$$

where I_{ds} is the drain current for the NMOSFET and C_L is the loading capacitance such as gate, junction and fringing capacitances. Accounting for high electric field effects, an analytical I_{ds} model[105] could be used

$$I_{ds} = \frac{W\mu C_{ox}\left(V_G' - \frac{V_{ds}}{2}\right)V_{ds}}{L\left(1 + \frac{V_{ds}}{E_C L}\right)} \quad \text{in linear region where } V_{ds} \leq V_{dsat}$$

$$I_{ds} = \frac{W\mu C_{ox}(V_G' - V_{dsat})E_C}{2} \quad \text{in saturation region where } V_{ds} > V_{dsat}$$

where $V_{dsat} = \frac{E_C L V_G'}{E_C L + V_G'}$ and integrating 1 from $V_0 = 0.9V_{DD}$ to $V_{out} = V_{dsat}$ and substituting into the above equation

$$\Rightarrow t_{f1} = t_2 - t_1 = C_L \int_{0.9V_{DD}}^{V_{dsat}} dV_{out} + \int_{0.9V_{DD}}^{V_{dsat}} I_{DS} dt = \frac{2C_L(0.9V_{DD} - V_{dsat})}{W\mu_n C_{ox}(V_G' - V_{dsat})}.$$

When the N-channel device begins to operate in the linear region, the discharge current is no longer constant. The time τ_{f2} taken to discharge the capacitor is obtained by substituting and integrating from V_{dsat} to $0.1V_{DD}$.

$$\Rightarrow t_{f2} = t_3 - t_2 = C_L \int_{0.1V_{DD}}^{V_{dsat}} dV_{out} + \int_{0.1V_{DD}}^{V_{dsat}} I_{DS} dt = \frac{C_L L}{W\mu C_{ox}}\left[\frac{1}{V_G'}\ln\left(\frac{V_{dsat}}{0.1V_{DD}}\right)A + \frac{2}{E_C L}\ln A\right]$$

where $A = \frac{2V_G' - 0.1V_{DD}}{2V_G' - V_{dsat}}$. The combination of τ_{f1} and τ_{f2} yields the complete fall time as a result of load C_L

$$t_f = \tau_{f1} + \tau_{f2} = \frac{C_L}{W\mu C_{ox}}\left[\frac{2(0.9V_{DD} - V_{dsat})}{(V_{Gn}' - V_{dsat})E_C} + \frac{1}{V_{Gn}'}\frac{V_{dsat}}{V_{DD}}A + \frac{2}{E_C L}\ln A\right] \quad (44)$$

2.16.2 RISE TIME DETERMINATION

Due to the symmetry of CMOS circuits a similar approach could be used to derive the rise time which is controlled by the P-channel device as

$$t_r = \frac{C_L}{W\mu C_{ox}}\left[\frac{2(0.9V_{DD} - V_{dsat})}{(V_{Gp}' - V_{dsat})E_C} + \frac{1}{V_{Gp}'}\frac{V_{dsat}}{V_{DD}}A + \frac{2}{E_C L}\ln A\right]. \quad (45)$$

Hence the average delay of an inverter stage which is dominated by the output rise and fall times could be approximated as:

$$\tau_{AV} = \tau_{PD} = \frac{t_r + t_f}{4} = kC_L V_{DD}\left(\frac{1}{W_N I_{dsn}} + \frac{1}{W_P I_{dsp}}\right) \quad (46)$$

where k is a constant, C_L is the total load capacitance, W is the transistor width and I_{ds} is current in MOSFET at $V_{gs} = V_{ds} = V_{DD}$. The load capacitance could then be expressed as

$$C_L = C_G x F.O + C_j + C_{Al}. \quad (47)$$

As shown by the derivations for τ_f, τ_r and τ_{PD}, the switching performance of a CMOS inverter has strong dependence on the loading capacitance C_L, which is mostly influenced by the gate capacitance C_G and the total number of fan-outs.

CHAPTER 3

EXPERIMENT

3.1 INTRODUCTION

This chapter describes my experimental procedure including device modeling, transistor fabrication, device aging and measurements. The primary goal of this thesis was to quantify the effects of thinner gate oxides on MOSFET switching performance. I therefore had to design experiments that will satisfy the modeled device characteristics.

In carrying out the experimental plan, at the N-tub step the threshold voltage had to be adjusted for the PMOS devices. This is because the blanket adjust used for the NMOS devices automatically alters the p-channel threshold. As a result the N-tub arsenic implant was used to correct for the V_T shift in the PMOS devices. This is because in the N-tub Arsenic engineering, the peak of the profile is placed just below the maximum depletion width edge. As a result the straggle of Arsenic profile can influence the surface charge enough to modulate the threshold voltage of the PMOSFET. This was done for all variations of the gate oxide thickness.

The gate oxidation process was performed in dry O_2 in presence of TCA with diluent N_2 to inhibit oxide growth during the ramp-up cycle to ensure thickness uniformity across the wafer. After the gate PBr_3 doping and patterning, the polysilicon gate was etched with reactive sputter etch plasma chemistries which had an enhanced ability of stopping on the thin gate oxide thus preventing gate oxide breakthrough into bulk silicon. The wafers were then recombined and processed with the flow yet to be described. The discussions in this chapter will encompass device simulation approach used, device fabrication, aging and I-V parametric measurements.

3.2 DEVICE SIMULATIONS

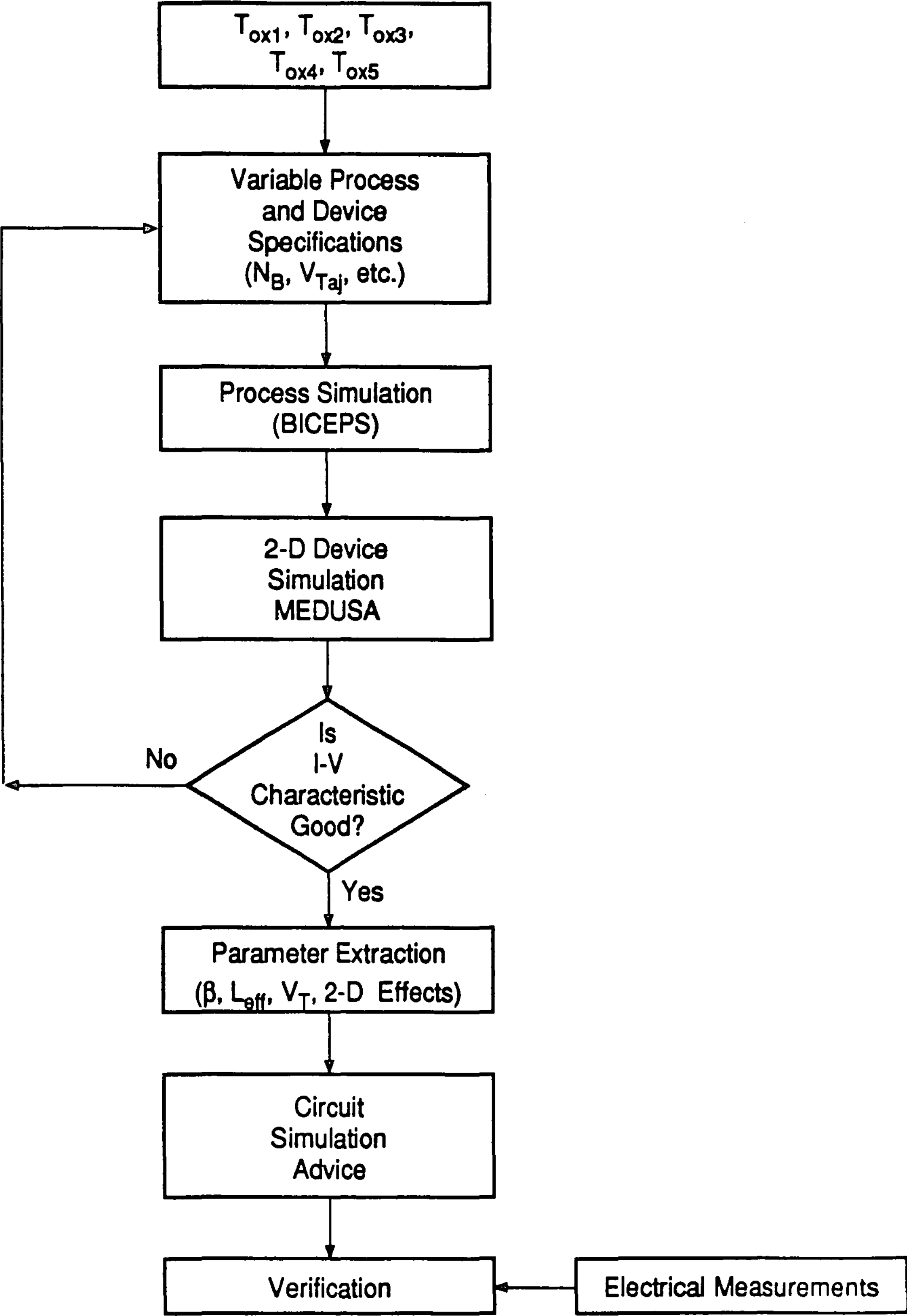


Figure 43: Device simulations flow diagram.

During the developmental stages of the experimental plan, I had to use the 2-D process simulator BICEPS to model the fabrication process the wafers were expected to see in the manufacturing line. The output of BICEPS which is the doping profiles for the devices was fed into the device simulator MEDUSA which was used to fully characterize the MOSFETS before actual processing. This device simulator takes into account all the 2-D devices encountered in the sub-micron regime such as short channel, narrow width, DIBL phenomena etc.

This highly computer intensive loop was carried out until the desired device characteristics were achieved for every variation of oxide thickness. As discussed earlier with the MOS device equations, thinning the gate oxide tends to cause a decrease in the threshold voltage of devices. To correct this the substrate doping must be increased to make it more difficult to turn the device on which results in an increase in the threshold voltage. But increasing the substrate doping also results in a decreased mobility and hence a decrease in device drive current. As a result, these must be carefully balanced for efficient drive capability of the transistors and also the ability to make one-to-one comparison of performance as a function of gate oxide thickness.

3.3 CMOS DEVICE FABRICATION SEQUENCE

3.3.1 NTUB FORMATION

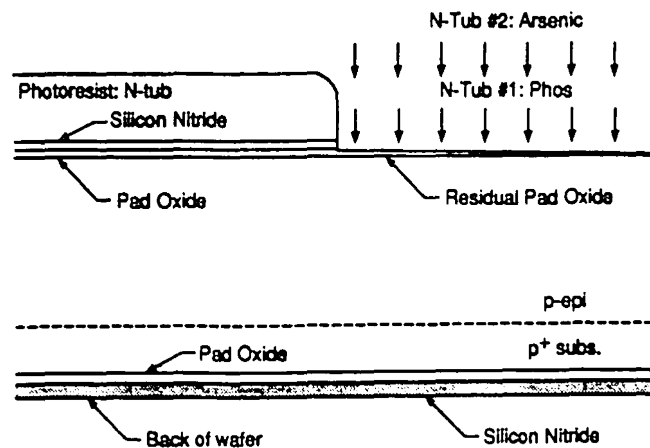


Figure 44: N-tub formation.

The starting material for the devices was lightly doped p-type epitaxial layer grown on p-type bulk silicon with <100> crystalline orientation which generates a minimum number of surface states. The N-tub processing begins with a thin pad oxidation, followed by Si_3N_4 deposition and spin on photoresist for tub patterning. After exposure, develop and bake the exposed the Si_3N_4 is etched off stopping on the pad oxide which serves as an implant mask as shown in fig. 44. The wafers are then implanted with phosphorus followed by arsenic whose peak depth is engineered so as to coincide with the depth of the LDD, thus serving as an effective punchthrough suppression implant for minimizing short channel $-V_{TP}$ roll-off in PMOS devices.

3.3.2 PTUB FORMATION

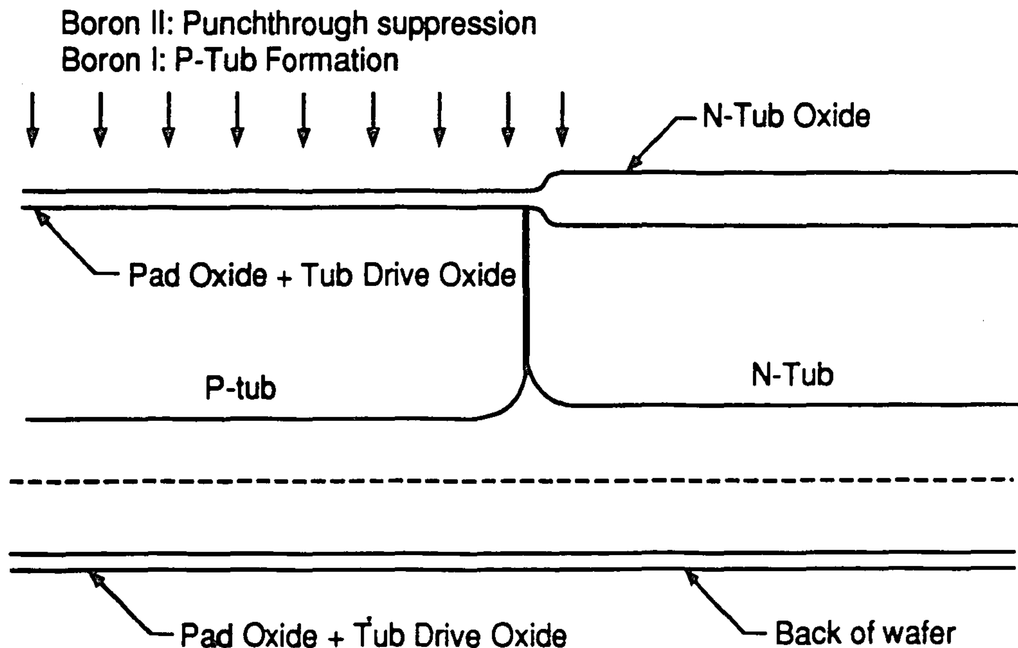


Figure 45: P-tub formation.

After N-tub implants the photoresist is stripped and a thick oxide is grown, which drives the N-tub into the silicon and set the tub junction. The Si_3N_4 is then stripped to

exposes the P-tub region which receives two boron implants for P-tub formation and NMOS punchthrough suppression as shown in fig. 45.

3.3.3 DEVICE ISOLATION

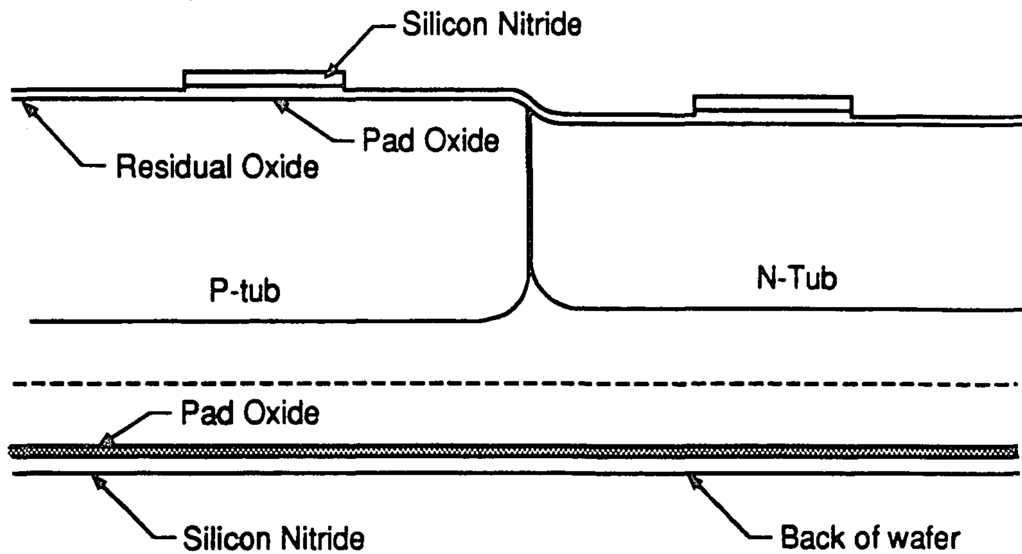


Figure 46: Active device area (Thinox) formation.

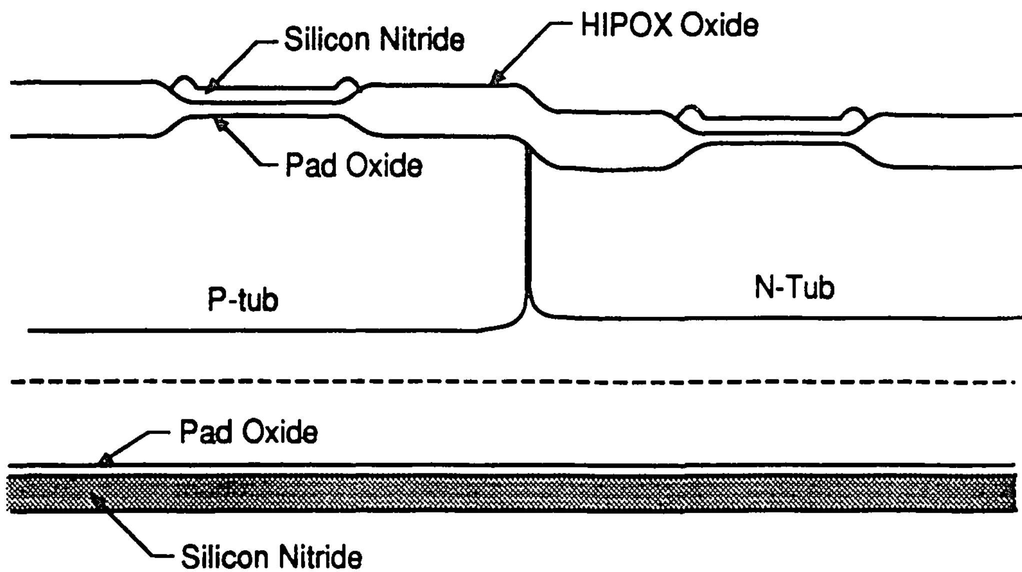


Figure 47: High pressure oxidation (HIPOX) for field oxide formation.

This sequence allows for the patterning and definition of the active regions of the transistors. Si_3N_4 is deposited and patterned to define the active regions and the field left exposed as shown in fig. 46. The field oxide is now grown as shown in fig. 47 using a high pressure oxidation scheme. This process is advantageous because of its reduced thermal budget which limits the oxidation-enhanced diffusion of dopants thus maintaining good isolation between two adjacent devices in the same tub.

3.3.4 THRESHOLD ADJUST IMPLANTS

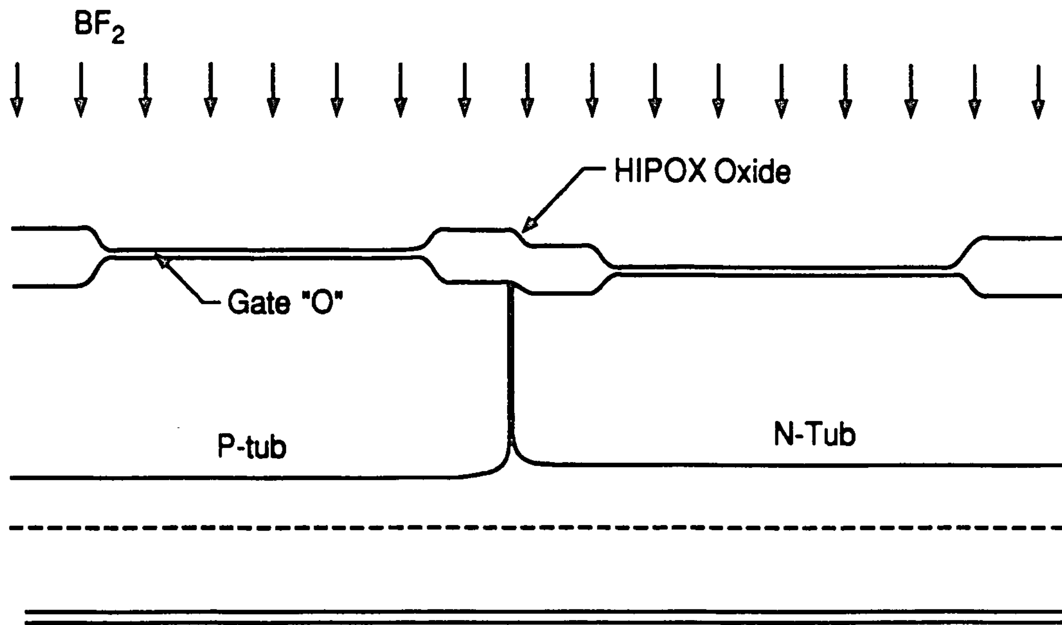


Figure 48: Threshold adjust ion implantation.

At this stage the Si_3N_4 had been stripped off and the transistor active region defined as shown in fig. 48. The pad oxide is stripped off and a sacrificial oxide of equal thickness as the actual gate oxide is grown. This is used as an implant screen to prevent implant ion channeling and silicon damage. A blanket BF_2 is implanted to set threshold voltage for the NMOS and PMOS devices. Five experimental cells were made each with

different BF_2 dose estimated from device simulations to accurately correct for the V_T shift due to the variation in gate oxide thickness.

