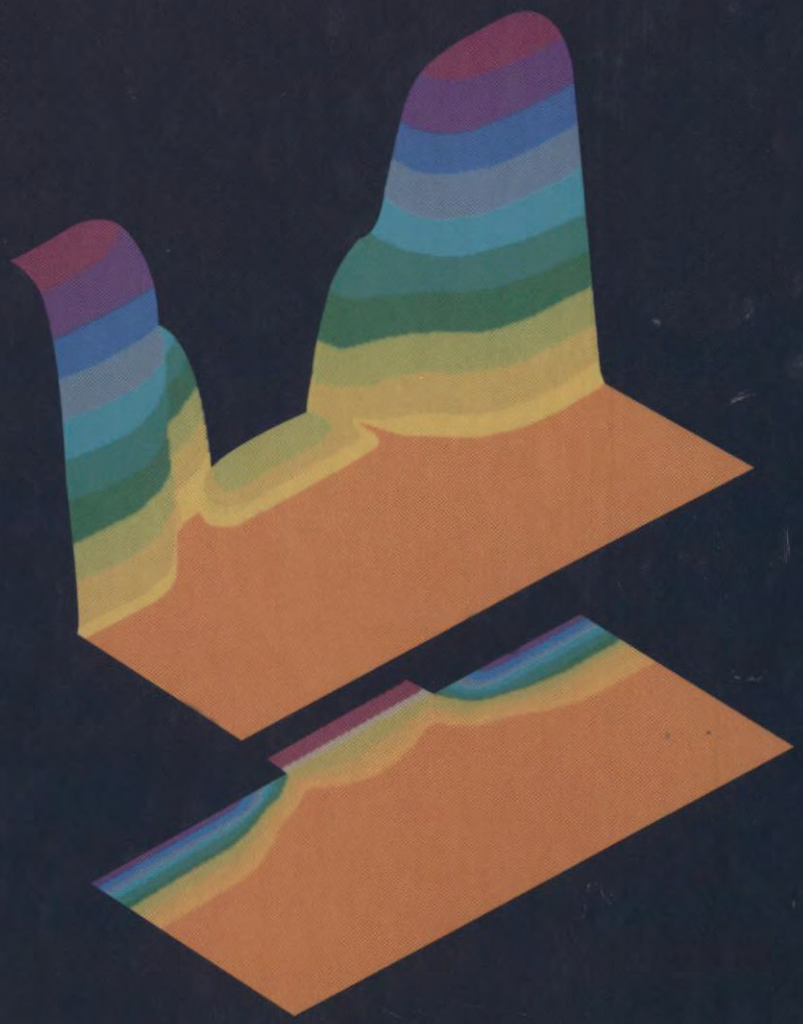
Silicon Processing

for the VLSI Era

Volume 3 - The Submicron MOSFET



S. Wolf

SILICON PROCESSING

FOR

THE VLSI ERA

VOLUME 3:

THE SUBMICRON MOSFET

STANLEY WOLF Ph.D.
Professor, Department of Electrical Engineering
California State University, Long Beach
Long Beach, California

LATTICE PRESS

Sunset Beach, California

DISCLAIMER

This publication is based on sources and information believed to be reliable, but the authors and Lattice Press disclaim any warranty or liability based on or relating to the contents of this publication.

Published by:

Lattice Press,
Post Office Box 340
Sunset Beach, California 90742, U.S.A.

Cover design by Roy Montibon, Visionary Art Resources, Inc., Los Angeles, CA.

Copyright © 1995 by Lattice Press.
All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system without written permission from the publisher, except for the inclusion of brief quotations in a review.