3.3.5 GATE DEFINITION

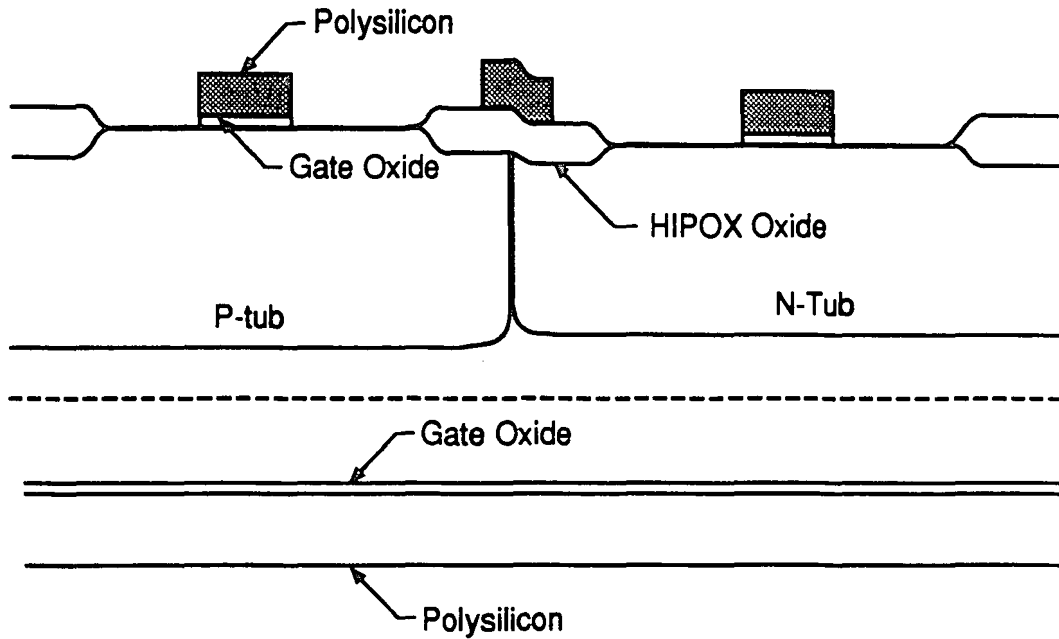


Figure 49: Polysilicon gate definition.

After the threshold adjust implant, the sacrificial gate oxide was etched and the true gate oxide was grown for the five cells of varying thickness, after which the polysilicon gate was deposited, PBr_3 doped and then patterned. This was then etched using dry etch chemistry to define the transistors and polysilicon interconnects.

3.3.6 LDD AND SPACER FORMATION

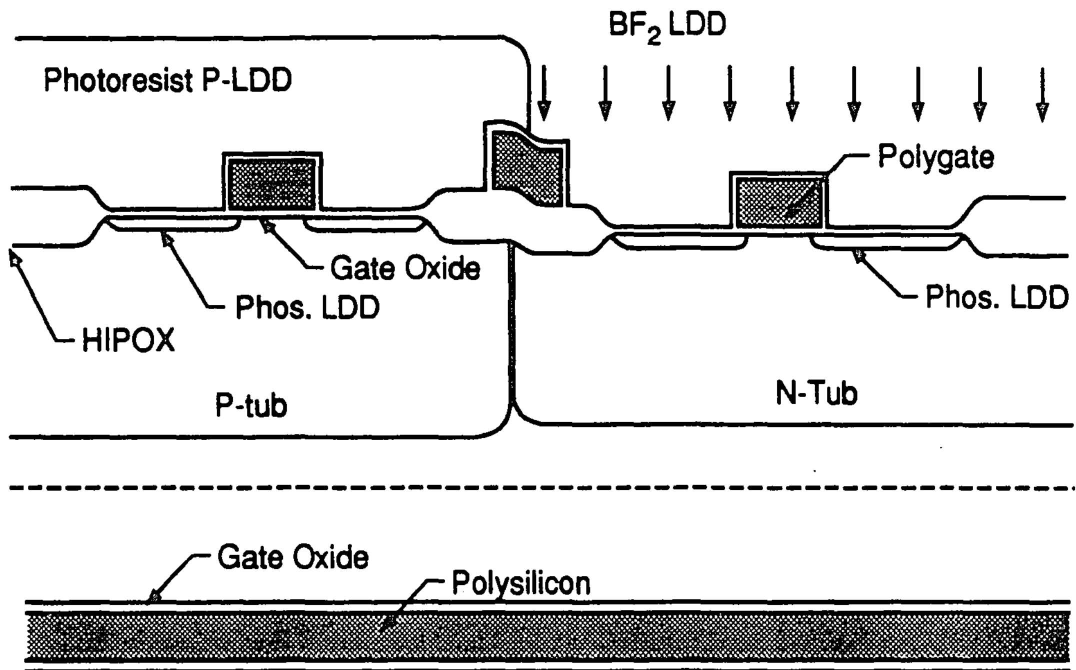


Figure 50: LDD formation.

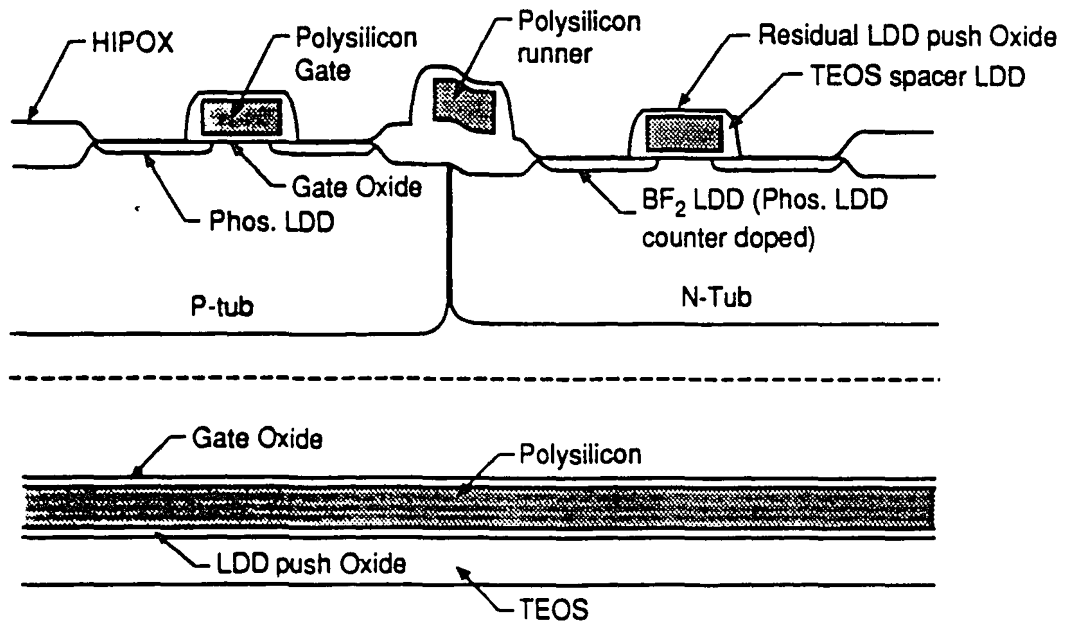


Figure 51: Spacer formation.

The LDD structures are formed by a blanket phosphorus implant which dopes both the n and p channel S/D regions. This is followed by a thermal cycle to set the NLDD junction and activate dopants. After this the n-channel regions are masked as shown in fig. 50 and the PLDD BF_2 implant done which is of high enough dose to counter-dope the phosphorus.

Following the PLDD implants comes the deposition of TEOS oxide which is etched anisothropically to form the spacers which protects the LDD regions from the S/D implants as shown in fig. 51.

3.3.7 SELECTIVE N^+ AND P^+ S/D FORMATION

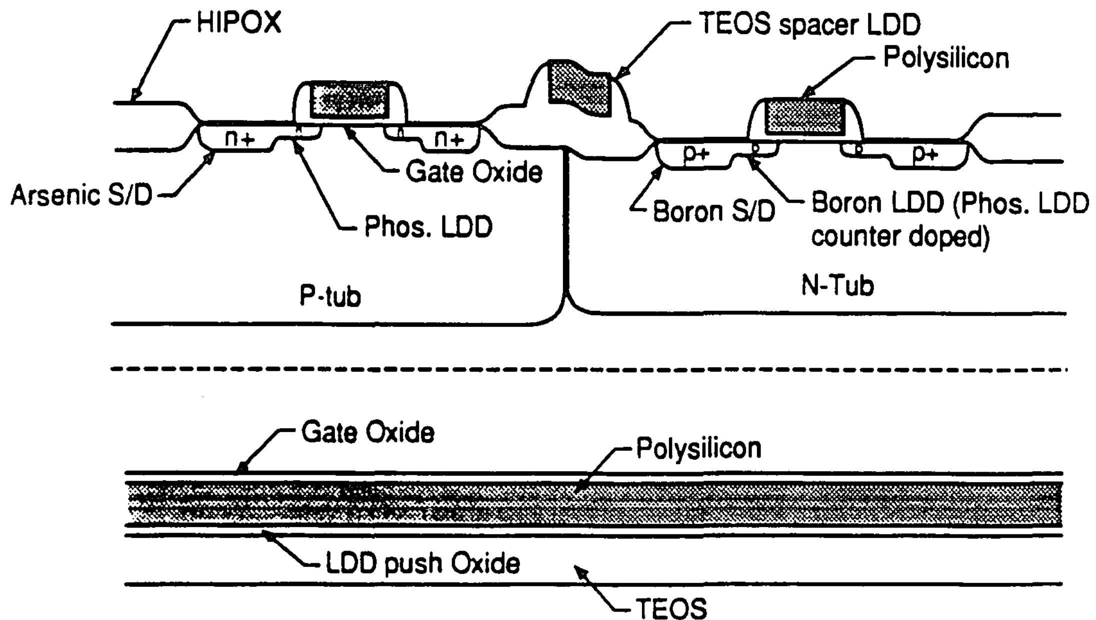


Figure 52: Selective N+ and P+ S/D ion implantation.

Photoresist is used to selectively mask the n and p-channel regions for Arsenic and BF_2 S/D implants respectively. This is followed by a high temperature anneal to set the S/D junctions, effective channel length and also to activate the dopants.

3.3.8 SELF-ALIGNED SILICIDATION

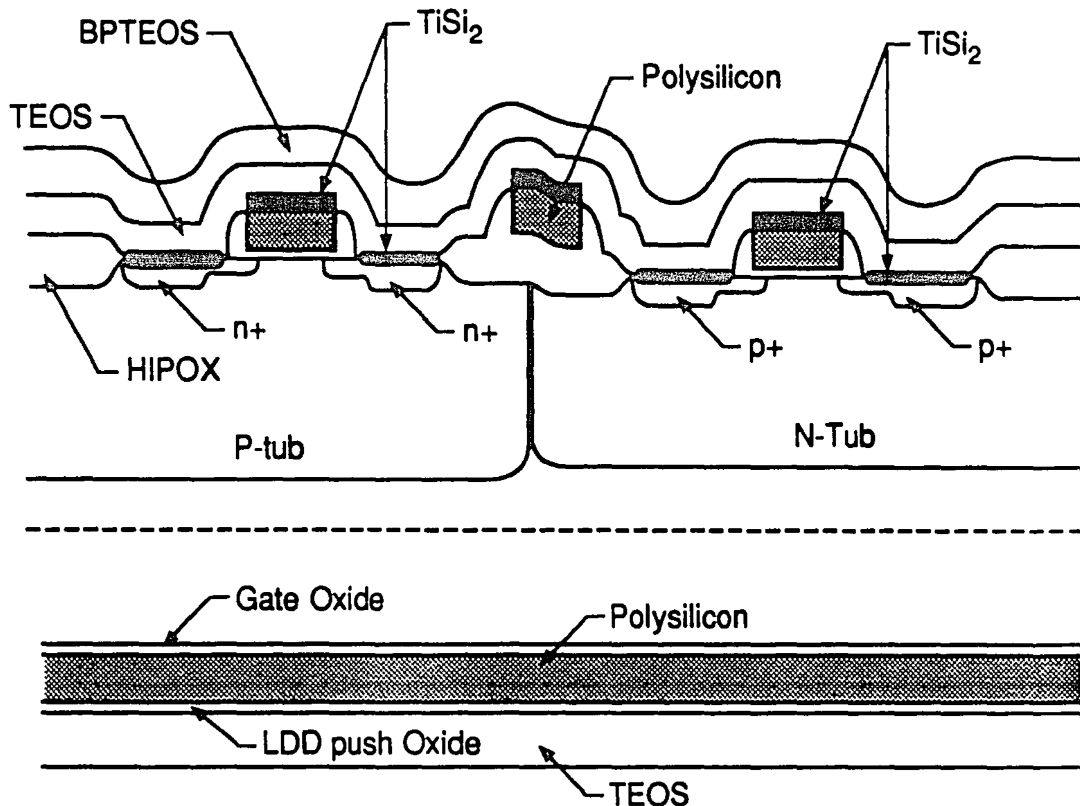


Figure 53: Self aligned selective silicidation.

$TiSi_2$ films are now formed on the S/D and gates only, by depositing titanium and subjecting it to an RTA heat treatment to form a C-49 phase $TiSi_x$ film. A selective chemical etch is then used to remove the unreacted Ti/TiN films. A second RTA is then performed to transform the C-49 to C-54 phase stable $TiSi_2$ film. This is followed by the deposition and flow of undoped TEOS and doped BPTEOS inter-level dielectric between the polysilicon gate and aluminum I levels as shown in fig. 53.

3.3.9 CONTACT, METALLIZATION AND PASSIVATION

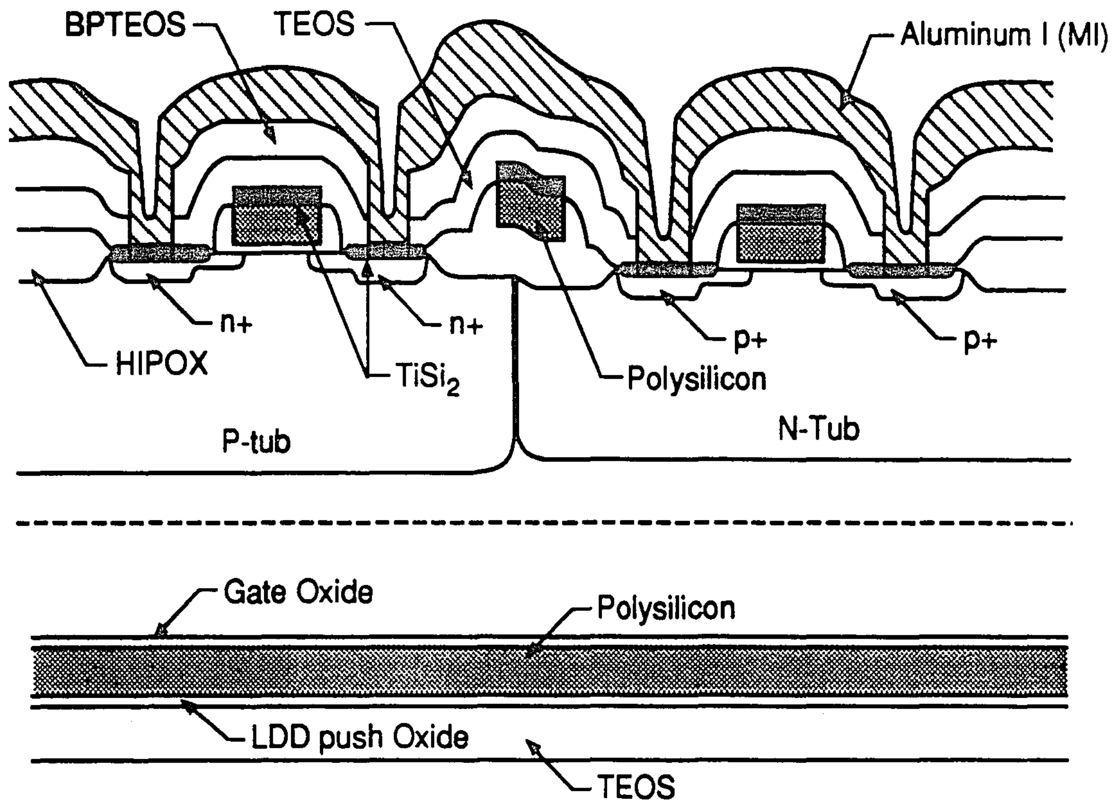


Figure 54: Dielectric I, contact opening and Aluminum I metallization.

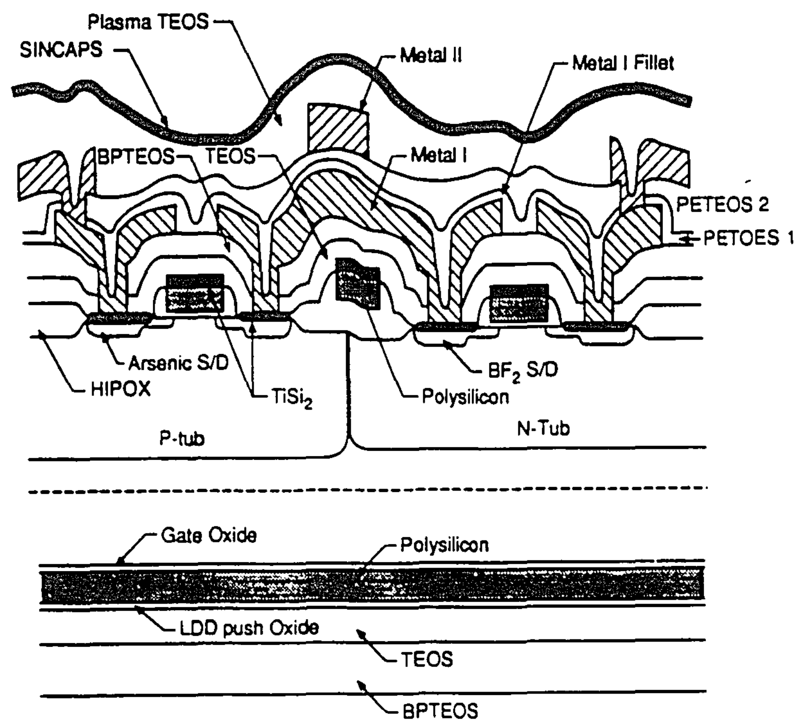


Figure 55: Dielectric II, Window II, Aluminum II and Passivation CAP.

The transistors are connected to bond pads with a sandwich of aluminum and Titanium Nitride interconnect. The contact to the transistor terminals are made by cutting windows in the inter-level dielectric I, after which titanium is deposited and given an RTA treatment in a nitrogen ambient to form the titanium nitride. Aluminum is then deposited, patterned and etched as shown in fig. 54. This process sequence is repeated with the deposition of inter-level dielectric II which is used to isolate the first and second level metallization. Window II's are then cut in the oxide and aluminum II is deposited, patterned and etched as shown in fig. 55. After this an oxide-Si_3N_4 sandwich is deposited, patterned and etched resulting in a passivation layer, which is followed by a low temperature sintering cycle. This is followed by back grinding, after which the wafers were ready for device characterization.

3.4 TEST STRUCTURES

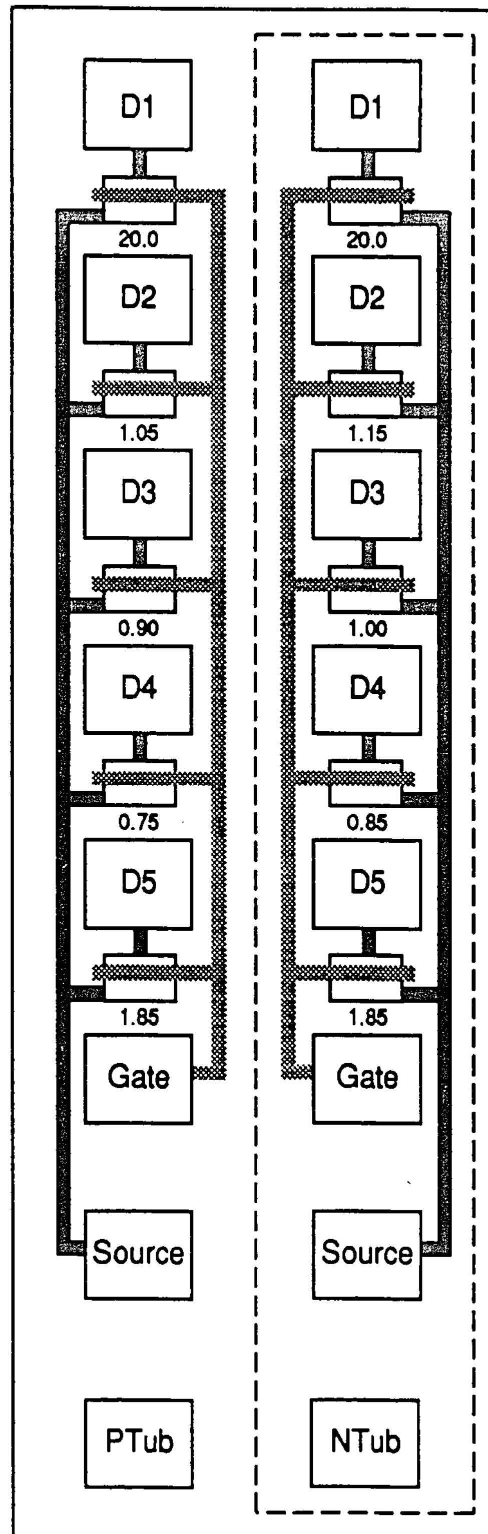


Figure 56: Test structure for I-V and aging characterization.

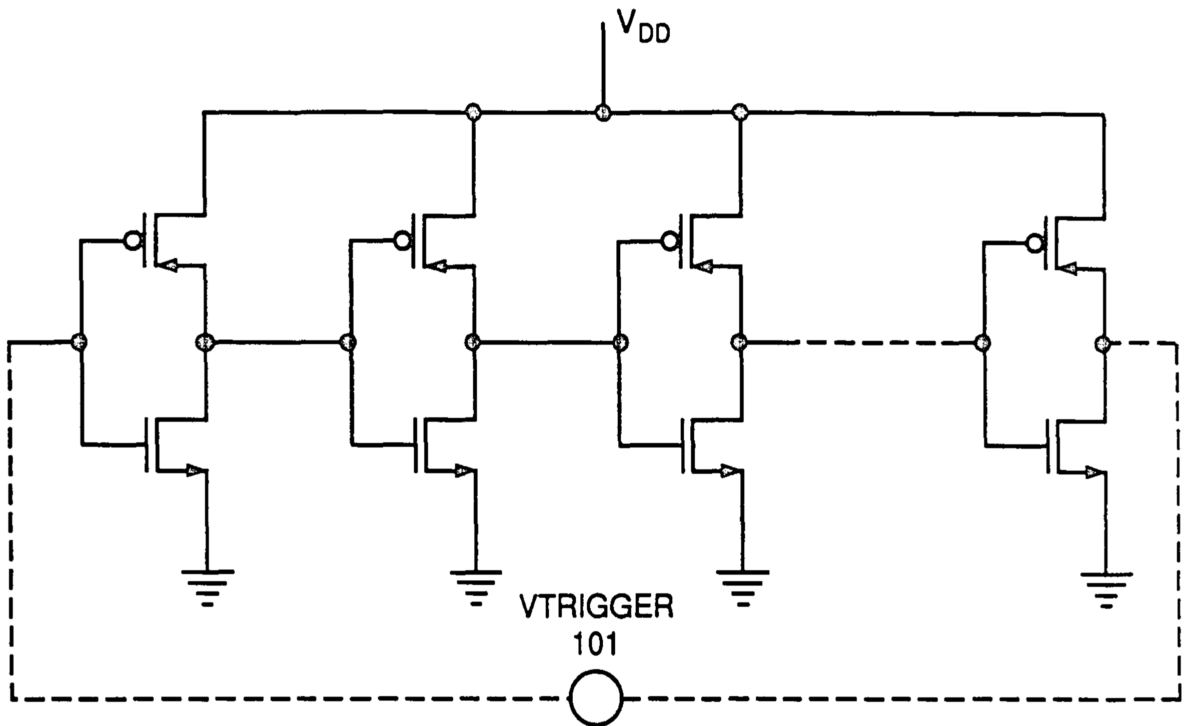


Figure 57a: Schematic diagram of 101 cascaded inverter chain.

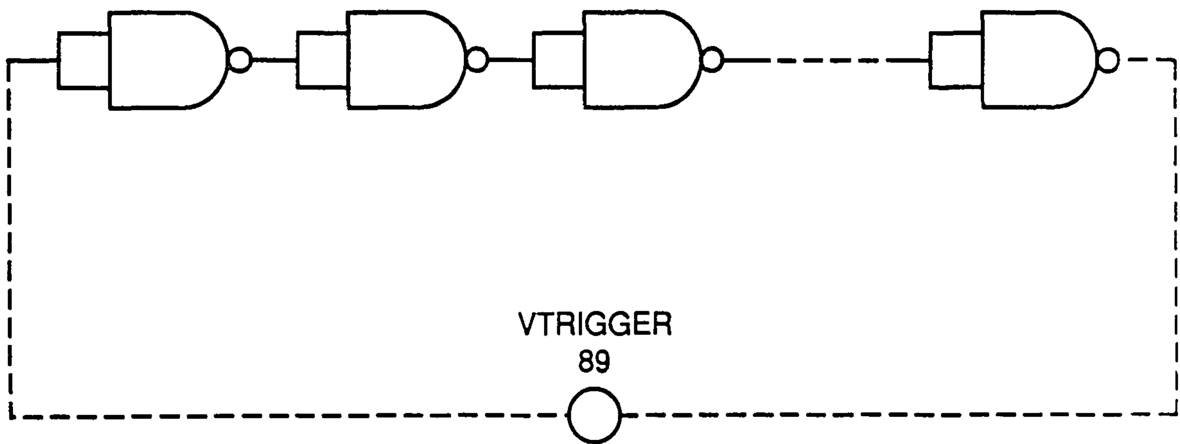


Figure 57b: Schematic diagram of 89 cascaded NAND chain.

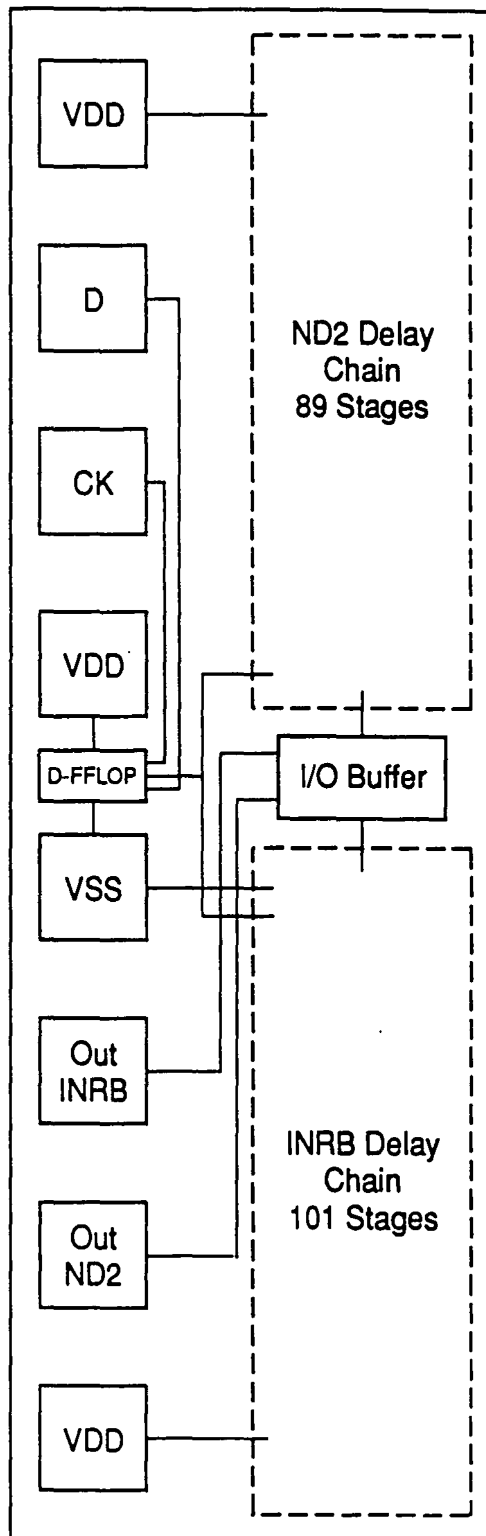


Figure 58: Test structure for performance characterization.

The devices used for transistor and performance characterization were fabricated with the process flow previously described. The transistor tester had NMOFETs of coded lengths 0.75um, 0.9um, 1.05um and 20um, and PMOSFETs of 0.85um, 1.0um, 1.15um and 20um gate lengths. As shown in the diagram above fig. 56, the transistors had a common source, gate and substrate nodes with individual drain contacts. As already discussed the effects of gate oxide on device parameters and hot carrier degradation were characterized using this test structure. Fig. 57 shows the schematic capture of the delay chains for both the INRB and ND2 with no external loading effects.

The fig. 58 contains a) the CMOS inverter delay chain with 101 stages and no external capacitive load, and b) the nand gate delay chain of 89 stages. These were both used in the quantification of the effect of thin gate oxides on switching delays.

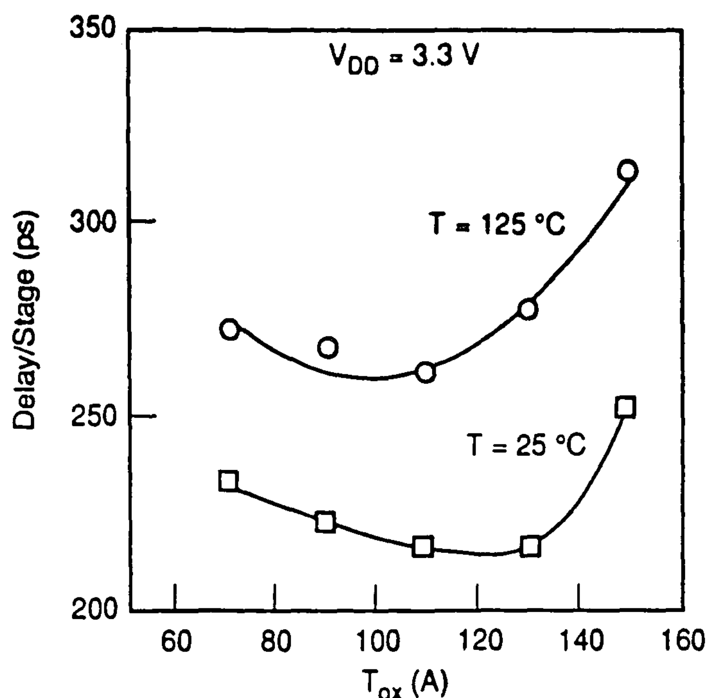
The electrical test equipment's used for the device characterization consisted of an HP model 4145A Semiconductor Parametric Analyzer, a probe station equipped with thermal chuck and an HP7470A plotter. For the hot carrier characterization similar equipment's were used, but one that had the capability of computer interface to run the "aging program" and report to the plotter the changes in device characteristics.

CHAPTER 4

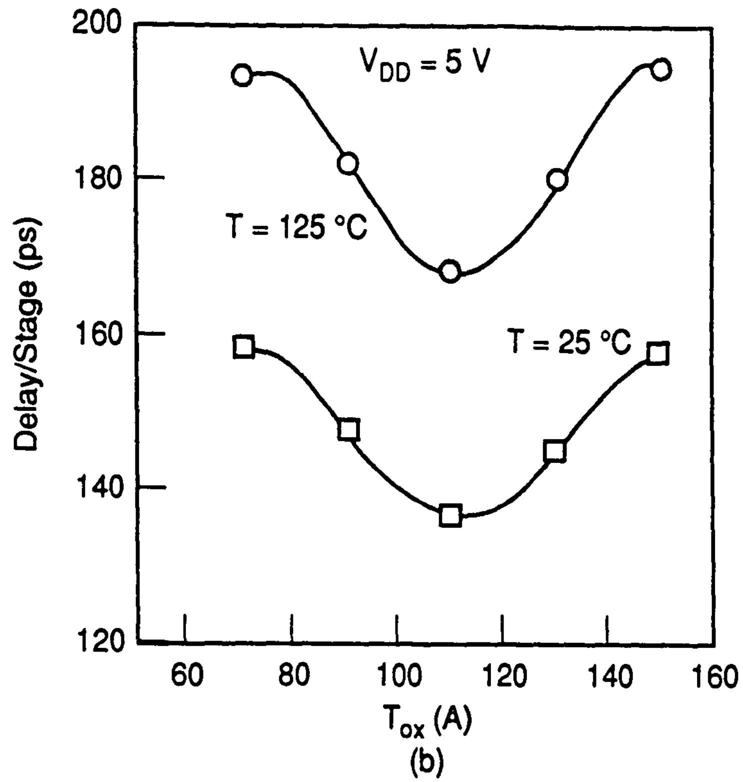
RESULTS AND CONCLUSIONS

4.1 DELAY CHARACTERIZATION

As shown in figure 59 and 60, the propagation delay was quantified for two different ring-oscillator delay chains, namely INRB and ND2 as described in the experiment section which was made of 101 and 89 stages respectively. Fig. 59a shows the INRB delay/stage at 3.3V operation as a function of gate oxide thickness and varying operating temperature. The delay/stage with a fan-out load of 1 is larger at the higher operating temperature, this is a manifestation of degraded device characteristics at those temperatures. However a characteristic "bell " shape dependence of the delay/stage on gate oxide thickness is present at both low and high temperatures. Fig. 59b shows the same result at the operating bias of 5V, the shape is still evident even with the lower τ_{PD} .

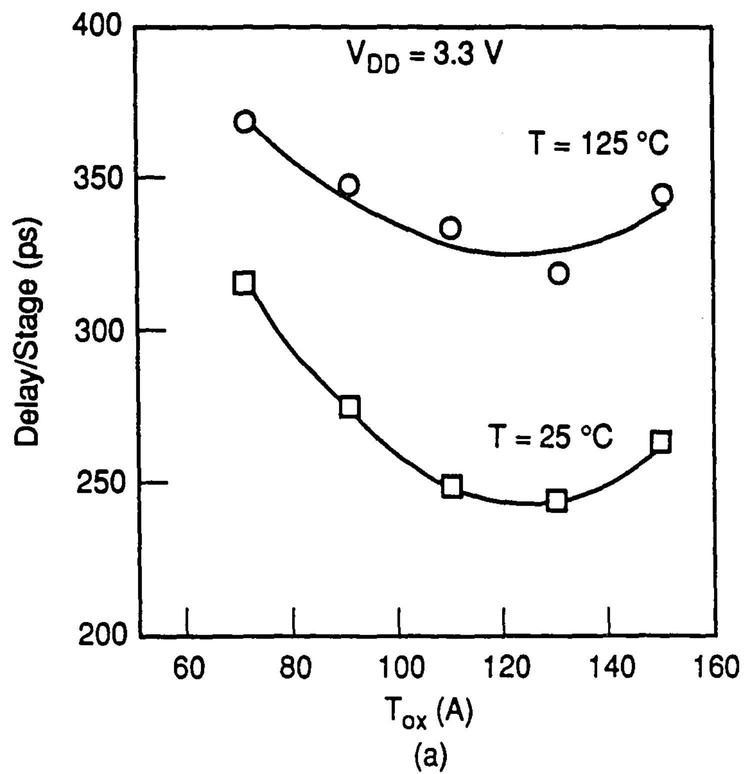


(a)



(b)

Figure 59: INRB delay/stage as a function of gate oxide thickness.



(a)

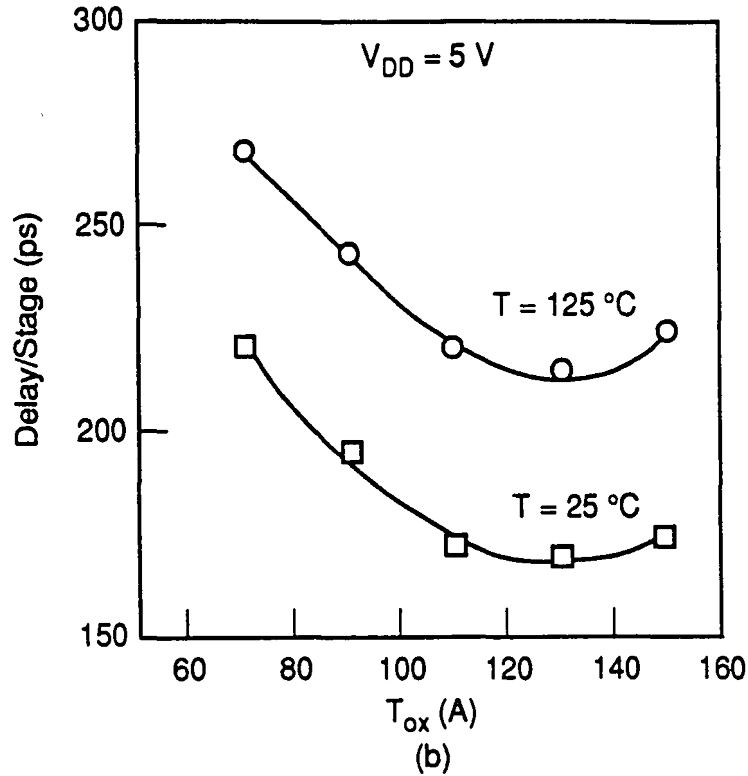


Figure 60: ND2 delay/stage as a function of gate oxide thickness.

Figure 60a and 60b shows a similar data for the ND2 delay chains even though the absolute delay is higher due to the larger load of the NAND. This convincingly defines a "minima" in gate oxide thickness necessary to achieve optimum circuit performance. The above experimental results can be explained by the equations developed in the theory section. Recalling from the theory section, τ_{PD} could be compactly expressed as

$$\tau_{PD} = kC_L V_{DD}\left(\frac{1}{W_{effn} I_{DSn}} + \frac{1}{W_{effp} I_{DSp}}\right)$$

where $$C_L = C_G x F.O + C_j + C_{AL},$$

from these simplified equations one can see why the propagation delay τ_{PD} is dependent on V_{DD} as shown in figs. 59 and 60, and also dependent on C_L which is mostly influenced by the gate capacitance C_G for a given fan-out. But C_G is a complex function of C_{ox} which increases with decreasing T_{ox} as developed in other equations in the theory section. This

unequivocally shows why after a certain "minima" in gate oxide thickness the propagation delay/stage begins to increase with decreasing gate oxide thickness. Another contributing factor also is the decrease in mobility as T_{ox} decreases which coupled with the velocity saturation due to the high critical electric field decreases after a certain gate oxide "minima".

As a result, to enhance device performance during scaling, the optimum T_{ox} must be chosen so as to minimize τ_{PD} for a given gate length. Also since the fan-out (F.O) is typically a design parameter the C_j and C_{Al} must also be reduced to enhance performance. This is a challenge because with scaling dimensions, higher substrate impurity concentration is required to prevent short-channel effects and also sufficient chanstop boron is required to maintain adequate n^+/p^+ isolation. However the higher substrate concentration also causes an increase in C_j, especially when using twin tub processing. It has been reported elsewhere that performing the chanstop implant after the isolation cycle serves to reduce C_j due to the small amount of boron lost from the substrate as a result of redistribution[105]. This technique is also effective in reducing the narrow width effects since the diffusion of the chanstop boron into the active device area is limited.

In other to minimize C_{Al}, multilevel metallization could be employed to effectively decrease interconnect parasitics due to the reduction of total wiring length thus reducing substrate-to-metal capacitance.

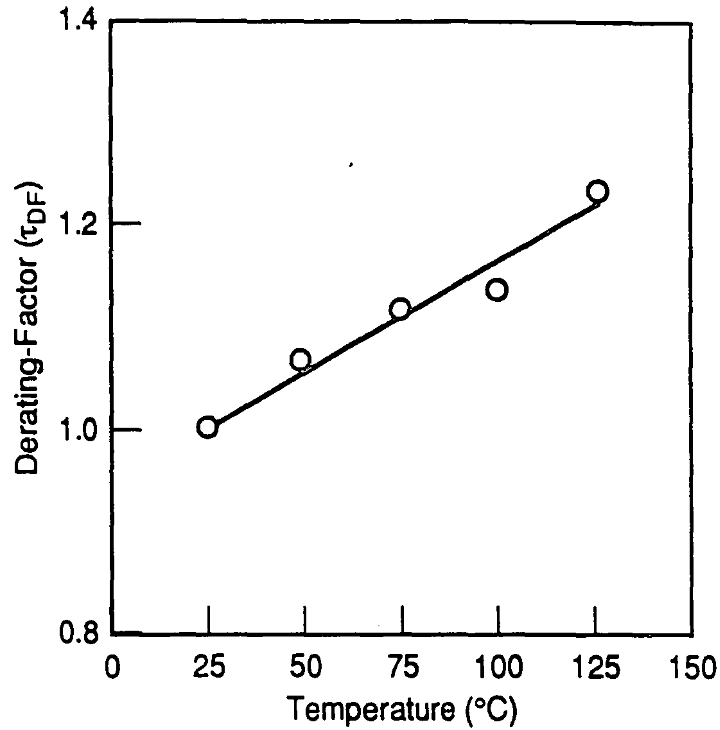
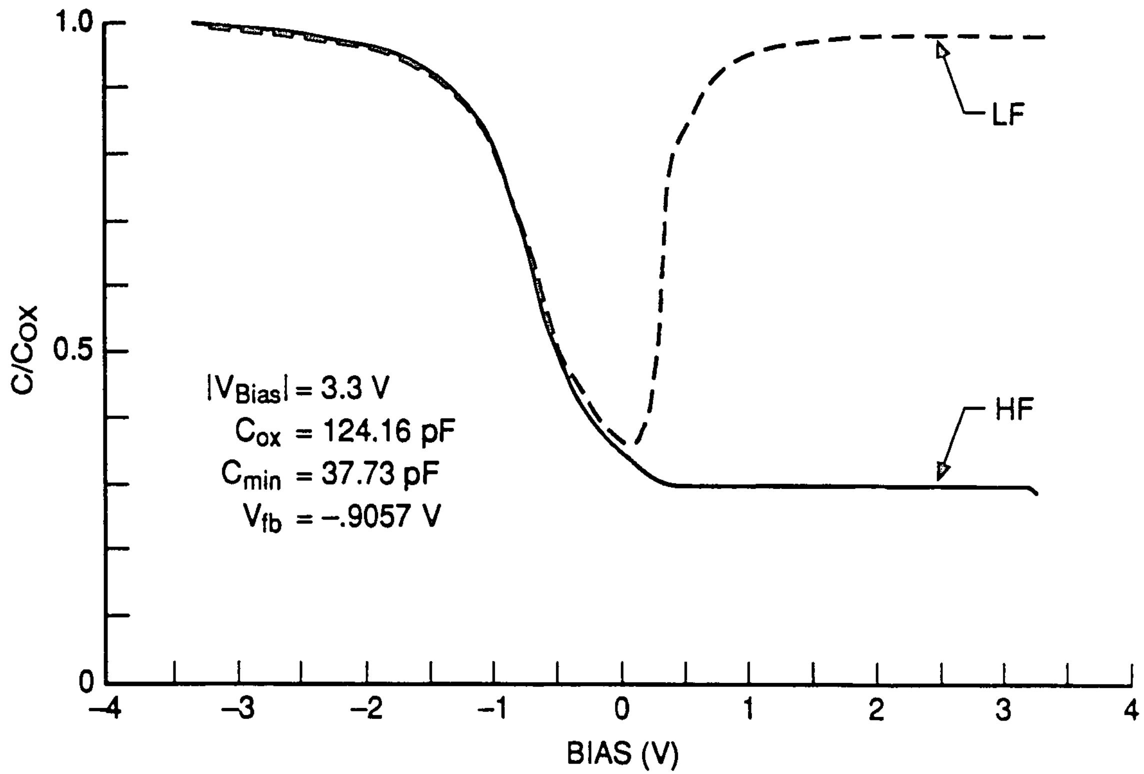


Figure 61: Inverter derating factor as a function of operating temperature.

Figure 61 shows the derating factor(τ_{DF}) for the INRB ring oscillator as a function of operating temperature. An increase of about 24% was evident as the operating temperature increased from 25°C to 125°C. It was also determined experimentally that the τ_{DF} was independent of operating supply voltages of 3.3V and 5V.