Library of Congress Cataloging in Publication Data
Wolf, Stanley

Silicon Processing for the VLSI Era
Volume 3 : The Submicron MOSFET

Includes Index
1. Integrated circuits-Very large scale integration. 2. Silicon. I. Title

ISBN 0-961672-5-3

9 8 7 6 5 4 3 2

PRINTED IN THE UNITED STATES OF AMERICA

Engn
TK
7874
W635
1986
v.3

CONTENTS

PREFACE

Chap. 1 - THE ROLE OF PROCESS AND DEVICE MODELS IN MICROELECTRONICS TECHNOLOGY 1

1.1 MODELS IN MICROELECTRONIC TECHNOLOGY 1

1.1.1 Development of Physically Based Models 3
1.1.2 Development of Empirically Based Models 4
1.1.2.1 Empirically Based versus Physically Based Models.
1.1.2.2 Deriving Empirically Based Models.
1.1.2.3 Role of Empirical Models in Semiconductor Simulation.
1.1.2.4 Examples of Quantitative Empirically Based Models.
1.1.2.5 PREDICT-1: A Process Simulator Based on Empirical Models.

1.2 MATHEMATICAL FORMULATION OF THE PHYSICALLY-BASED MODELS OF SEMICONDUCTOR PROCESSES AND DEVICES 9

1.2.1 The Physically Based Model of Impurity Diffusion in Silicon - Fick's Second Law Equation 11
1.2.2 The Physically Based Model of Semiconductor Device Behavior 12
1.2.3 The Poisson Equation 15

1.3 CLOSED-FORM ANALYTICAL SOLUTIONS VERSUS NUMERICAL SOLUTIONS OF THE PDES THAT ARE THE PHYSICALLY BASED MODELS OF MICROELECTRONICS 16

1.3.1 Closed-Form Analytical Solutions - Mathematical Aspects 16
1.3.2 Numerical Solutions - Mathematical Aspects 16
1.3.3 Comparing the Analytical and Numerical Solutions in Semiconductor Simulation Applications 17
1.3.4 Role of Analytical Solutions in Submicron Semiconductor Technology 17
1.3.5 Analytic Solutions of Fick's 2nd Law Equation 19

1.3.6 Analytical Solutions of the One-Dimensional Form of the Poisson Equation in the Depletion Approximation 23

1.4 CIRCUMSTANCES UNDER WHICH THE ANALYTIC SOLUTIONS OF THESE PDES BECOME INADEQUATE FOR SIMULATION PURPOSES 24

1.4.1 The Inadequacy of the Analytic Solutions of Fick's 2nd Law Equation for Simulating Diffusion in Submicron Device Structures 24
1.4.2 The Inadequacy of Using a One-Dimensional Form of the Poisson Equation and the Depletion Approximation to Model Submicron MOSFET Behavior 27

REFERENCES 28

Chap. 2 - NUMERICAL METHODS FOR SOLVING THE PARTIAL DIFFERENTIAL EQUATIONS WHICH MODEL THE PHYSICS OF SUBMICRON DEVICES AND PROCESSES 29

2.1 NUMERICAL SOLUTIONS OF DIFFERENTIAL EQUATIONS 31

2.2 NUMERICAL METHODS FOR SOLVING NONLINEAR PARTIAL DIFFERENTIAL EQUATIONS 32

2.3 GRID GENERATION AND REFINEMENT PROCEDURES 34

2.4 DISCRETIZATION OF DIFFERENTIAL EQUATIONS 36

2.4.1 Space Discretization 39
2.4.2. The Finite-Difference Method 39
2.4.2.1 The Difference Operator Method
2.4.2.2 Taylor Series.
2.4.2.3 One-Dimensional Finite Difference Approximations of Derivatives.
2.4.2.4 Two-Dimensional Space Discretization.
2.4.3 The Finite Element Method 45
2.4.4 Time Discretization 47
2.4.4.1 Explicit Time Discretization.
2.4.4.2 Implicit Time Discretization.