4.2 ELECTRICAL CHARACTERIZATION



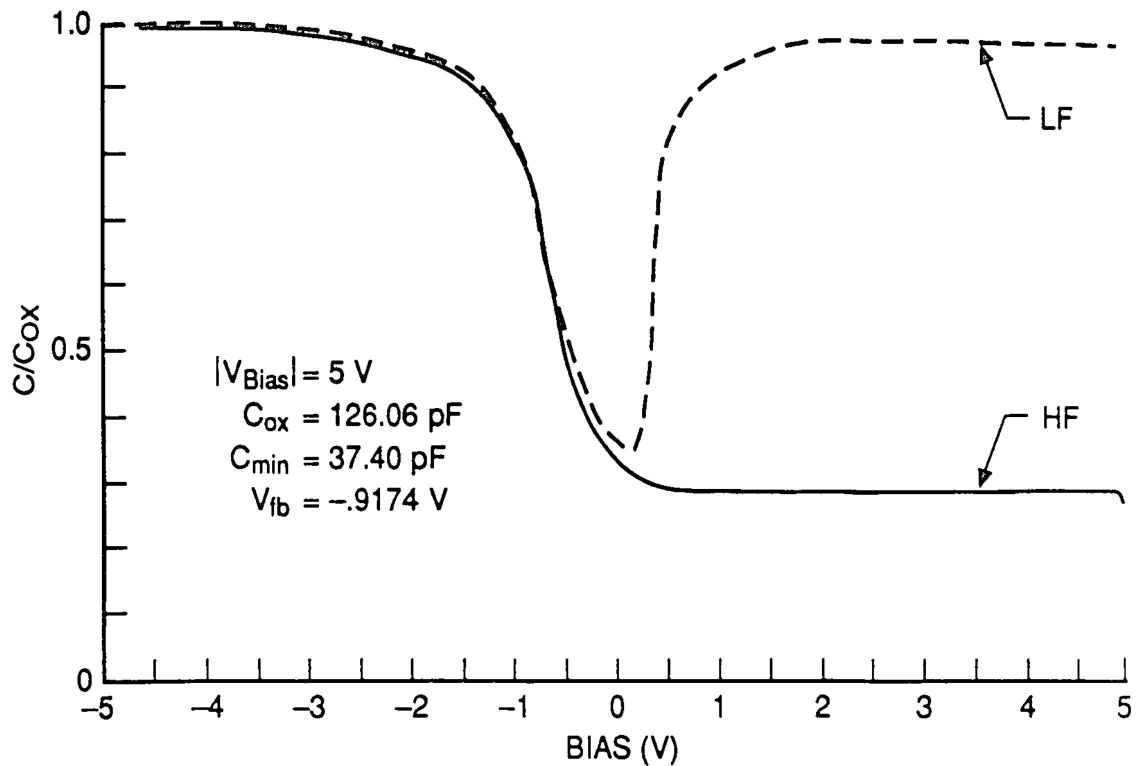
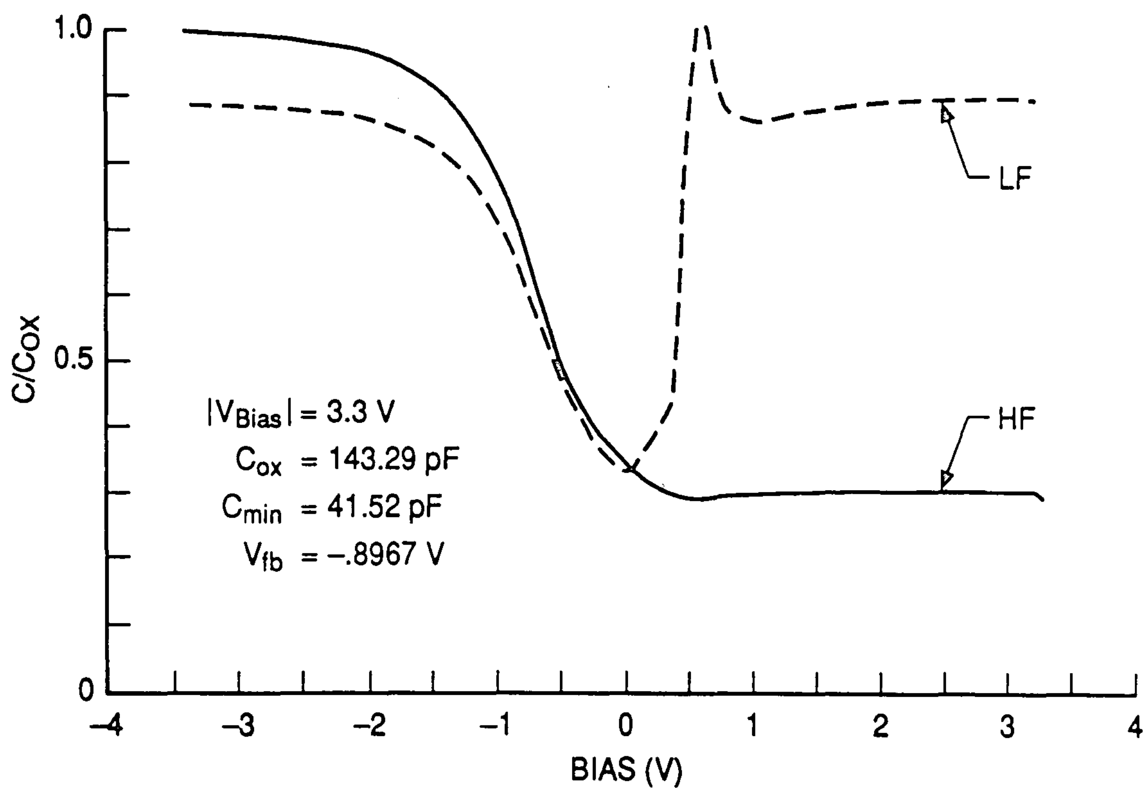


Figure 62: C-V characteristics for 150Å gate oxide.



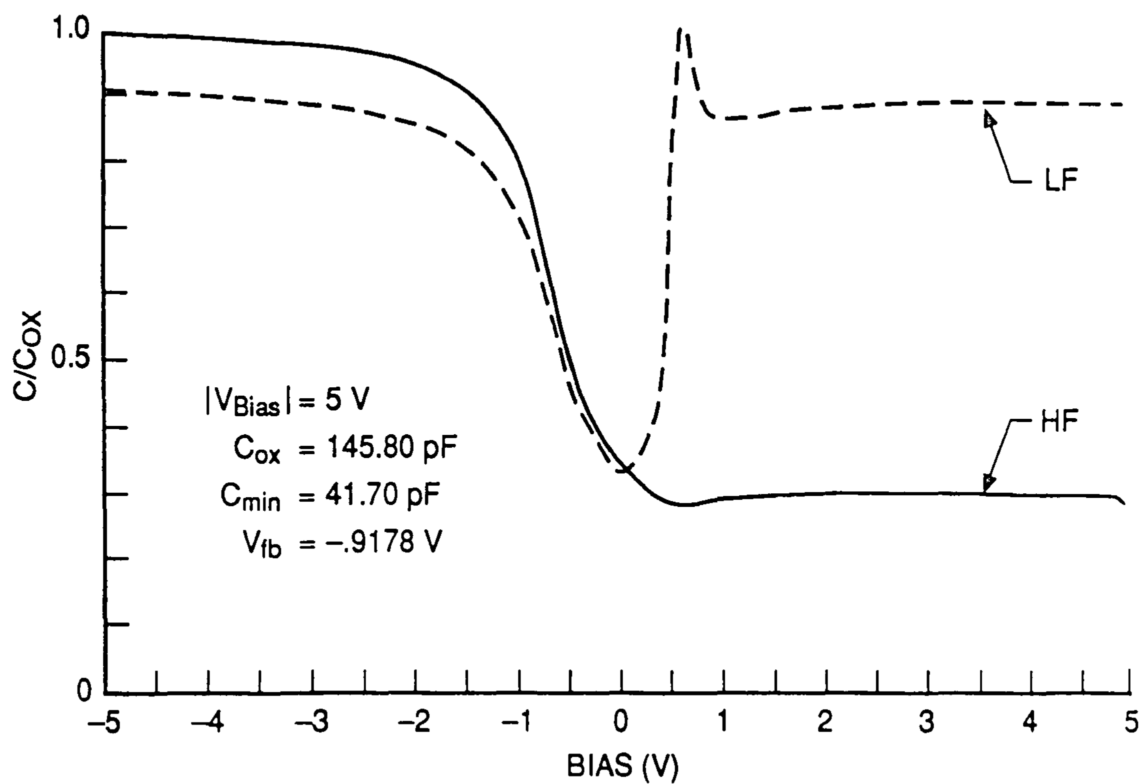
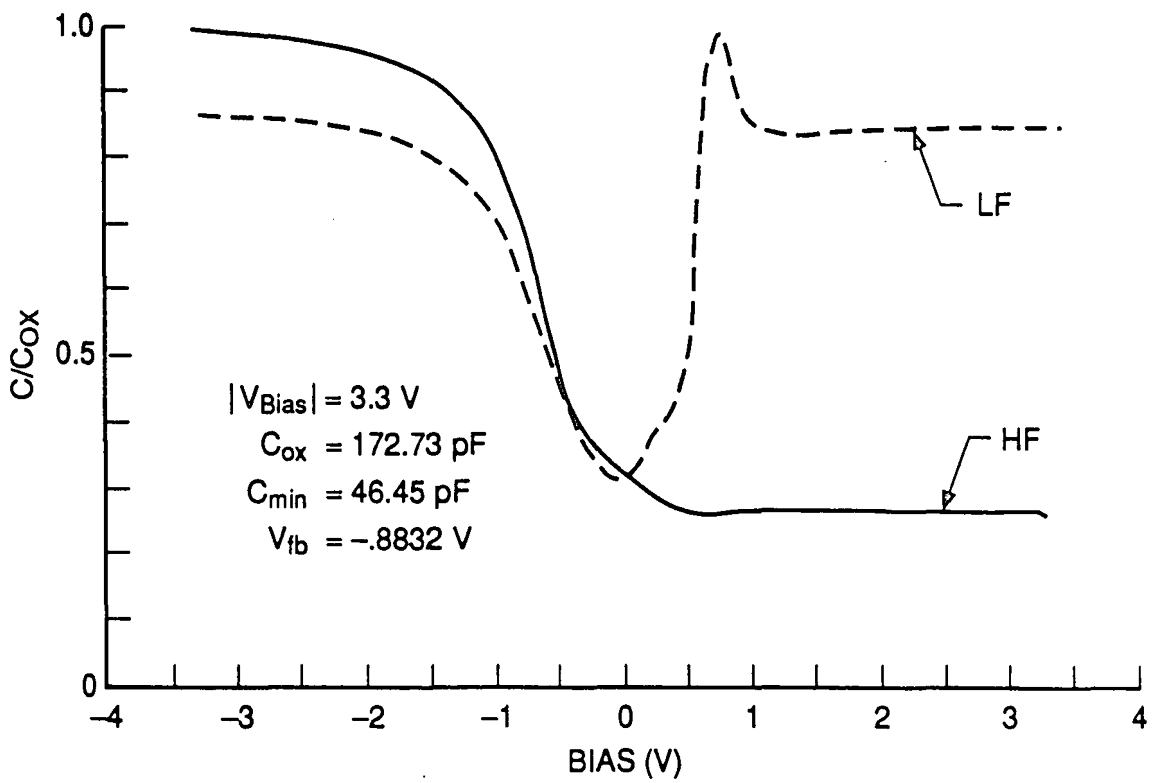


Figure 63: C-V characteristics for 130Å gate oxide.



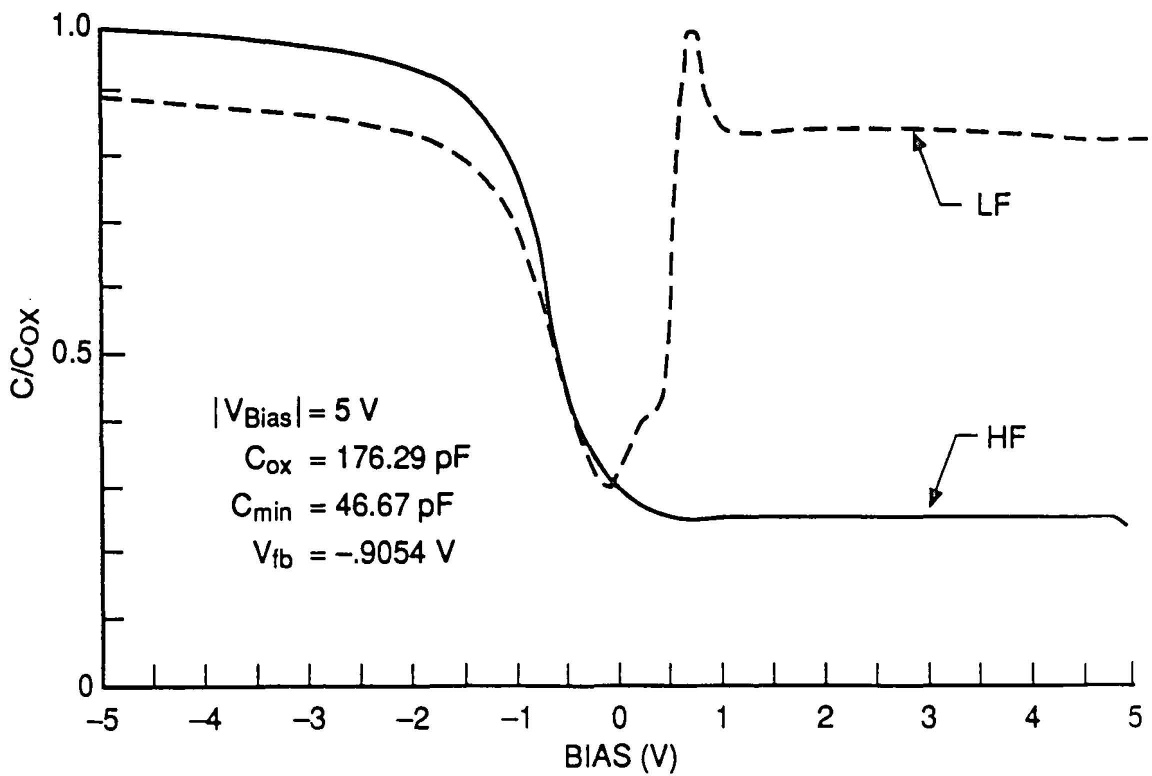
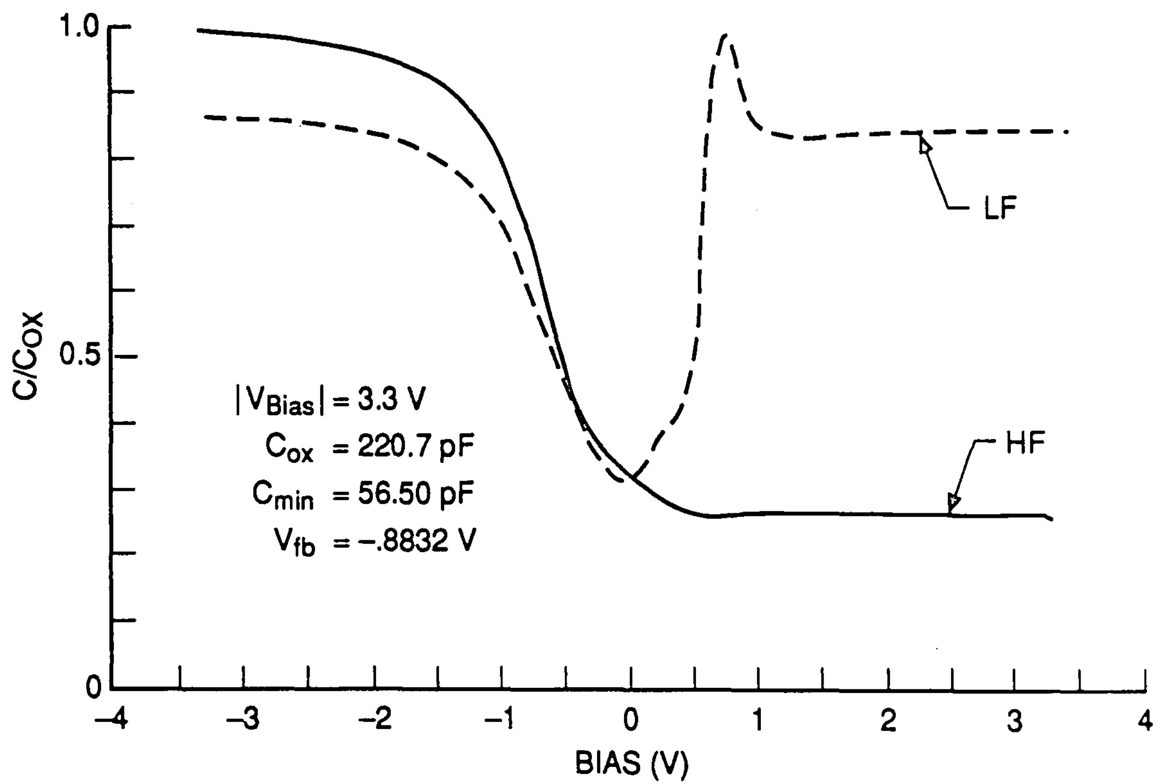


Figure 64: C-V characteristics for 110Å gate oxide.



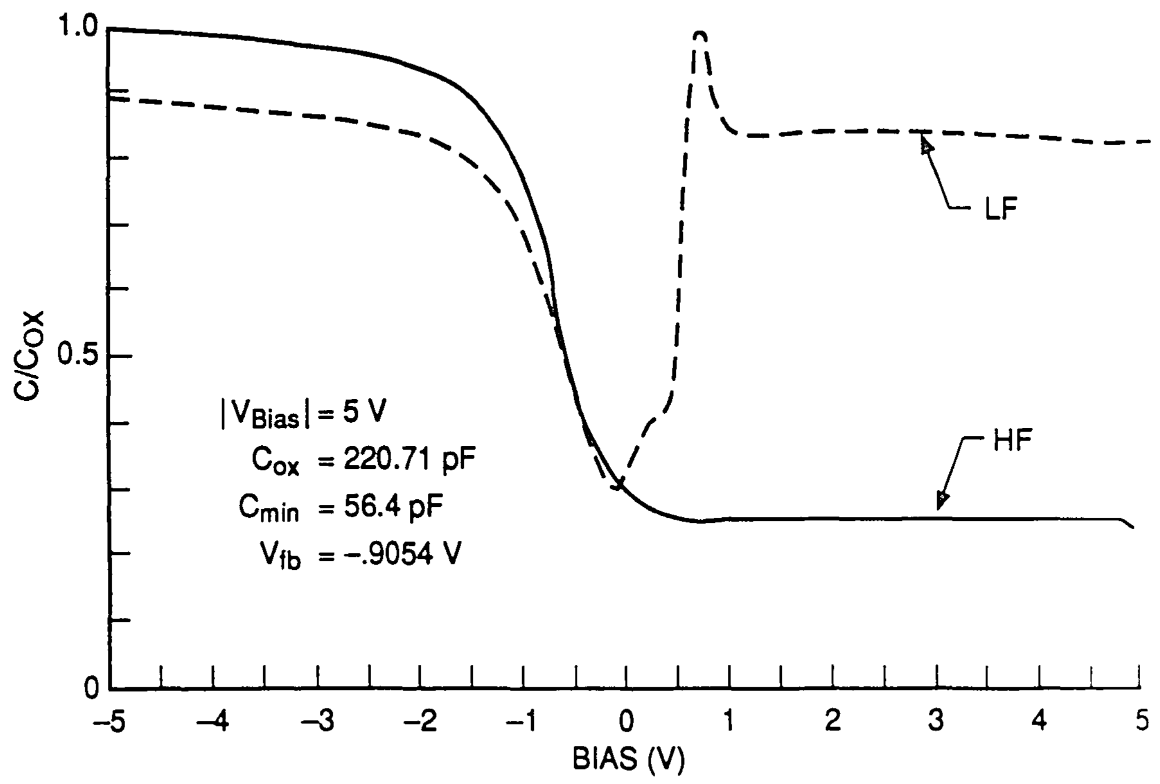
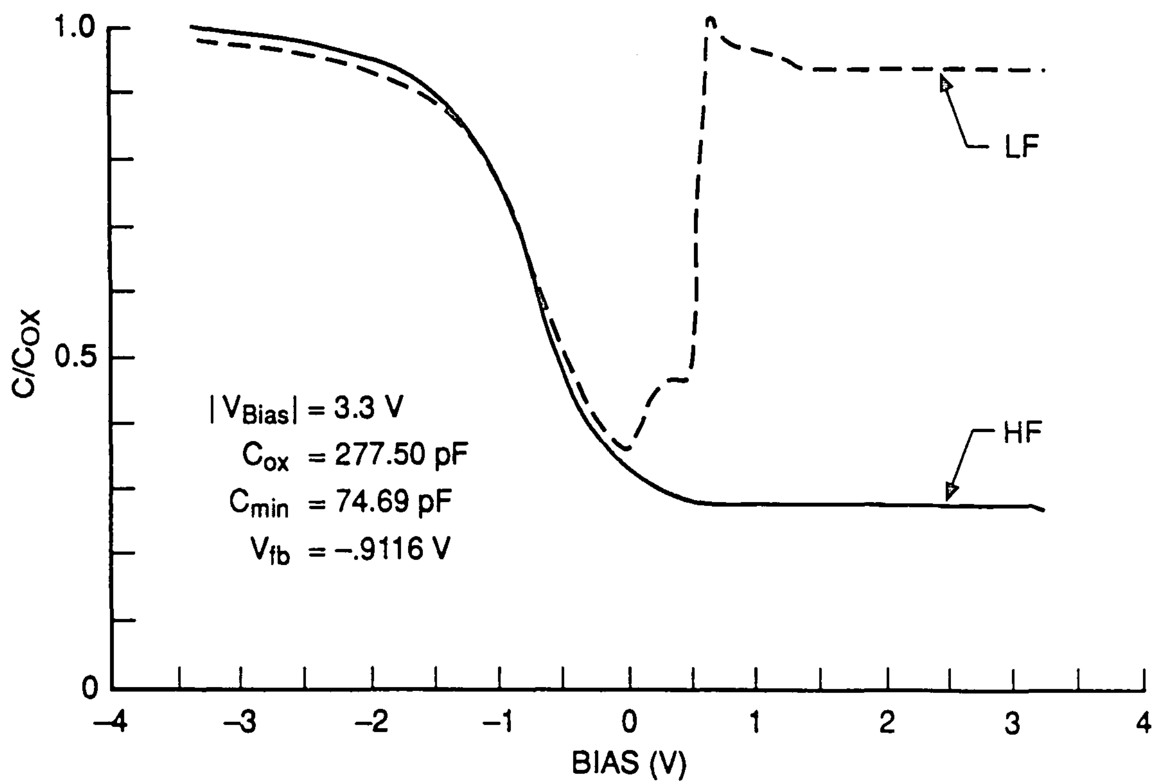


Figure 65: C-V characteristics for 90Å gate oxide.



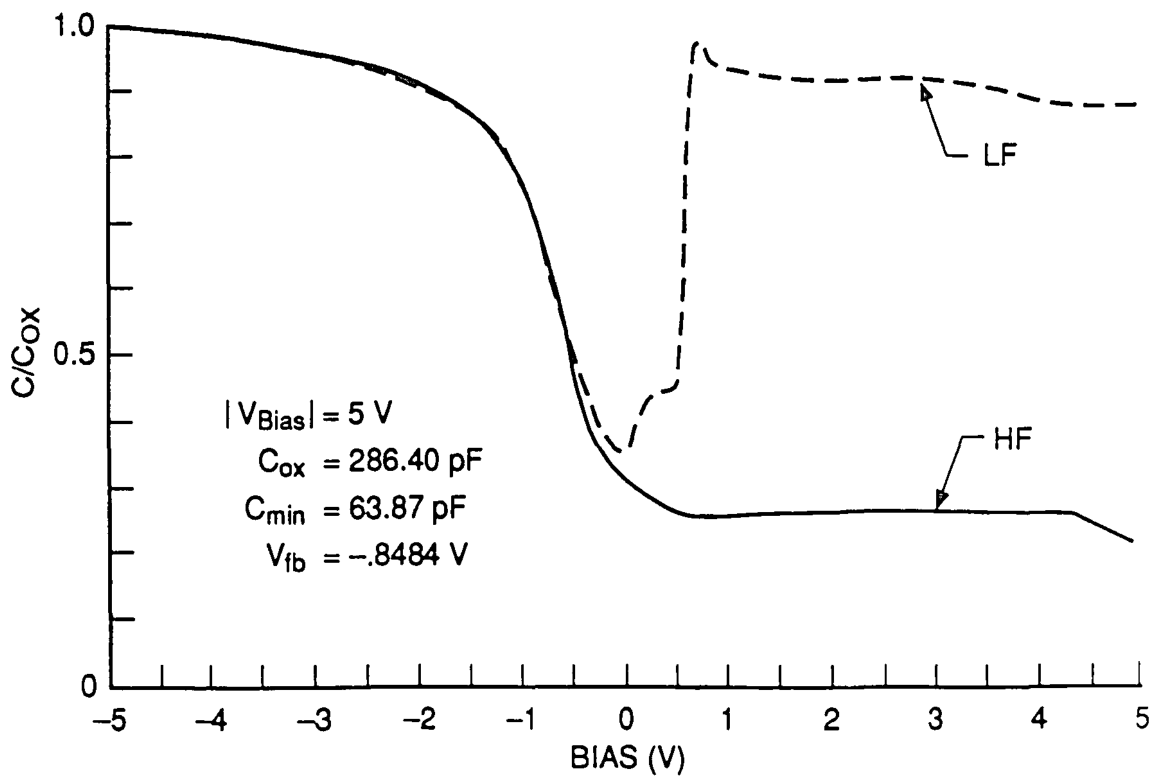


Figure 66: C-V characteristics for 70Å gate oxide.

The C-V curves at low and high frequencies were generated for the various oxide thickness as shown above in fig 62 to 66. These were generated to examine the variation of C_{min} and the percentage recovery of the oxides capacitance after bias. Each oxide thickness was stressed at 5V and 3.3V. From the data gathered, it can be seen that within experimental error the C_{ox}, and the percent recovery were independent of applied voltage. However as theoretically expected, both C_{ox} and C_{min} increases with decreasing oxide thickness while the percent recovery decreased with thickness as tabulated below.

Tox (Å)	C_{ox} (pF/cm^2)	C_{min} (pF/cm^2)	**Recovery** (%)
150	126.1	37.4	98
130	145.8	41.7	91
110	176.3	46.7	88
90	220.7	56.5	90
70	286.4	63.9	91

Table 1: Summary of C-V measurement data for various gate oxide thicknesses

Assuming linear scaling of oxide capacitance with thickness, C_{ox} was found to increase at about 2.3pF/Å increase in gate oxide thickness.

In a attempt to verify the quality and reliability of the gate oxides grown, the breakdown voltage(V_{BD}) and charge to breakdown(Q_{BD}) as a function of T_{ox} were generated as shown in fig 67 and 68. Fig 67 confirms theoretical expectations of decreasing oxide breakdown voltage with decreasing T_{ox} as a result of increased Fowler-Nordheim tunneling gate current induced by the increased vertical fields across the oxide Similarly fig. 68 shows that the charge to breakdown decreases with decreasing oxide thickness.

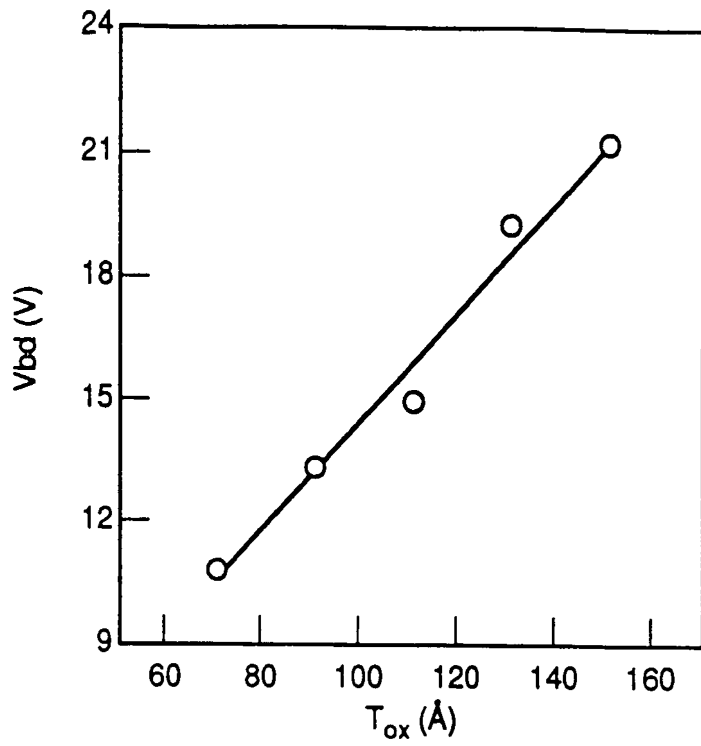


Figure 67: Breakdown voltage versus T_{ox}.

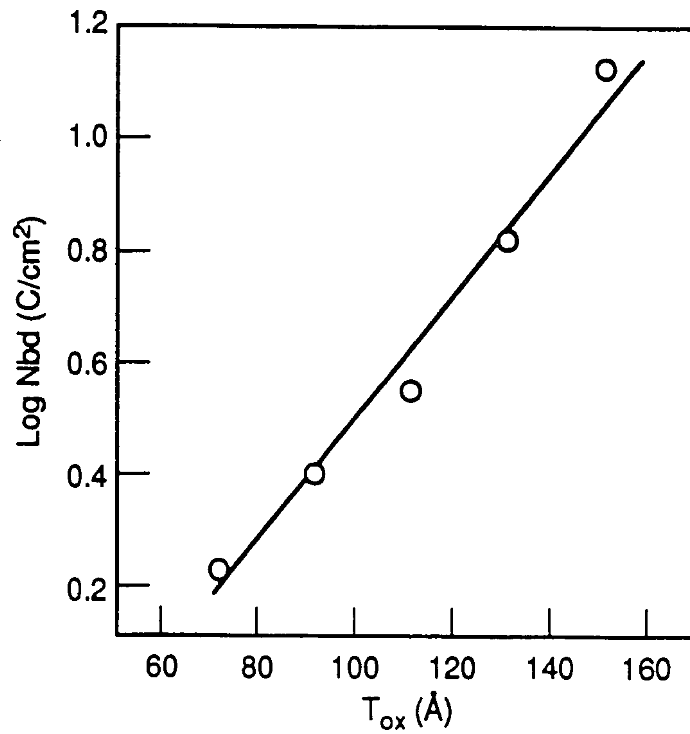


Figure 68: Charge to breakdown versus T_{ox}.

The devices were fabricated with specific interest in the submicron devices. As shown in fig. 69 a consistent difference exists between the 0.9um and 0.7um NMOSFET devices. The Vsnap or lateral punchthrough voltage of the above mentioned L_{coded} devices was found to decrease with gate oxide thickness. This is because of the increased depletion width as T_{ox} decreases thus decreasing the effective channel length and making the device punchthrough susceptible.

Standard I-V characteristics were extracted for all the oxide thicknesses as shown in fig 70-74. This was done for the 0.75um NMOSFET device. As shown, the I_{ds} increased with decreasing gate oxide thickness, but beyond 110Å of gate oxide it appears to decrease a bit. This is believed to be the result of increased gate capacitance C_G due to decreased T_{ox} which impacts the drive capability of transistors.

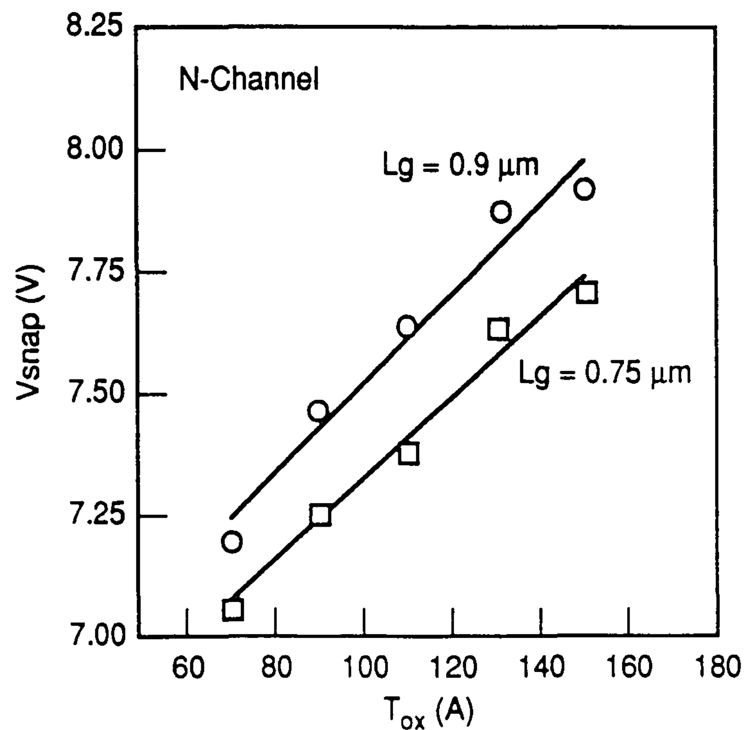


Figure 69: Snapback voltage versus T_{ox}.

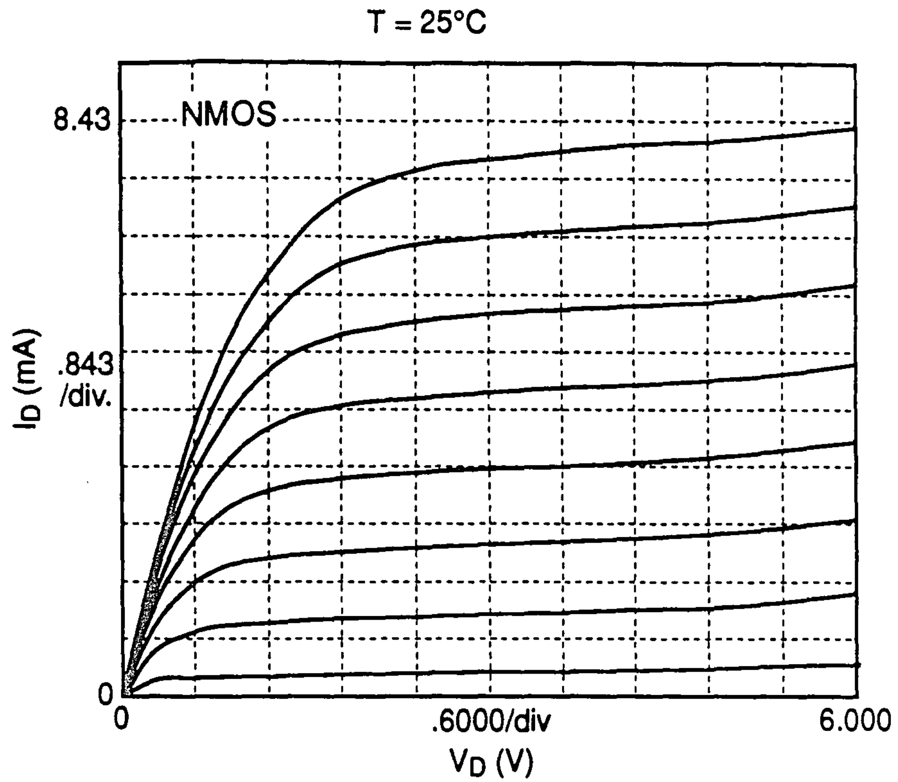


Figure 70a: I_{DS} vs V_{DS} characteristics for 150Å gate oxide thickness.

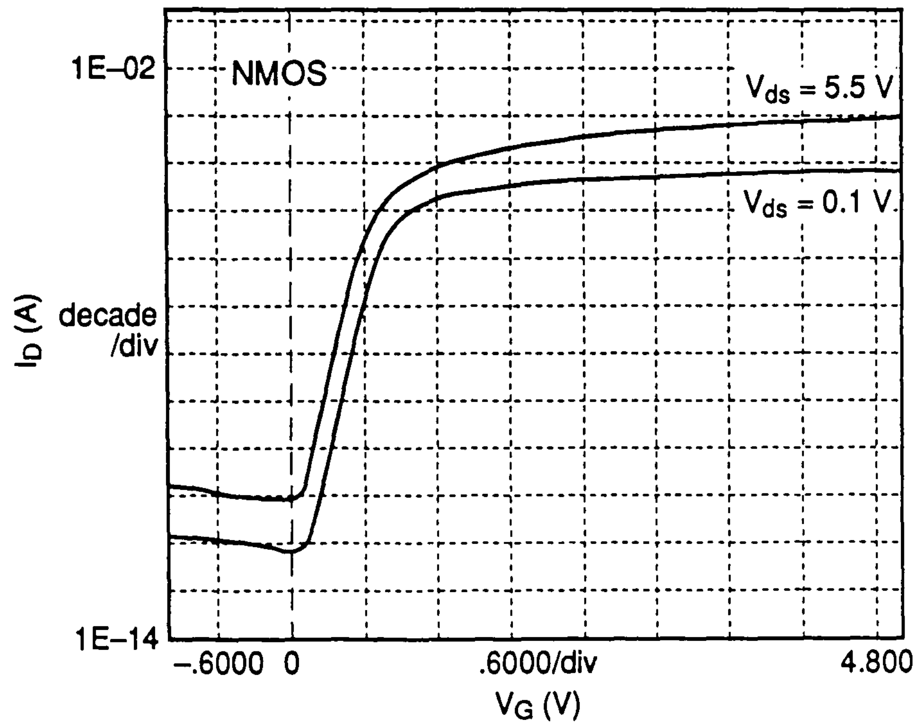


Figure 70b: Subthreshold characteristics for 150Å gate oxide thickness.

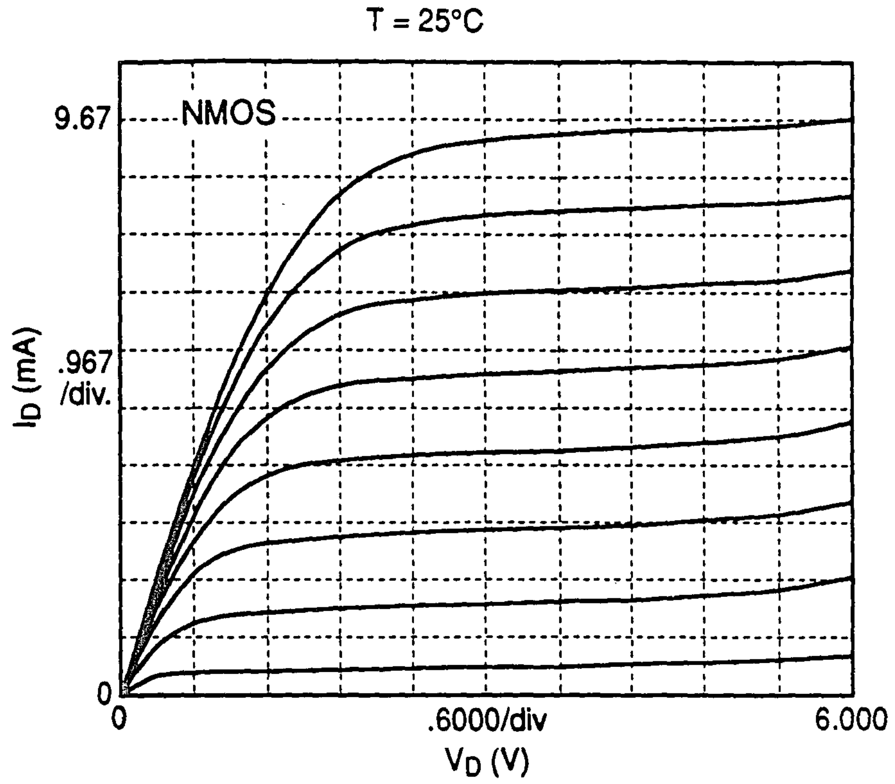


Figure 71a: I_{DS} vs V_{DS} characteristics for 130Å gate oxide thickness.

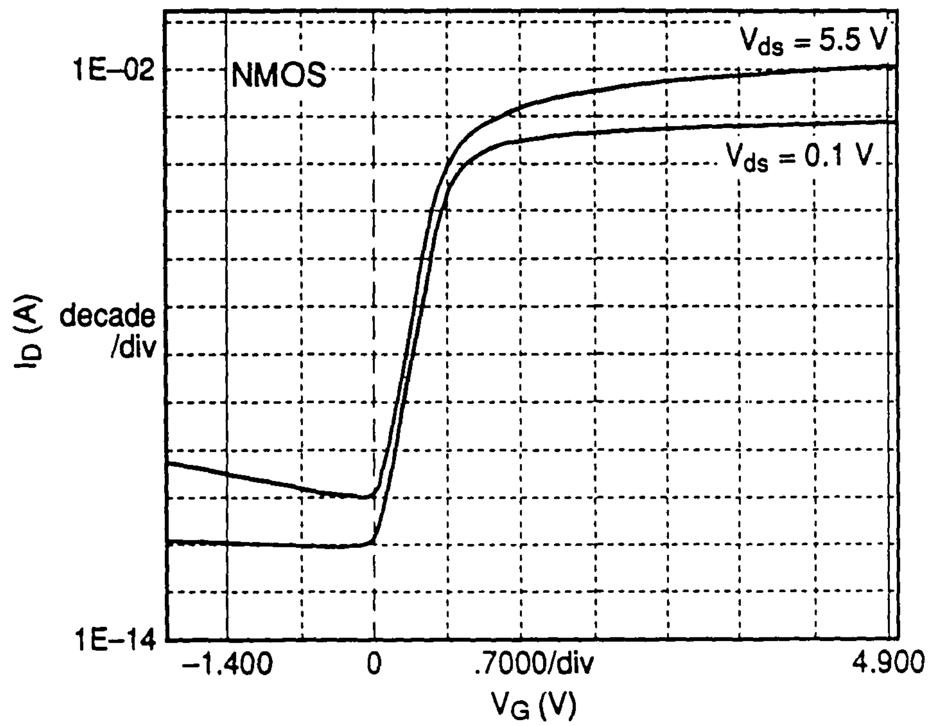


Figure 71b: Subthreshold characteristics for 130Å gate oxide thickness.

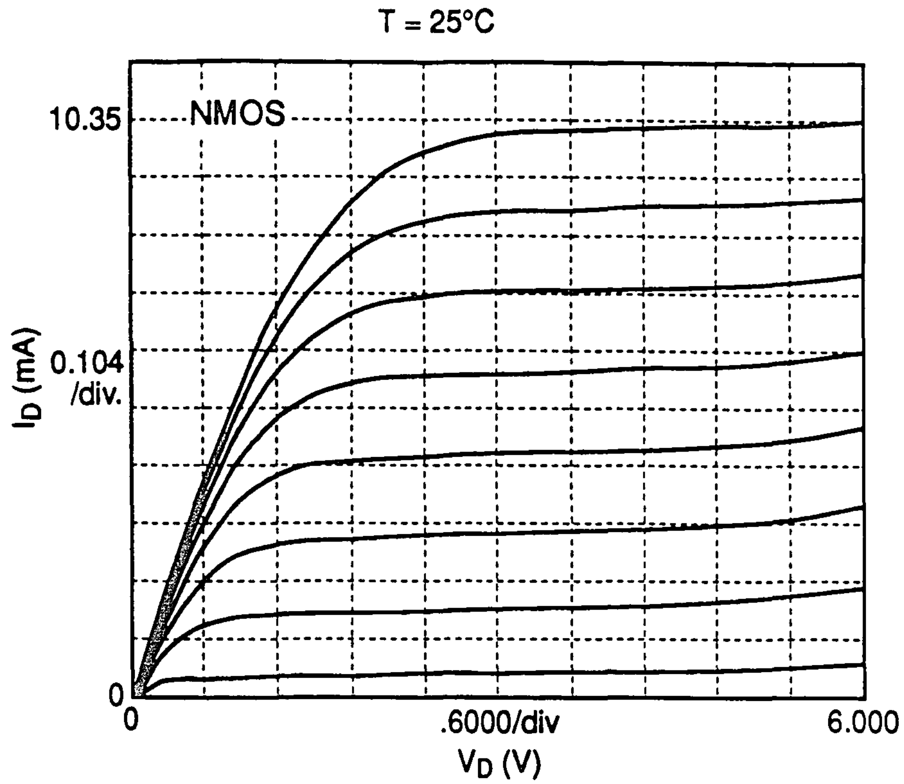


Figure 72a: I_{DS} vs V_{DS} characteristics for 110Å gate oxide thickness.