2.5 ORDERING THE SET OF DISCRETIZED PDEs INTO APPROPRIATE MATRIX FORM 50

2.6 LINEARIZING NONLINEAR ALGEBRAIC EQUATIONS 52

2.7 MATRIX METHODS FOR SOLVING SIMULTANEOUS LINEAR ALGEBRAIC EQUATIONS 55

2.8 VISUALIZATION OF THE RESULTS 57

2.9 DATA BASE OF THE SIMULATION PACKAGES 58

2.10 EXAMPLES OF THE PROCEDURES USED TO NUMERICALLY SOLVE THE MODEL PARTIAL-DIFFERENTIAL EQUATIONS OF MICROELECTRONICS 59

2.10.1 Outline of a Numerical Analysis Procedure for Solving the Two-Dimensional Poisson Equation to Simulate the Electrostatic Potential in a MOSFET 59

2.10.1.1 Example of How Device Simulation Programs are Used to Obtain the Two-Dimensional Solution of the Poisson Equation in a MOSFET Structure Operating in Subthreshold.

2.10.2 Example of a Numerical Procedure for Solving the One-Dimensional Form of Fick's 2nd Law Equation as a Model for Impurity Diffusion in Silicon 70

2.11 SYSTEMS OF COUPLED EQUATIONS 74

2.11.1 Gummel's Iterative Method 75
2.11.2 Newton's Direct Method 76

2.12 EVOLUTION OF MOSFET DEVICE SIMULATORS 77

2.12.1 CADDET 78
2.12.2 MINIMOS 78
2.12.3 GEMINI 78
2.12.4 PISCES-II 80
2.12.5 Three-Dimensional Device Simulators 80

REFERENCES 81

Chap. 3 - MOS TRANSISTOR DEVICE PHYSICS: PART 1 - BASICS MOS PHYSICS and THE MOS CAPACITOR **83**

3.1 BASIC MOS THEORY and PHYSICS of IDEAL MOS CAPACITORS (MOS-C) 84

3.1.1 The Structure of the MOS-C and the MOSFET 85

3.1.1.1 Comparison of MOS-Cs and Conventional Capacitor Structures.

3.1.1.2 Why is the MOS-C Studied First?

3.2 QUALITATIVE DESCRIPTION OF MOS-C DEVICE BEHAVIOR 89

3.2.1 Energy-Band Diagram Representation of Semiconductor Devices 90
3.2.2 Energy Band Diagram of the Ideal MOS Capacitor 91
3.2.2.1 Accumulation.
3.2.2.2 Depletion.
3.2.2.3 Inversion.
3.2.2.4 Strong Inversion.
3.2.2.5 Microscopic Perspective of the Mobile Charge Behavior in a MOS-C.

3.3 QUANTITATIVE DESCRIPTION OF MOS-C BEHAVIOR 96

3.3.1 The Electrostatic Potential in MOS Structures 98
3.3.2 General Approach to Deriving MOS-C Characteristics 100
3.3.3 Finding the Quantitative Characteristics of a MOS-C when the Depletion Approximation is Invoked 103
3.3.4 More Accurate and General Model for MOS-C Behavior 106
3.3.5 Charge-Sheet Model for Determining the Characteristics of the Ideal MOS-C 112
3.3.5.1 Approximation of Q_I in Weak Inversion.
3.3.5.2 Approximation of Q_I in Strong Inversion.

3.4 THRESHOLD VOLTAGE OF IDEAL MOS CAPACITORS 116

3.5 PHYSICS OF NON-IDEAL MOS CAPACITORS (n^+PolySi-SiO_2-Si MOS Capacitors) 117

3.5.1 Workfunction Difference Between Gate and Semiconductor 117
3.5.2 Oxide Charges and Traps 119
3.5.2.1 Mobile Ionic Charges.
3.5.2.2 Fixed Oxide Charge.
3.5.2.3 Interface Trapped Charge.
3.5.2.4 Oxide Trapped Charge.
3.5.2.5 Combined Effect of Oxide Charges on V_T of Submicron MOSFETs.

3.6 THE THRESHOLD VOLTAGE EQUATION OF Si-SiO_2 MOS CAPACITORS 127

3.7 VISUALIZATION OF MOS-C AND MOSFET PHENOMENA USING POTENTIAL VERSUS POSITION DIAGRAMS 129

REFERENCES 130

Chap. 4 - MOS TRANSISTOR DEVICE PHYSICS: PART 2 - LONG-CHANNEL MOSFETS 134

4.1 TERMINOLOGY OF MOSFET DEVICE STRUCTURES 135

4.1.1 Basics of MOSFET Operation 136

4.2 BIASING THE INVERSION LAYER (BODY EFFECT) 138

4.2.1 Qualitative Discussion of the Effect of Applying a Bias to the Inversion Layer of a MOS Structure 138

4.2.1.1 Case When Both V_{GB} and $V_{CB} = 0V$.

4.2.1.2. Case When $V_{GB} \geq V_T$ and $V_{CB} = 0V$.

4.2.1.3. Case When $V_{GB} = 0V$, But $V_{CB} > 0V$.

4.2.1.4. Case When $V_{GB} \geq V_T$ and $V_{CB} > 0V$.

4.2.1.5 Summary of the Behavior of the Three-Terminal MOS Structure and the Significance of Such Behavior on the Operation of the MOSFET.

4.2.1.6 Body Effect.

4.2.2 Quantitative Description of Effects of Biasing the Inversion Layer of MOS Structures 144

4.2.2.1 General Case.

4.2.2.2 Strong Inversion Case.

4.3 DEFINITION OF THE TERM: LONG-CHANNEL MOSFET 147

4.3.1 The Gradual Channel Approximation (GCA) 149

4.3.1.1 Verifying the Validity of the GCA as a Function of the Geometrical Factors of the MOSFET Device Structure and Bias Conditions.

4.3.2 Circuit Characteristics of Long-Channel MOSFETs 152

4.4 QUALITATIVE DESCRIPTION OF MOSFET OPERATION 153

4.4.1 MOSFET Behavior in Response to Gate Bias Alone 153

4.4.2 Qualitative MOSFET Behavior when both Gate and Drain Bias are Applied 158

4.5 QUANTITATIVE BEHAVIOR OF MOSFETS: DC CIRCUIT MODELS OF LONG-CHANNEL MOSFETS 162

4.5.1 General Approach to Deriving the DC Drain-Current Relationships in Long-Channel MOSFETs 163

4.5.2 Pao-Sah Model 166

4.5.3 The Charge Sheet Model 170

4.5.4 Variable Depletion-Charge (or Bulk-Charge) Model 173

4.5.5 Square-Law Model 175
4.5.6 Comparison of the Four MOSFET Circuit Models 177
4.5.6.1 MOSFET Models of the SPICE Circuit Simulator.

4.6 THRESHOLD-VOLTAGE CONTROL IN MOSFETS 182

4.6.1 Measuring Threshold Voltage 182
4.6.2 Adjusting Threshold Voltage 182
4.6.3 Ion Implantation for Adjusting Threshold Voltage 183
4.6.3.1 Impact of the V_T-adjust Implant on the Body Effect.
4.6.4 Impact of V_T-Adjust Implant on Other MOSFET Characteristics 188