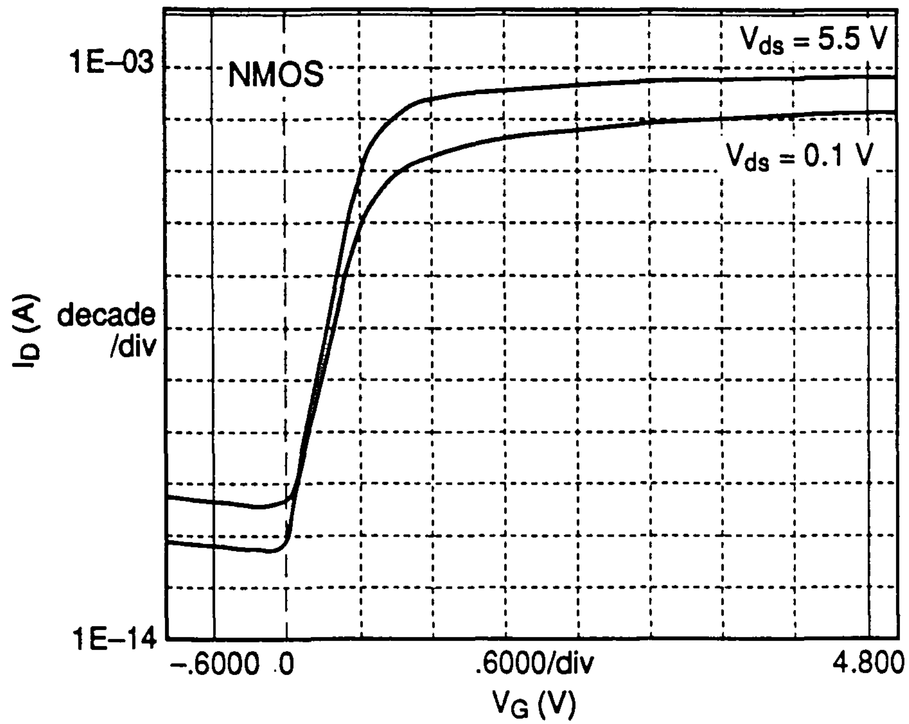


Figure 72b: Subthreshold characteristics for 110Å gate oxide thickness.

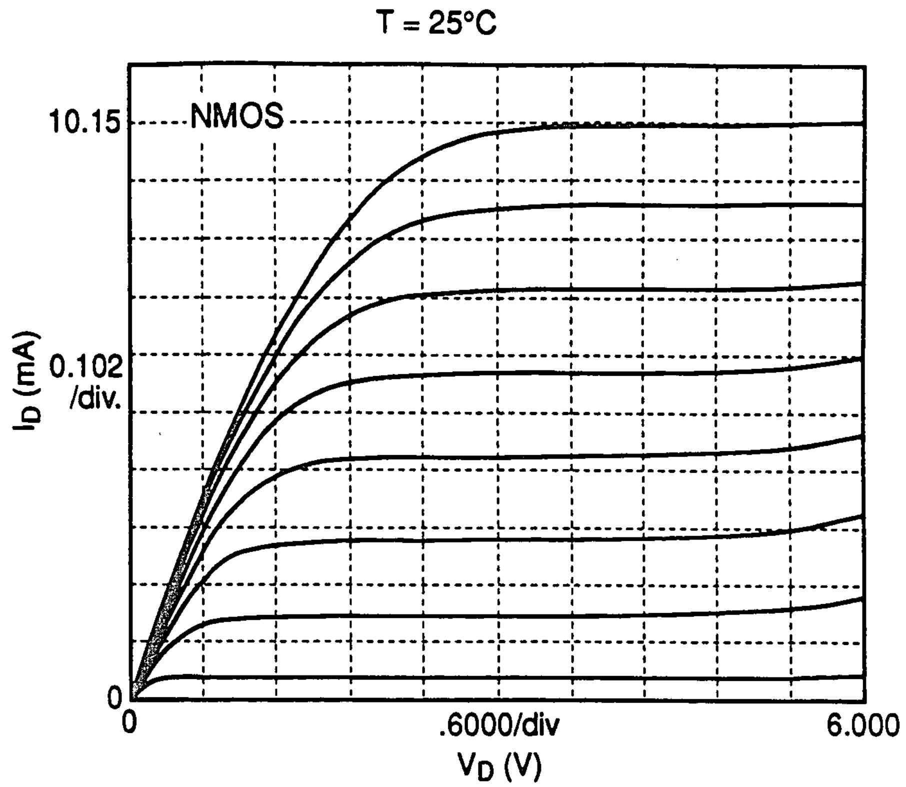


Figure 73a: I_{DS} vs V_{DS} characteristics for 90Å gate oxide thickness.

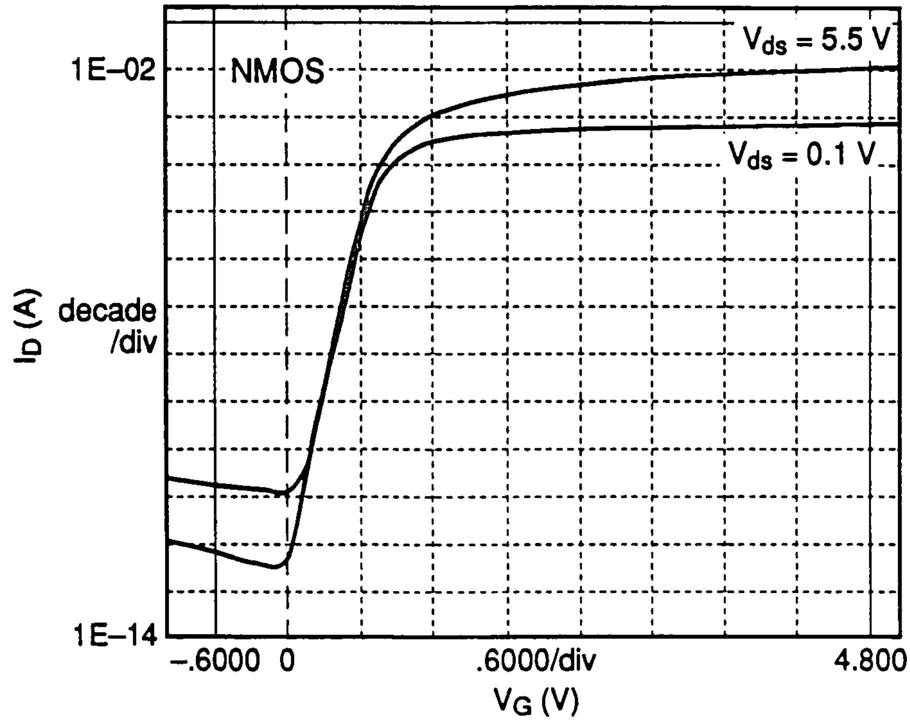


Figure 73b: Subthreshold characteristics for 90Å gate oxide thickness.

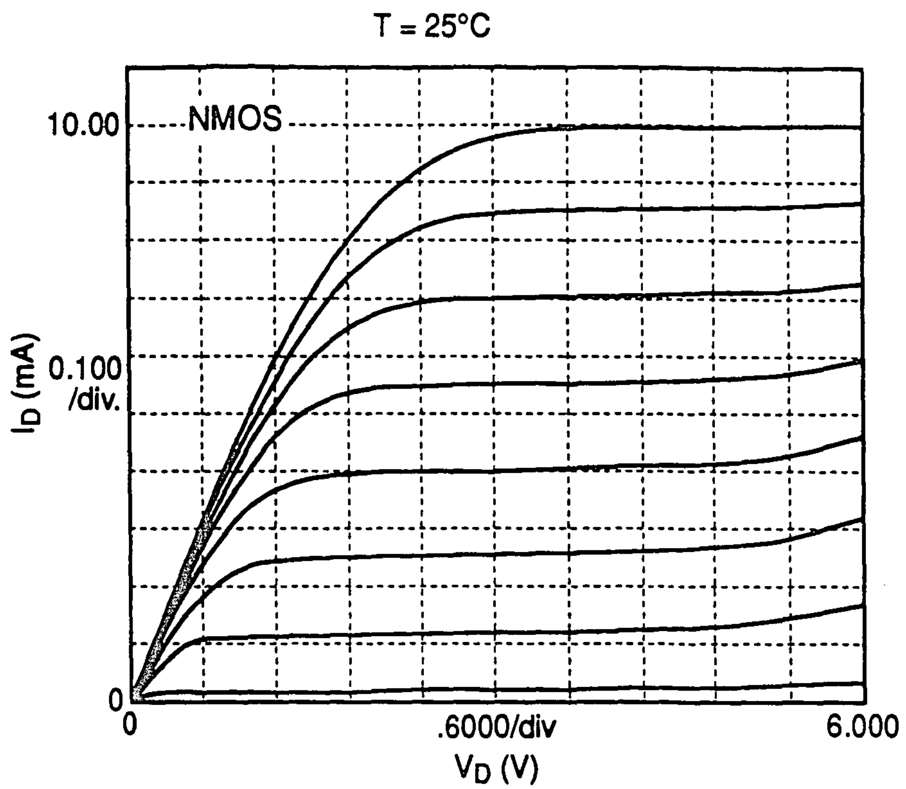


Figure 74a: I_{DS} vs V_{DS} characteristics of 70Å gate oxide thickness.

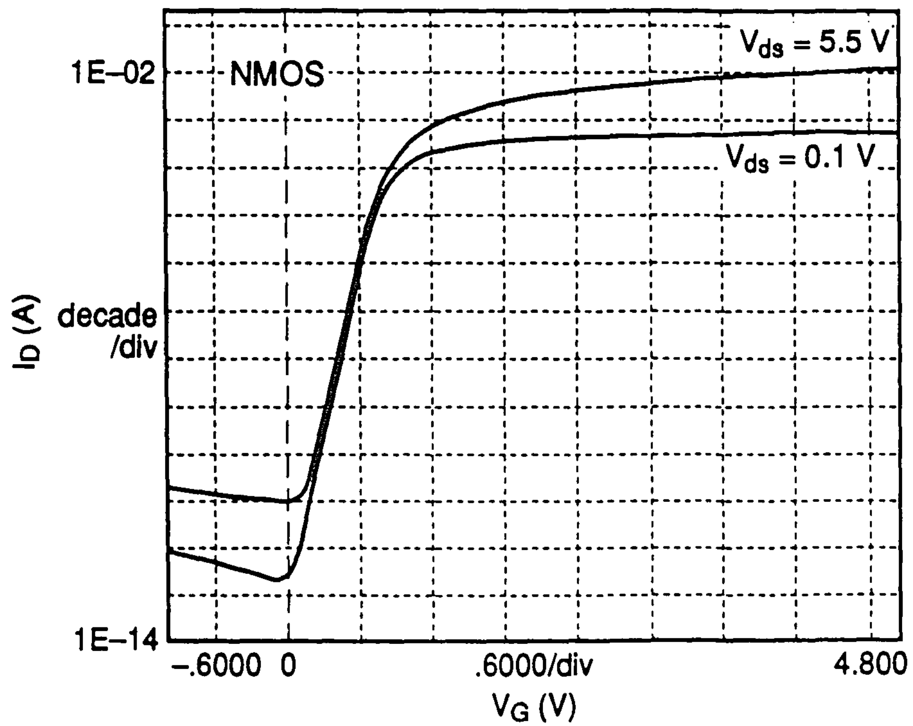


Figure 74b: Subthreshold characteristics for 70Å gate oxide thickness.

The I_{ds} appears to have decreased about 16% as you go from 25°C to 125°C operating condition, while the Ioff in both cases satisfied the typical 1nA/um at 125C criterion for leakage in VLSI technologies. Tabulated below in table 2 are the I_{ds}, and Ioff for N and PMOSFETs which appears to support the previously discussed "bell" shape phenomena of propagation delay.

Tox (Å)	Ion@25°C (mA)	Ion@125°C (mA)	Ioff@25°C log(A)	Ioff@125°C log(A)
150	8.44	6.83	11.16	10.72
130	9.67	8.07	10.27	10.66
110	10.36	8.56	11.32	10.67
90	10.15	9.31	11.48	10.72
70	10.1	8.65	11.50	10.70

Table 2: Ion/Ioff data for NMOS of L_{CODED} = 0.75um as a function of gate oxide thickness

Tox (Å)	Ion@25°C (mA)	Ion@125°C (mA)	Ioff@25°C log(A)	Ioff@125°C log(A)
150	3.73	3.20	11.70	10.80
130	4.49	4.09	10.72	10.79
110	5.18	4.60	11.70	10.60
90	5.0	4.33	11.58	10.80
70	4.81	4.20	10.89	10.68

Table 3: Ion/Ioff data for PMOS of L_{CODED} = 0.85um as a function of gate oxide thickness

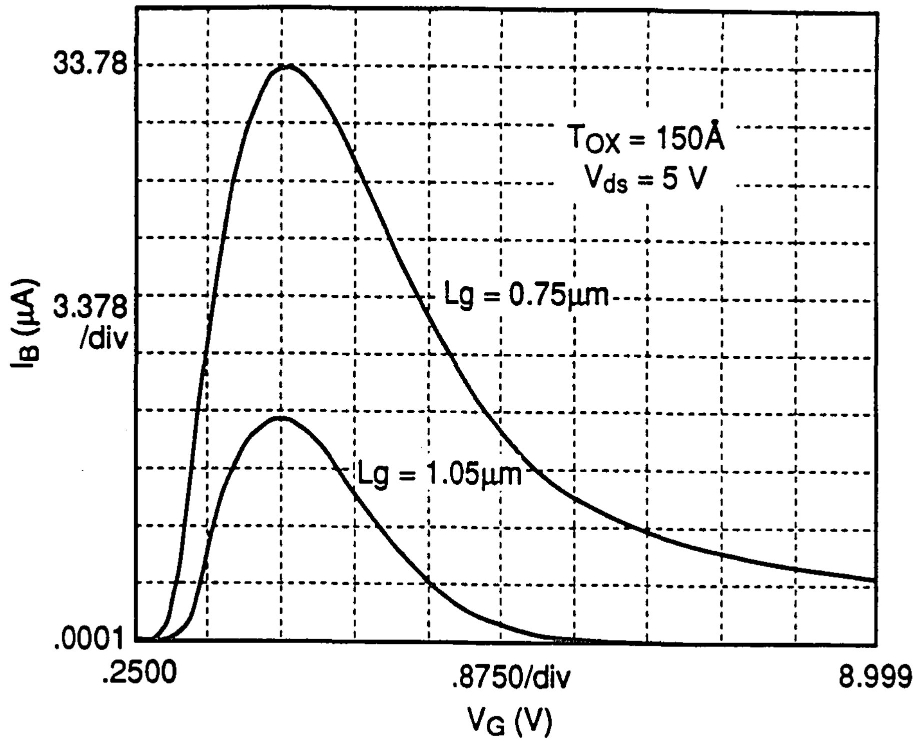
4.3 DEVICE AGING CHARACTERIZATION

Transistor aging results is of paramount importance in VLSI technology. This is results can be used to predict the device lifetime thus providing prolonged reliability information about devices for a given technology.

Figure 75 to 79 shows the substrate currents as a function of the applied gate voltage at constant drain voltage with V_{ds}=5V for different coded gate lengths. It is observed that the peak substrate current increases with decreasing L_{coded} and increased with decreasing gate oxide thickness. These curves were used to judiciously determine the optimum gate voltage for a given drain voltage where impact ionization due to hot electron injection was a maximum. This was then employed in the hot carrier aging characterization of the transistors to establish some reliability dependence of device lifetime on gate oxide thickness .

Figures 80 through 84 shows graphically the time to 10% g_m degradation as a function of device substrate current for a fixed V_{gs}=2V and V_{ds}=7V, for the various gate oxide thicknesses. The data was plotted on a log-log scale to facilitate the extraction of the a and b coefficients which allows for the computation of the lifetime τ_{LIFE} for a given substrate current. Recalling from the theory section, the lifetime of a device could be related to the substrate current for a given gate oxide thickness and L_{gate} as

$$\tau_{LIFE} = a\left(I_{sub}\right)^{-b}.$$



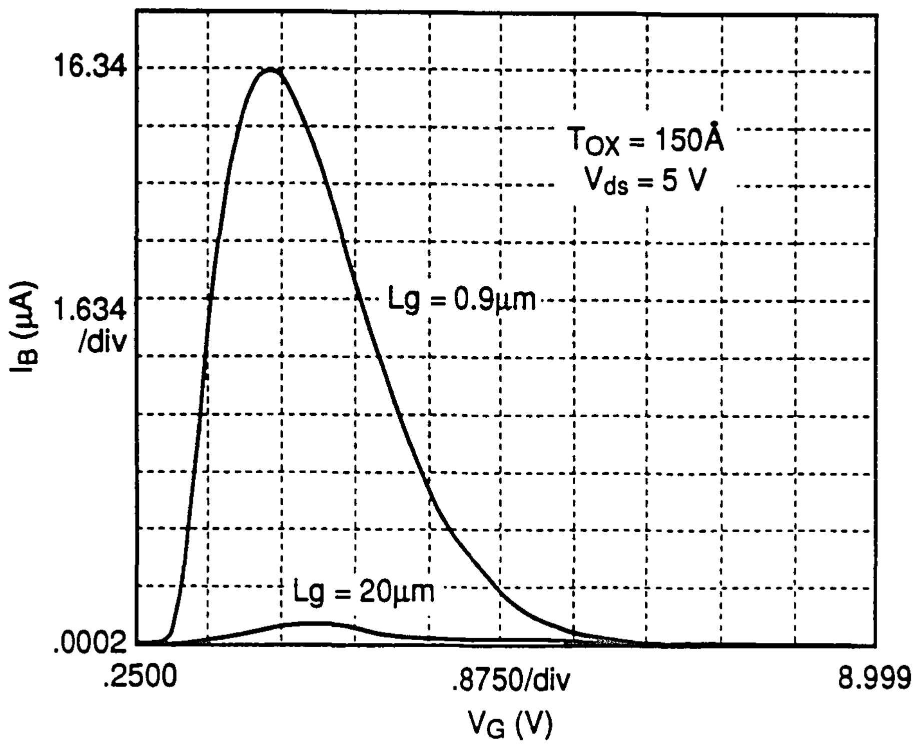
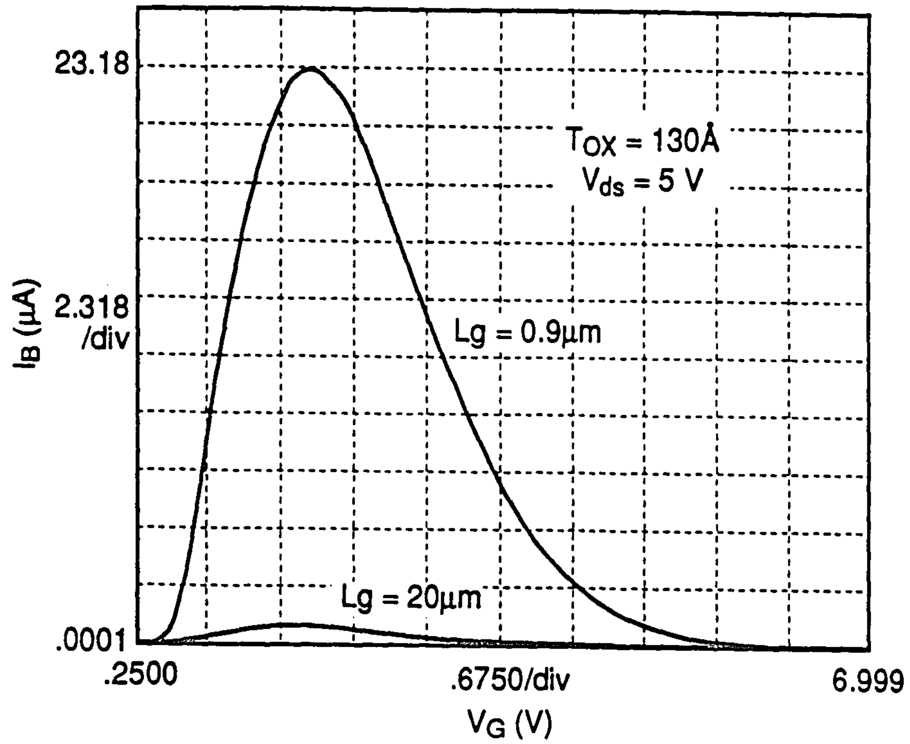


Figure 75: Substrate current as a function of gate voltage for variable L_{gate} and T_{ox}.