4.7 SUBTHRESHOLD CURRENTS IN LONG-CHANNEL MOSFETs (I_{Dst} when $V_{GS} < |V_T|$) 191

4.7.1 Subthreshold Swing S_t 196
4.7.2 Gate-Induced Drain Leakage (GIDL) 198

4.8 SUMMARY OF LONG-CHANNEL MOSFET BEHAVIOR 200

REFERENCES 201

Chap. 5 - MOS TRANSISTOR DEVICE PHYSICS: PART 3 - THE SUBMICRON MOSFET **205**

5.1 DEVICE CHARACTERISTICS OF LONG-CHANNEL VERSUS SHORT-CHANNEL MOSFETS 207

5.2 EFFECT OF GATE DIMENSIONS ON THRESHOLD VOLTAGE 208

5.2.1 Short-Channel Threshold-Voltage Shift 208
5.2.1.1 Numerical Solution of the 2-D Poisson Equation for Calculating ΔV_{Tsc}.
5.2.1.2 Charge-Sharing Models.
5.2.1.3 Barrier-Lowering Models and DIBL.
5.2.1.4 Quasi-Two-Dimensional Analytical Solution for ΔV_{Tsc}.
5.2.1.5 Measuring Surface-DIBL Effects.
5.2.1.6 Dependence of V_T on V_{BS} in Short-Channel MOSFETs.
5.2.2 Narrow Gate-Width Effects on Threshold Voltage 222
5.2.2.1 Devices Exhibiting Both Short- and Narrow-Width Effects.

5.2.3 Reverse Short-Channel Effects on Threshold Voltage 226

5.2.3.1 Model which Proposes RSCE is Due to Lateral Dopant Nonuniformity at the Channel Si-SiO_2 Interface (Arising from Enhanced Diffusion of Channel Dopants Caused by Interstitial Injection During Poly Reox, Salicide Formation, or Implant Damage).

5.2.3.2 Model Which Proposes RSCE Arises from Boron Segregation to Implant Damaged Regions at the Edge of the Channel.

5.2.3.3 Model Which Attributes RSCE to Transient Enhanced Diffusion of Channel Dopants to the Silicon Surface Arising from Implant Damage.

5.2.3.4 Model which Attributes RSCE to Silicon Interstitial Capture in the Gate Oxide.

5.3 SHORT-CHANNEL EFFECTS ON SUBTHRESHOLD SWING, S_t (Subsurface Punchthrough) 232

5.3.1 Experimentally Characterizing Punchthrough 240

5.3.2 Modeling Punchthrough 242

5.3.3 Modeling the Simultaneous Occurrence of Surface-DIBL and Subsurface-DIBL 245

5.4 SHORT-CHANNEL EFFECTS ON THE I-V CHARACTERISTICS OF MOSFETS OPERATED IN THE STRONG THRESHOLD REGIME 247

5.4.1 Impact of the MOSFET Electric Fields on Mobility 248

5.4.1.1 Models of Mobility Degradation Due to Gate-Induced Electric Fields.

5.4.1.2 Models of the Carrier-Velocity Dependence on Drain Electric Field (Velocity Saturation).

5.4.2 A Simple, Analytic DC Circuit Model of the Short-Channel MOSFET in Strong Inversion 254

5.4.2.1 DC Circuit Model of the Short-Channel MOSFET in the Linear Regime.

5.4.2.2 Effects of Scaling on g_{msat}.

5.4.2.3 Comparing the Values of I_D Calculated with the Long- and Short-Channel MOSFET Models as a Function of Channel Length.

5.4.3 Modeling the Saturation Region of Operation in Short-Channel MOSFETs 260

5.4.3.1 Modeling Short-Channel MOSFET Saturation Behavior Using the Abrupt Junction and Depletion Approximations.

5.4.3.2 Constant Field-Gradient Model.

5.4.3.3 Pseudo-Two-Dimensional Model.

5.5. SHORT-CHANNEL MOSFET MODELS USED IN THE *SPICE* CIRCUIT SIMULATOR 271

5.5.1 BSIM Model Equations 272
5.5.1.1 BSIM Threshold Voltage Equation.
5.5.1.2 BSIM Equations for I_D when the MOSFET is Operated in Strong Inversion.
5.5.1.3 BSIM Equations for I_D when the MOSFET is Operated in Weak Inversion (Subthreshold Regime).