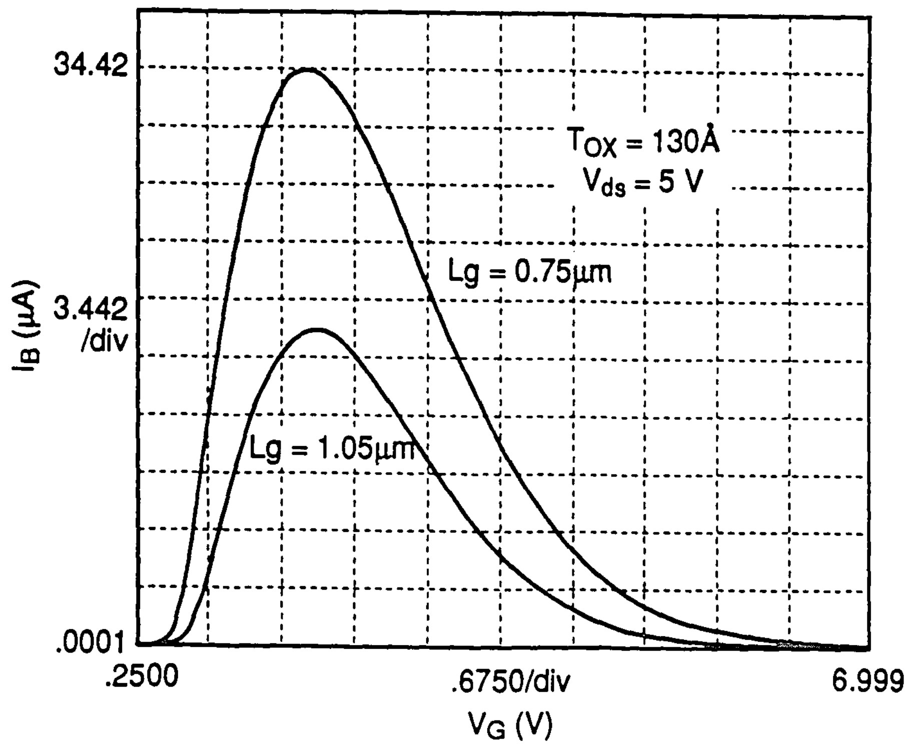
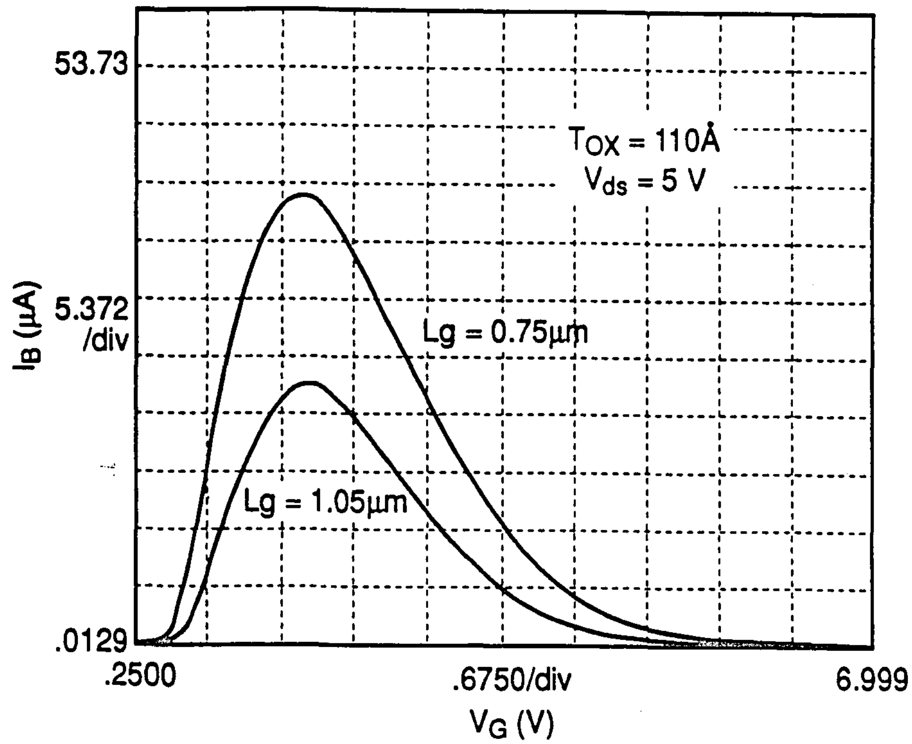


Figure 76: Substrate current as a function of gate voltage for variable L_{gate} and T_{ox}.



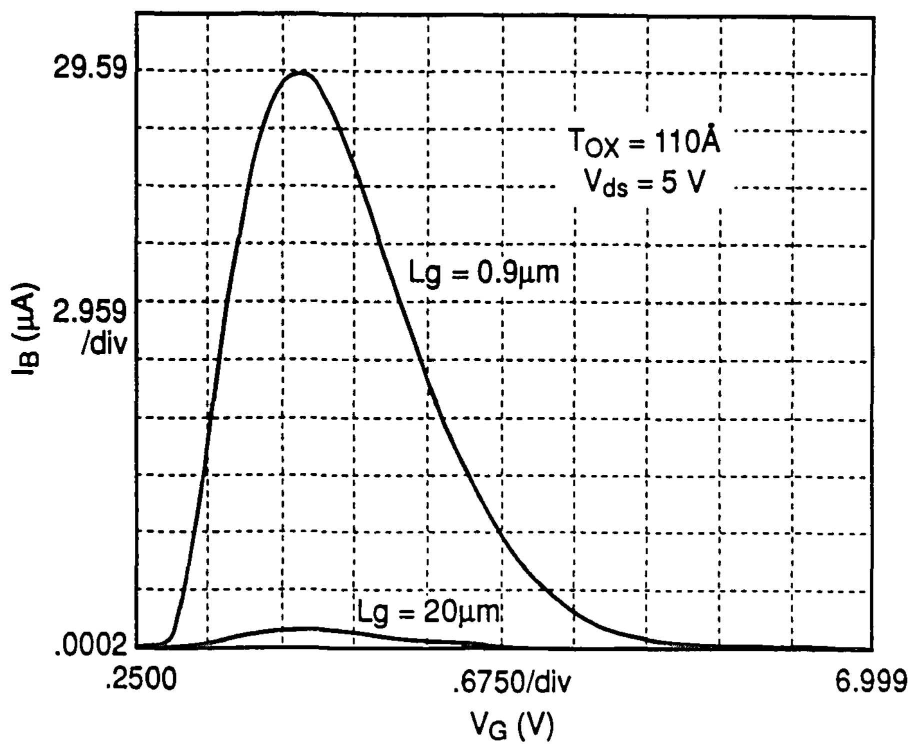
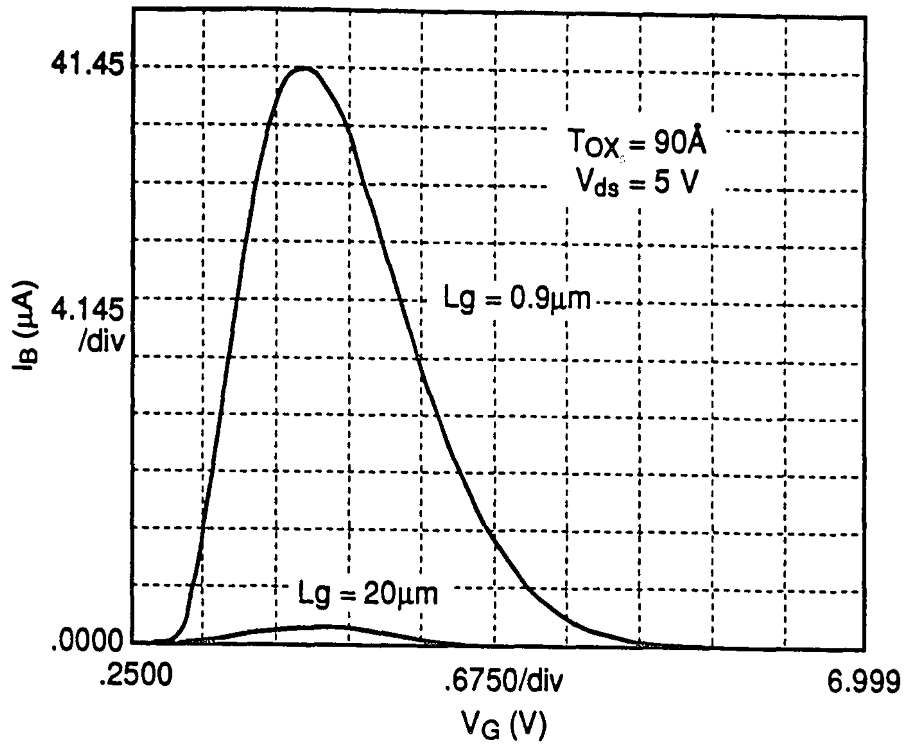


Figure 77: Substrate current as a function of gate voltage for variable L_{gate} and T_{ox}.



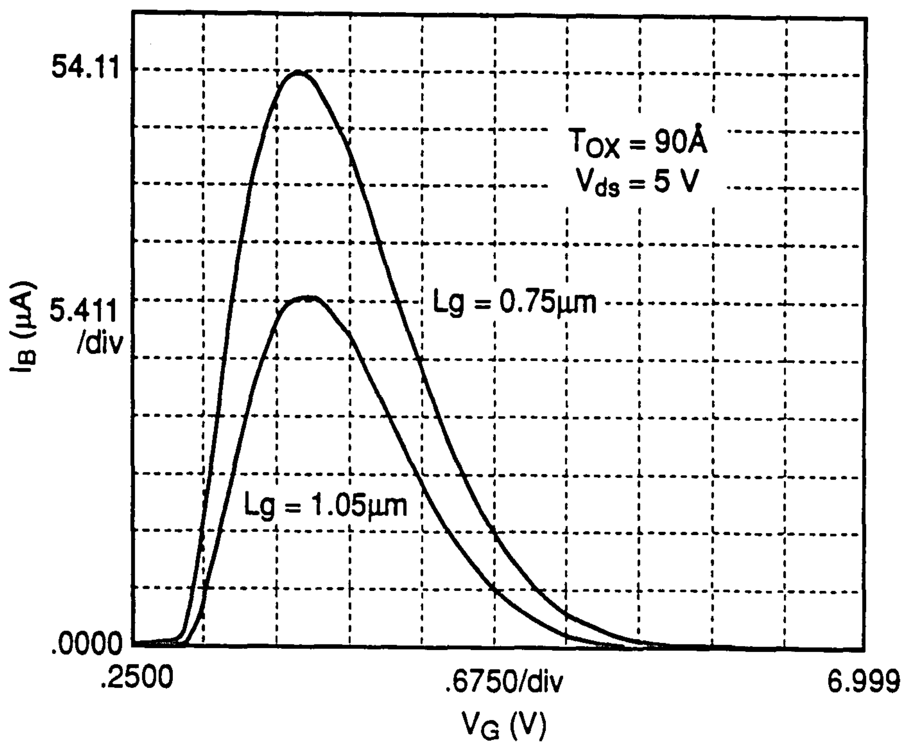
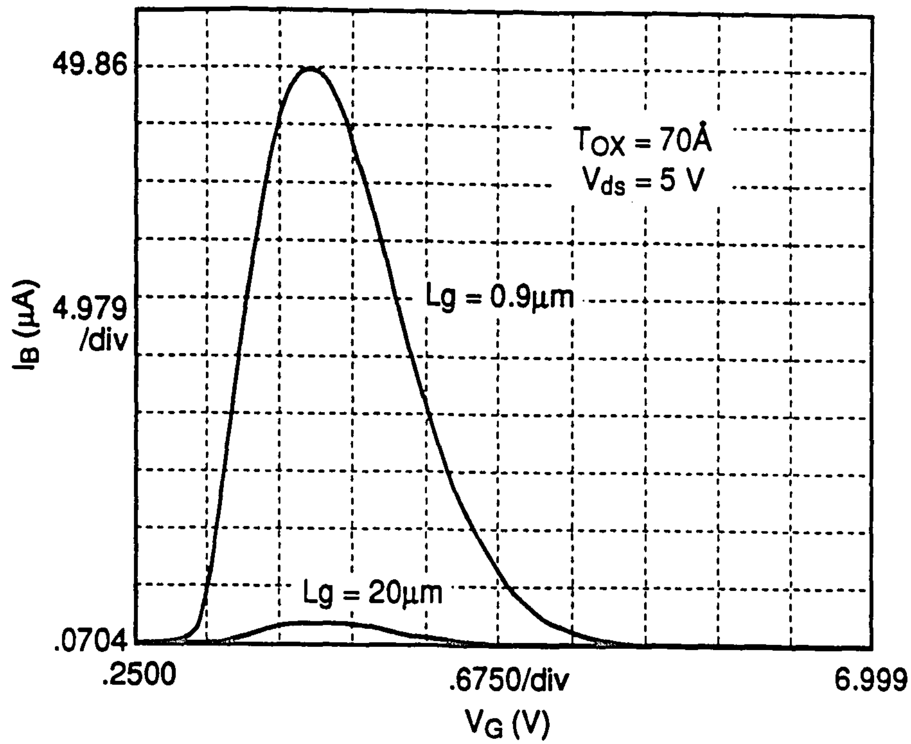


Figure 78: Substrate current as a function of gate voltage for variable L_{gate} and T_{ox}.



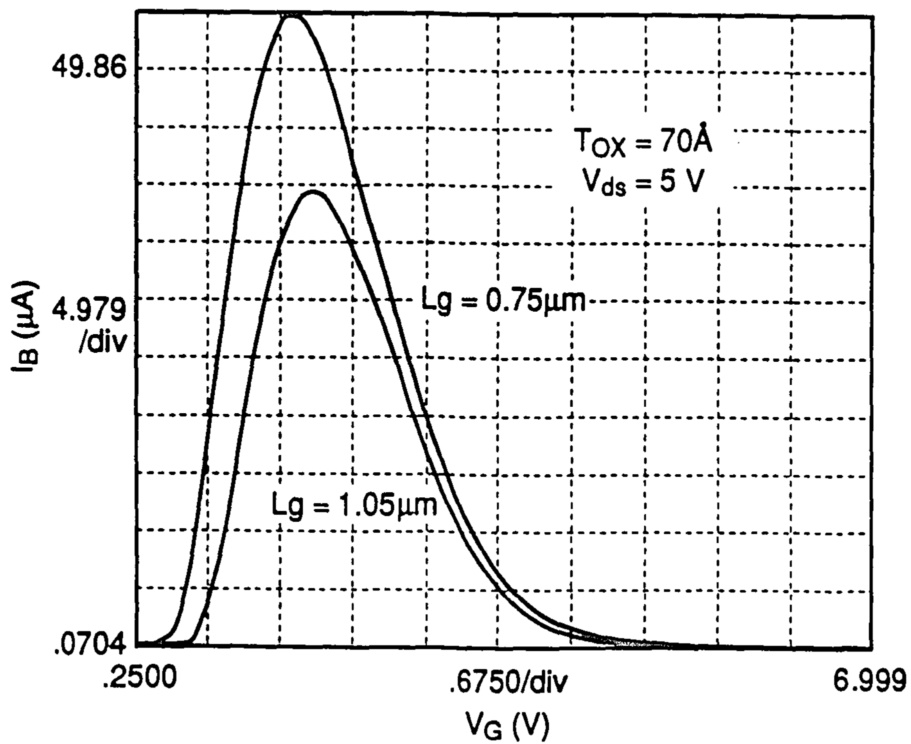


Figure 79: Substrate current as a function of gate voltage for variable L_{gate} and T_{ox}.

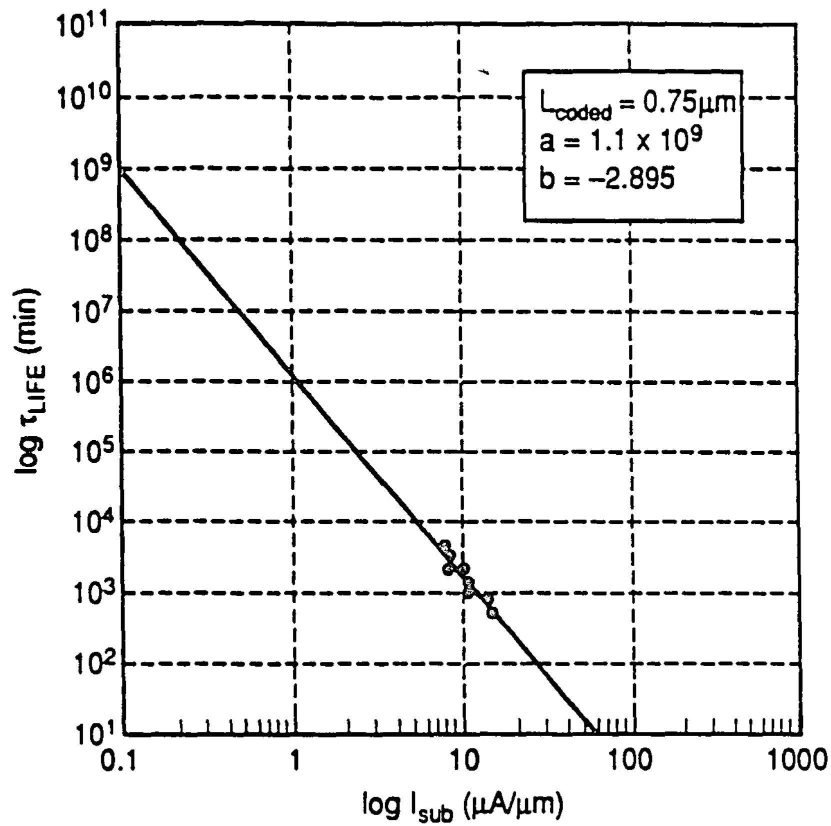


Figure 80: 10% transconductance (g_m) degradation as a function of I_{sub} for 150Å gate oxide.

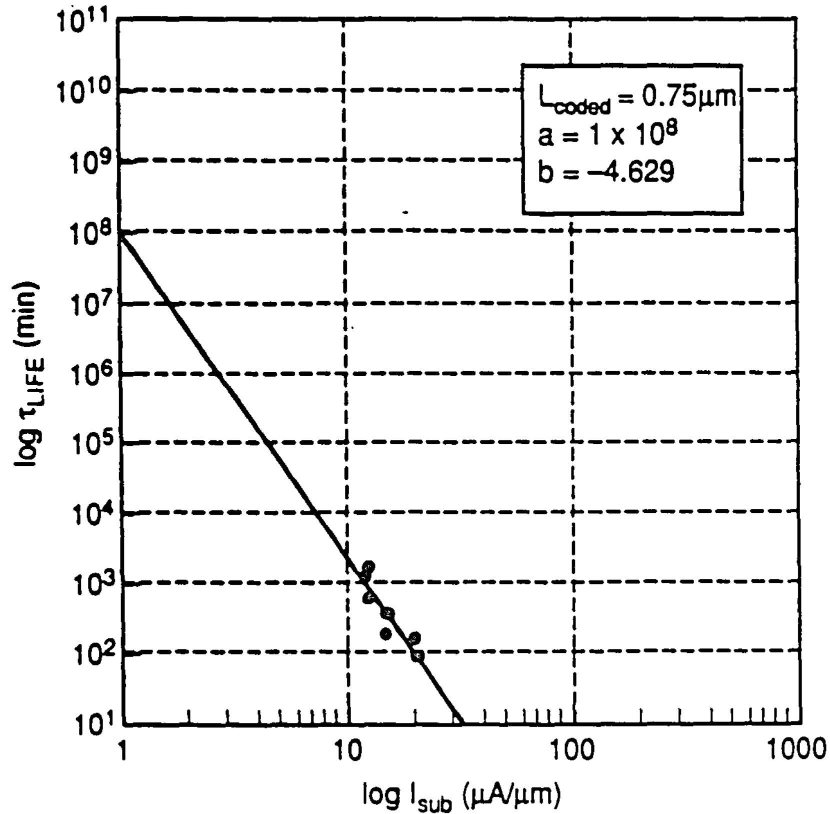


Figure 81: 10% transconductance (g_m) degradation as a function of I_{sub} for 130Å gate oxide.

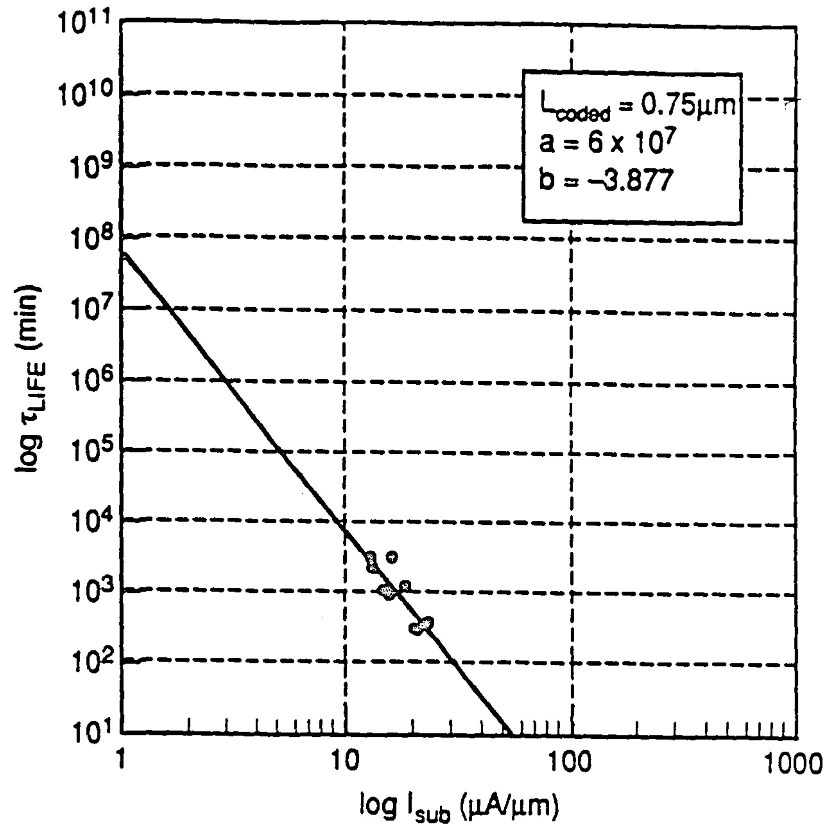


Figure 82: 10% transconductance (g_m) degradation as a function of I_{sub} for 110Å gate oxide.

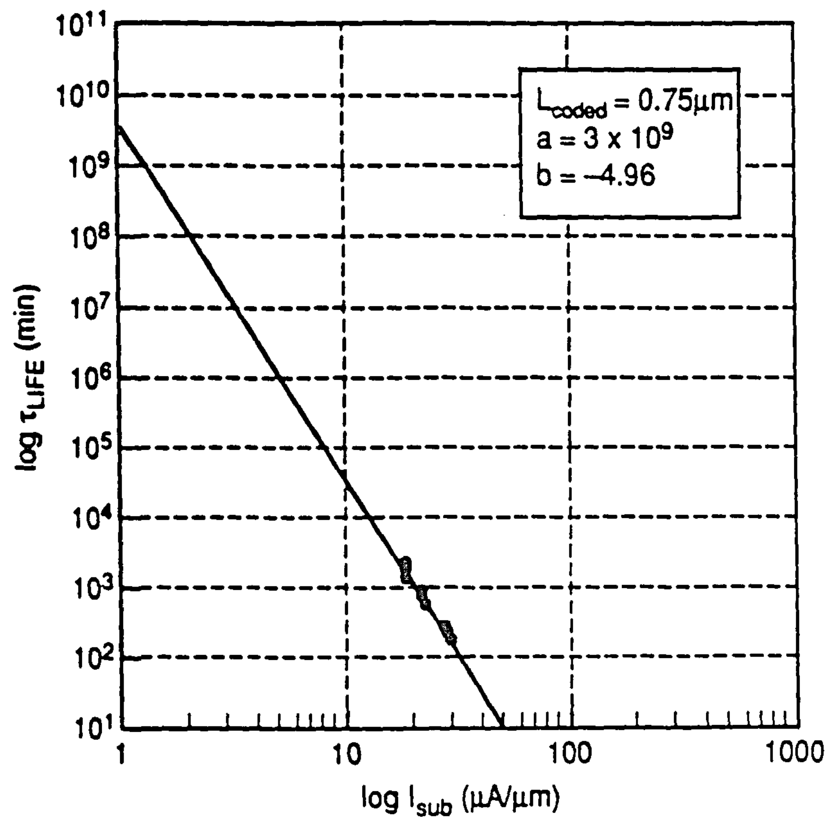


Figure 83: 10% transconductance (g_m) degradation as a function of I_{sub} for 90Å gate oxide.

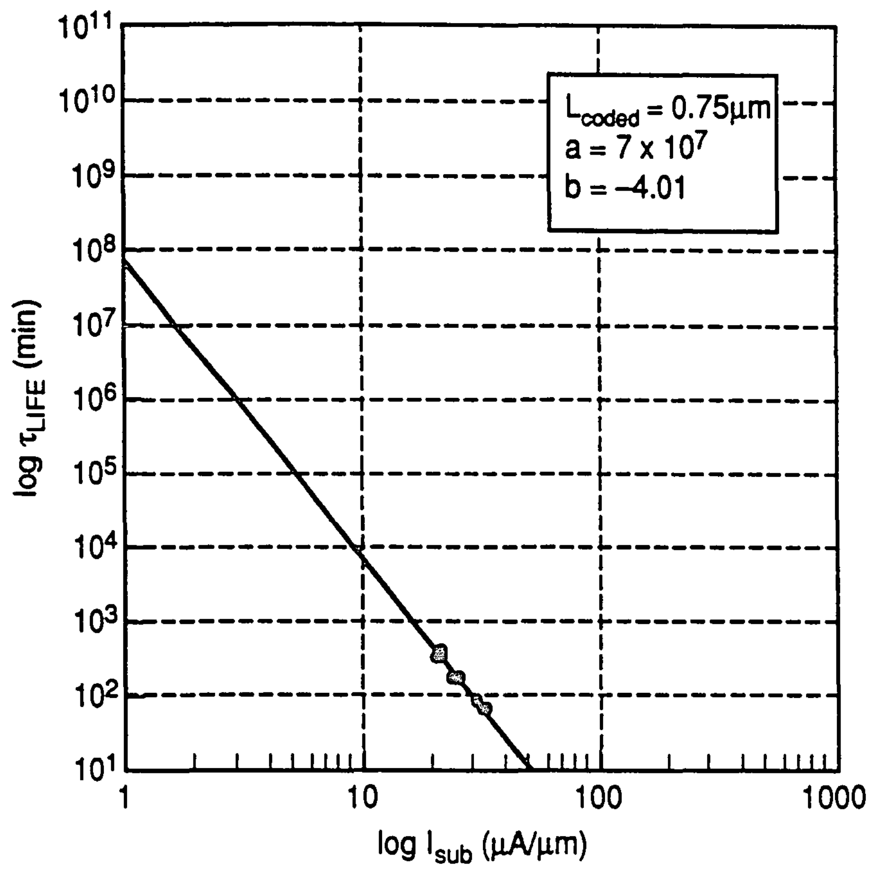


Figure 84: 10% transconductance (g_m) degradation as a function of I_{sub} for 70Å gate oxide.

The **a** and **b** coefficients are shown on all the above mentioned figures. Tabulated below is the comparison of τ_{LIFE} and I_{sub} for various gate oxide thicknesses and operating bias conditions.

T_{OX} (Å)	$I_{SUB}@V_{ds}=5V$ $V_{gs}=2V$	$I_{SUB}@V_{ds}=7V$ $V_{gs}=2V$	$\tau_{LIFE}@V_{ds}=5V$ $V_{gs}=2V$	$\tau_{LIFE}@V_{ds}=7V$ $V_{gs}=2V$
150	$1.28\mu A/\mu m$	$15.77\mu A/\mu m$	$1024 yrs$	$0.707 yrs$
130	$1.74\mu A/\mu m$	$20.5\mu A/\mu m$	$14.5 yrs$	84.47min
110	$2.12\mu A/\mu m$	$25.3\mu A/\mu m$	$6.12 yrs$	217.9min
90	$2.77\mu A/\mu m$	$30.5\mu A/\mu m$	$36.2 yrs$	130min
70	$3.24\mu A/\mu m$	$34.6\mu A/\mu m$	$1.19 yrs$	47.14min

Table 4: Summary of Device degradation for NMOS of L_{CODED} = 0.75um as a function of V_{ds} and gate oxide thickness.

The data in table 3 clearly shows the strong dependence of τ_{LIFE} on substrate current which increases with increasing V_{ds} for a given V_{gs}. This is because of the dependence of I_{sub} on I_{ds} which is a function of V_{ds} as depicted in the equation below

$$I_{SUB} = C_1 I_{ds} e^{-\frac{\xi_i}{\lambda_e E_m}} = \frac{I_{ds} A E_{\max} e^{-\frac{B}{E_{\max}}}}{B[dE(x)/dx]}$$

where C_1 is determined experimentally, λ_e is the electron mean free path, ξ_i is the impact ionization coefficient and E_m the peak electric field at the drain end.

From the tabulated data in table 4 one can see that as the gate oxide thickness decreases the device becomes more susceptible to impact ionization. This gives rise to higher substrate currents and subsequently lower lifetime τ_{LIFE}. The 90Å cell apparent defiance of this trend could be attributed to a superior oxide growth conditions. This

τ_{LIFB} dependence on I_{sub} for a given T_{ox} reinforces the fact that power supply voltage should be scaled together with T_{ox} so as to maintain reasonable device reliability.

REFERENCES

1. B.E. Deal and A.S. Grove, J. Appl. Physics; 36 pg. 3770 (1965)

2. E.A. Irene and Y.J. Van der Mevlen, J. Electrochemical society 123 pg.1380 (1974)

3. N. Cabrera and N.F. Mott, Rep. Prog. Physics 30 pg. 3770 (1948)

4. P. Jorgensen, J. Chem. Physics, 37 pg. 874 (1962)

5. A.S. Grove, "Physics and Technology of Semiconductor Devices", (1967)

6. M.Hamasaki, Solid-State Electronics, 25 pg. 479 (1982)

7. Y.Z. Lu and Y.C. Chen, J. Appl. Physics, 56 pg. 1608 (1984)

8. S.M. Hu, Appl. Physics Letters, 42 pg. 872 (1983)

9. A.G. Revesz and R.J. Evans, J. Phys. Chem. Solids, 30 pg. 551 (1961)

10. S.W. Ing et al, J. Electrochemical Society 109 pg. 229 (1962)

11. A.G. Revesz, J. Electrochemical Society 80-2, pg. 897 (1980)

12. R.J. Jaccodine and W.A. Schlegel, J. Appl. Physics, 37 pg. 2429 (1966)

13. P.G. Borden, Appl. Phys. Letters, 36 pg. 829 (1980)

14. E.P EerNisse, ibid., 35 pg. 8 (1979)

15. E.A. Irene, J. Appl. Physics, 54 pg. 5416 (1983)

16. A. Fargeix et al, J. Appl. Physics, 54 pg. 2878 (1983)

17. G.D. Watkins, J. Appl. Physics, 53 pg. 7097 (1982)

18. H.Z. Massoud, J.D. Plummer, E.A. Irene, J. Electrochemical Society, 132 pg. 2693 (1985)

19. E.A. Irene, ibid., 125, pg. 1708 (1978)

20. M.A. Hopper et al, ibid., 122, pg. 1216 (1975)

21. A.G. Revesz, J. Electrochemical Society 81-1, pg. 701 (1981)

22. R.J. Jaccodine and W.A. Schlegel, J. Appl. Physics, 37, 2429 (1966)

23. P.G. Borden, Appl. Phys. Letters, 36, pg. 829 (1980)

24. E.P EerNisse, ibid., 35 pg. 12 (1977)

25. T. Hamamoto et al, "Asymmetry of the Substrate Current Characteristics Enhanced by the Gate birds beak", Proc. Symp. VLSI Technology, pg. 67-68 (1986)

26. E.A. Irene et al, J. Electrochemical Society, 129, 2594 (1982)

27. S.C. Sun and J.D. Plummer, "Electron mobility in inversion and accumulation on thermally oxidized silicon surfaces", IEEE Trans. Electron Devices, ED-27 pg. 1497-1508 (1980)

28. E.H. Snow, "Fowler-Nordheim tunneling in SiO_2 films", Solid State Communication; vol 5 pg. 813-815 (1967)

29. C.S. Jeng et al, "High-field generation of electron traps and charge trapping in ultra-thin SiO_2", IEDM Tech. Dig. pg. 388-391 (1981)

30. M.S. Liang et al, "MOSFET degradation due to stressing of thin oxide", IEDM Tech. Dig. pg 50-53 (1982)

31. Y. N. Cohen et al, "Characterization of simultaneous bulk and interface high-field trapping effects in SiO_2", IEDM Tech. Dig. pg. 182-185 (1983)

32. S. Manzini, "Effects of coulomb scattering in n-type silicon inversion layers", J. Appl. Physics vol 57 (1985)

33. T. Tscuhiya et al, "Hot carrier degradation mechanism in Si NMOSFETs", Tech. Dig. Int. Conf. Solid State Devices and Materials, pg. 21-24 (1985)

34. Y. Nissan-Cohen et al,"Measurement and analysis of hot-carrier-stress effect on NMOSFETs using substrate current characterization", IEEE Electron Device Letters, EDL-7, pg. 451-453 (1986)

35. H. Gesch, J.P. Leburton and G.E. Dorda, "Generation of interface states by hot hole injection in MOSFETs", IEEE Trans. Electron Devices, ED-32 pg. 691-699, (1985)

36. H.E. Maes and G. Groeseneken, "Determination of spatial surface state density distribution in MOS and SIMOS transistors after channel hot electron injection", Electron Letters, vol 18 pg. 913-918 (1982)

37. K.R. Hofmann et al, "Hot-electron and hole-emission effects in short n-channel MOSFETs", IEEE Trans. Electron Devices, ED-32, pg. 691-699 (1985)

38. G. Hu and R.H. Bruce, "Design tradeoffs between surface and buried channel PFETs", IEEE Trans. Electron Devices, ED-32 pg 584 (1985)

39. J.Y.C. Sun et al, "Study of boron penetration through thin oxide with P+ polysilicon gate", Dig. Int. Symp. on VLSI Technology pg. 17 (1989)

40. F. Bager et al, "The influence of fluorine on threshold voltage instabilities in P+ polysilicon gated p-channel MOSFETs", IEDM Tech. Dig., pg. 443 (1989)

41. J. Sung et al, "Fluorine effect on boron diffusion of P+ gate devices", IEDM Tech. Dig. pg. 447 (1989)

42. S.M. Hu J. Appl. Physics, 55 pg. 4095 (1984)

43. M. Morita et al, "Fluorine-Enhanced Thermal Oxidation of silicon in the presence of NF_3", Appl. Phys. Letters, 45(12) pg. 1312-1314 (1984)

44. M. Morita et al, "Low-temperature SiO_2 growth using fluorine-enhanced thermal oxidation", Appl. Phys. Letters, 47(3) pg. 253-255 (1985)

46. M. Morita et al, "A new SiO_2 Growth by fluorine-enhanced thermal oxidation", Proc. of IEDM pg. 144-147 (1984)

47. P.F. Schmidt et al, "Chemically enhanced thermal oxidation of silicon", Materials Letters pg. 235-238 (1985)

48. U.S. Kim and R.J. Jaccodine, "Fast shrinkage of oxidation staking faults during O_2 / NF_3 oxidation of silicon", Appl. Phys. Letters, 44(18) pg 1201-1203 (1986)

49. U.S. Kim, T. Kook and R.J. Jaccodine, "Diffusion of boron in silicon during O_2 / NF_3 oxidation", J. Electrochemical Society 135(1) pg. 270-271 (1988)

50. T. Sakurai and T. Sugano, "Theory of continuously distributed trap states at the $Si-SiO_2$ interfaces", J. Appl. Physics, 52(4) pg. 2889-2896 (1981)

51. R.T. Sanderson, Chemical Bonds and Bond Energy, Academic Press (1971)

52. J. P. Stagg, "Drift mobilities of Na^+ and K^+ Ions on SiO_2 films", Appl. Phys. Letters 31(8) pg. 532-533 (1977)

53. W.A. Harrison, Electronic Structure and the Properties of Solids (1980)

54. P.J. Wright, M. Wong, K.C. Saraswat, "The Effects of Fluorine on Gate Dielectric Properties", Proc. IEDM Tech. Dig. pg 574-577 (1987)

55. T. Ito, T. Nozaki, H. Ishikawa, "Direct Thermal Nitridation of Silicon Dioxide Films in Anhydrous Ammonia Gas", J. Electrochemical Society 127(9) pg. 2053-2057 (1980)

56. T. Kusaka, A. Hiraiwa and K. Mukai, "Mobility Degradation of Nitrided Oxide MISFETs", J. Electrochemical Society 135(1) pg. 166-172 (1988)

57. D.K. Shih et al, "MOS Characteristics of Rapid Thermal Nitrided Thin Oxides", Appl. Phys. Letters 52(20) pg. 1698-1700 (1988)

58. T. Hori, H. Iwasaki, k. Tsuji, "Charge-Trapping Properties of Ultrathin Nitreded Oxidex Prepared by Rapid Thermal Annealing", IEEE Trans. on Electron Devices, ED-35(7) pg. 904-910 (1988)

59. T. Hori, H. Iwasaki, Y. Naito, H. Esaki, "Electrical and Physical Characterization of Thin Nitrided Oxides Prepared by Rapid Thermal Nitridation", IEEE Trans. on Electron Devices, ED-34(11) pg 2238-2245 (1987)

60. M.A. Schmidt et al, "Inversion Layer Mobility of MOSFETs with Nitrided Oxide Gate Dielectrics", IEEE Trans. on Electron Devices, ED-35(10) pg 1627-1632 (1988)

61. P. Pan and C. Paquette, "Effect of Nitridation Pressure on Properties of Nitrided Oxide Films", Semiconductor Silicon (1986)

62. P. Wright Thesis, "The Physics and Technology of Gate Dielectric for ULSI" (1988)

63. T. Hori, H. Twasaki, "Ultra-Thin Re-Oxidized Nitrided Oxides Prepared by RTP". Proc. IEDM pg. 570-573 (1987)

64. A.E.T. Kuiper et al, "Hydrogenation during Thermal Nitridation of Silicon Dioxide," J. Appl. Physics 59(8) pg. 2765-2772 (1986)

65. B.E. Deal, "The current understanding of charges in the thermally oxidized silicon structure", J. Electrochemical Society 121(6) pg. 198C-205C (1974)

66. C.M. Dzuiba, M.S. Thesis, "Effects of presence of fluorine on aging characteristics of N-channel transistors with respect to hot electron degradation of the $Si - SiO_2$ interface (1991)

67. M. Yamis, IEEE Trans. Electron Devices, ED-12, pg. 88 (1965)

68. M. Kuhn and D.J. Sliversmith, J. electrochemical Society, 118 pg. 966 (1971)

69. N.J. Chou, J. Electrochemical Society, 118, pg. 601 (1971)

70. J.C. Lee et al, "Modeling and Characterization of Gate Oxide Reliability", IEEE Tran. Electron Devices, pg. 2268-2277 (1988)

71. I.C. Chen, S.Holland and C. Hu, "Electrical breakdown in thin gate oxide and tunneling oxides", IEEE Trans. Electron Devices, ED-32 pg. 413 (1985)

72. T.H. DiStefano and M. Shatzkes, "Impact ionization model for dielectric instability and breakdown", Appl. Physics Letters, vol 25 pg. 685 (1974)

73. E. Harari, "Dielectric breakdown in electrically stressed thin films of thermal SiO_2", J. Appl. Physics, vol 49 pg. 2478 (1978)

74. B. Ricco, M. Azbel and M. Brodsky, "Novel mechanism for tunneling and breakdown of thin SiO_2 films", Phys. Rev. Letter, vol 51 pg. 1795 (1983)

75. I.C. Chen, S. Holland and C. Hu, "Oxide breakdown dependence on thickness and hole current-enhanced reliability of ultra thin oxides", IEDM Tech. Dig. pg. 660 (1986)

76. H. Abe et al, "Analysis of defects in thin SiO_2 thermally grown on Si Substrate", IEDM Tech. Dig., pg. 372 (1985)

77. K. Honda, A.Ohawa and N. Toyokura, "Breakdown in silicon oxides correlation with Cu precipitates", Appl. Physics Letters, vol 45 pg. 270 (1984)

78. S.M. Sze, "VLSI Technology", pg. 599-638 (1983)

79. E.H. Nicollian and J.R. Brews, "MOS Physics and Technology", (1982)

80. A.S. Grove et al, "Investigation of thermally oxidized silicon surfaces using MOS structures", Solid State Electronics, vol 8 pg. 14-163 (1965)

81. F.D. Nkansah, "ECE 483 Lecture notes", (1992)

82. J.R. Brews, "Physics of the MOS transistor", Applied Solid-State Science series, Academic Press (1981)

83. T. Skotnicki, W. Marciniak, "New Approach to Threshold Voltage Modeling of MOSFETs", Solid-State Electronics-29, pg. 1115-1128 (1986)

84. T. Skotnicki et al,"The voltage doping transformation, A new approach to modeling Short-channel MOSFETs, Proc. ESSDERC pg. 543-546 (1987)

85. J. R. Brews, "Subthreshold behavior of uniformly and non-uniformly doped long-channel MOSFET", IEEE Trans. Electron Devices, ED-26 pg. 1282 (1979)

86. S.M. Sze, "Physics of Semiconductor Devices", pg. 46 (1981)

87. Y.W. Sing and B. Sudlow, "Modeling and VLSI design constraints of substrate current", IEDM Tech. Dig. pg. 732-735 (1980)

88. I. Matsunaga et al, "Characterization of two step impact ionization and its influence on NMOS and PMOS VLSI's", IEDM Tech. Dig. pg. 736 (1980)

89. H. Gesch et al,"Generation of interface states by hot-hole injection in MOSFETs", IEEE Trans. Electron Devices, ED-29 pg. 913 (1982)

90. E. Takeda and N. Suzuki, "An empirical model for device degradation due to hot-carrier injection", IEEE Electron Device Letters, EDL-4 pg. 111-113 (1983)

91. F.C. Hsu and S. Tam, "Relationship between MOSFET degradation and hot-electron induced interface-state generation", IEEE Electron Device Letters, EDL-5 pg. 50 (1984)

92. Y.A. El-mansy and D.M. Caughey, "Modeling weak Avalanche Multiplication Currents in IGFETs and SOS transistors for CAD", IEDM Tech. Dig. pg. 31 (1975)

93. T.H. Ning et al, "1 μm MOSFET VLSI technology: Hot electron design constrains", IEEE Trans. Electron Devices, ED-26 pg. 346-353 (1979)

94. P.K. Ko et al, "A unified model for hot electron currents in MOSFETs", IEDM Tech. Dig. pg. 600 (1981)

94. S. Tam, P.K. Ko and C. Hu, "Lucky-electron model of channel hot electron injection in MOSFETs", IEEE Trans. Electron Devices, pg. 1116-1125 (1984)

95. E. Takeda et al, "Device performance degradation due to hot-carrier injection at energies below the $Si-SiO_2$ energy barrier", IEDM Tech. Dig. (1983)

96. J. Lohstroh, E. Seevinck and J. De Groot, "Worst-Case Static Noise Margin Criteria for Logic Circuits and Their Mathematical Equivalence", IEEE J. Solid-State Circuits, SC-18(6) pg. 803-806 (1983)

97. J. Maserjian, "Tunneling in Thin MOS Structures", J. Vac. Sci. Technology, 11(6) pg. 996-1003 (1974)

98. Z.A. Weinberg, "Tunneling in Metal Oxide Silicon Structures", J. Appl. Physics, 53(7) pg. 5052-5056 (1982)

99. C. Chang, R.W. Brodersen and C. Hu, "Direct and Fowler-Nordheim Tunneling in thin gate oxide MOS structure", Elsevier Science Publishers pg. 176-180 (1983)

100. M. Lenzlinger and E.H. Snow, "Fowler-Nordheim Tunneling into thermally grown SiO_2", J. Appl. Physics 40(1) pg. 278-283 (1969)

101. N. Wang, "Digital MOS Integrated Circuits Design for Applications" (1991)

102. H. El Kamchouchi and A.A. Zaky, "A Direct Method for the calculation of the Edge Capacitance", J. Appl. Physics, 8 pg. 1365-1371 (1975)

103. E.W. Greeneich, "An Analytical model for the Gate Capacitance of Small-Geometry MOS structures", IEEE Trans. Electron Devices, ED-30 pg. 1838-1839 (1983)

104. N. Weste and K. Eshraghian, "Principles of CMOS VLSI Design" (1985)

105. C.G. Sodini, P.K. Ko and J. L. Moll, "The effect of high field on MOS device and circuit performance devices", IEEE Trans. Electron Devices, ED-31 pg. 1386-1393 (1984)

106. F. H. Hsu and K. Y. Chiu, "A comparative study of tunneling, substrate and hot-electron and channel hot-electron injection induced degradation in thin-gate MOSFETs, IEDM Tech. Dig. pg. 96 (1984)

107. M. Koyanagi et al, "Optimum design of N+/N- double diffused drain MOSFET to reduce hot-carrier emission", IEEE Trans. Electron Devices, ED-32 pg. 562 (1985)

VITA

Franklin Daniel Nkansah was born in Larteh Ghana, West Africa. He enrolled at Kutztown University in Kutztown, Pennsylvania in the Fall of 1985 and received his Bachelor of Science Degree in Physics in the Fall of 1988. He was employed at Dunnington Company in the Spring of 1989, following which he joined AT&T Bell Laboratories in Summer of 1989, where he is currently working as a Technologist in the Advanced CMOS Technology Development Group. He began part-time graduate studies at Lehigh University in Bethlehem, Pennsylvania in the Fall of 1990 and is expected to graduate with a Master of Science Degree in Electrical Engineering in the Spring of 1993.

END

OF

TITLE