5.6 MOSFET SCALING 274

5.6.1 Design of Submicron MOSFETs: Scaling Guidelines 274
5.6.1.1 Constant Electric-Field, Constant Voltage & Electrostatic Scaling.
5.6.1.2 Subthreshold Scaling.
5.6.1.3 Off-Current Scaling Approach.
5.6.2 Manufacturing Considerations in Submicron MOSFET Device Design 285

5.7 DESIGN OF SUBMICRON MOSFETS 286

5.7.1 Design Methodology Based on the Subthreshold Scaling Equation 286
5.7.2 Design Methodology Based on Required Off-Current Value 287

5.8 *p*-CHANNEL DEVICES IN CMOS 288

5.8.1 PMOS Devices with n^+-Polysilicon Gates 289
5.8.1.1 Punchthrough Susceptibility.
5.8.2 Why PMOS Devices with n^+-Polysilicon Gates and Implanted *p*-Surface Layers Exhibit Enhanced Susceptibility to Short-Channel Effects 291
5.8.2.1 Threshold-Voltage Roll-Off in Buried-Channel PMOS.
5.8.2.2 Subthreshold Swing (S_t) in Buried-Channel PMOS.
5.8.2.3 Punchthrough Susceptibility of Buried-Channel PMOS.
5.8.3 Techniques Used to Suppress Short-Channel Effects in Buried-Channel PMOS 305
5.8.3.1 Techniques Used to Keep the Implanted p-Layer as Thin as Possible.
5.8.3.2 Shallow p^+ Source/Drain Junctions to Suppress Punchthrough in PMOSFETs.
5.8.3.3 LDD Structures in PMOS to Limit Punchthrough.
5.8.3.4 New PMOS Device Structures Designed to Suppress Punchthrough.
5.8.4 PMOS Devices with p^+--Polysilicon Gates 311

REFERENCES 313

Chap. 6 - ISOLATION STRUCTURES IN SUBMICRON CMOS 321

6.1 CHARACTERIZING MOS ISOLATION

6.1.1 Test Structures for NMOS Isolation Structures 322
6.1.2 Punchthrough Prevention between Adjacent Devices in MOS Circuits 325

6.2 THE BASIC LOCOS ISOLATION PROCESS 330

6.2.1 Details of the Basic Semi-recessed LOCOS Process 332
6.2.1.1 Pad-Oxide Layer.
6.2.1.2 CVD of Silicon Nitride Layer.
6.2.1.3 Mask and Etch Pad-Oxide/Nitride Layer to Define Active Regions.
6.2.1.4 Channel-Stop Implant.
6.2.1.5 Problems Arising from the Channel-Stop Implants.
6.2.1.6 Grow Field Oxide.
6.2.1.7. Factors Which Impact Bird's Beak Length and Shape.
6.2.1.8 Use of High-Energy Ion Implantation to Introduce the Channel-Stop Dopant After Field Oxide Growth.
6.2.1.9 Strip the Masking Nitride/Pad-Oxide Layer.
6.2.1.10 Regrow Sacrificial Pad Oxide and Strip (Kooi Effect).
6.2.2. Planarizing Semirecessed LOCOS Structures 348

6.3 MODELING THE TOPOGRAPHY OF LOCOS STRUCTURES 349

6.3.1 Empirical Models of 2-D Thermal Oxidation 350
6.3.2 Physically Based Models of 2-D Oxidation 351
6.3.3 Simulating Thermal Oxidation in 3-D Using Physical Models 355

6.4 POLY-BUFFERED LOCOS 357

6.4.1. Modified Poly-Buffered LOCOS Processes 361
6.4.1.1 Framed-Mask PBL (FMPBL)
6.4.1.2 Polysilicon-Encapsulated LOCOS (PELOX)
6.4.1.3 Nitride-Clad LOCOS (NCL)
6.4.1.4 Mitigating the Field-Thinning Problem of LOCOS

6.5 SHALLOW TRENCH AND REFILL ISOLATION 367

6.5.1 Early Shallow Trench and Refill Isolation 368
6.5.2 Improved Shallow Trench and Refill Isolation Processes 371

6.6 CMOS ISOLATION TECHNOLOGY 373

6.6.1 Qualitative Description of Isolation in CMOS 374

6.6.1.1 Leakage Currents in CMOS Gates Due to Active Device Phenomena and Intra-Well Isolation Loss

6.6.1.2 Leakage Currents in CMOS Gates Due to Loss of Isolation Among Opposite-Type FETs

6.6.1.3 Why n^+-to-p^+ Isolation Spacing is Greater than n^+-to-n^+ Spacing in NMOS (or p^+-to-p^+ Spacing in PMOS)

6.6.1.4 Effects on n^+-to-p^+ Spacing of Dopant Redistribution at the Well Borders

6.6.1.5 Threshold Voltage Rolloff, DIBL, and Lateral Punchthrough in Submicron CMOS Parasitic Field FETs

6.6.2 Quantitative Modeling of CMOS Isolation 384

6.6.3 Experimental Characterization of Isolation in CMOS 387

6.6.3.1 Experimental Characterization of Isolation in n^+-to-n-well Parasitic FETs.

6.6.3.2 Experimental Characterization of Isolation in p^+-to-p-substrate Parasitic FETs.

6.6.4 Establishing the Minimum n^+-to-p^+ Isolation Spacing in CMOS from Modeling and Experimental Data 398

6.6.4.1 Isolation Spacing Design Rules for 0.35 and 0.25 μm CMOS.

6.6.5 Vertical Isolation in CMOS 402

6.6.6 LOCOS-Based Isolation in CMOS 405

6.6.7 Trench Isolation for CMOS 407

REFERENCES 414

Chap. 7 - THIN GATE OXIDES: GROWTH AND RELIABILITY **421**

7.1 GATE OXIDE CHARACTERISTICS NEEDED FOR SUBMICRON MOSFETS 422

7.2 PHYSICAL PICTURE OF SiO_2 AND THE Si/SiO_2 INTERFACE 423

7.2.1 Physical and Chemical Properties of SiO_2 423

7.2.2 The Si/SiO_2 Interface 425

7.3 CARRIER INJECTION IN THE Si/SiO_2 SYSTEM 429

7.3.1 Channel Hot Electrons (CHE) 430

7.3.2 Injection of Drain Avalanche Hot Carriers (DAHC) 434

7.3.3 Fowler-Nordheim (FN) Tunneling 435

7.3.4 Direct Tunneling 437

7.4 TESTS FOR GATE OXIDE RELIABILITY 438

7.4.1 Measuring Dielectric Strength 438
7.4.2 Measuring Time-Dependent Breakdown Behavior 440
7.4.2.1 Time-to-Breakdown under Constant-Voltage Stressing.
7.4.2.2 Time-to-Breakdown under Constant-Current Stressing.
7.4.2.3 Charge to Breakdown (Q_{BD}).
7.4.3 Mathematics of Reliability Characteristics 441
7.4.3.1 The Failure Unit (FIT).
7.4.3.2. The Cumulative Distribution Function.
7.4.3.3 Probability Density Function.
7.4.3.4 The Instantaneous Failure Rate.
7.4.3.5 The Mean Time to Failure.
7.4.3.6 The Weibull Cumulative Distribution Function.
7.4.3.7 The Log-Normal Cumulative Distribution Function.
7.4.3.8 Characterizing the Reliability of Oxides in Terms of the Mathematical Reliability Functions.

7.5 THE PHENOMENON OF OXIDE BREAKDOWN 448

7.5.1 Qualitative Physical Models of Intrinsic Oxide Breakdown 451
7.5.1.1 Hole Generation and Trapping Model.
7.5.1.2 Wolters' Electron Lattice-Damage Model.
7.5.2 Qualitative Models for the Physical Origin of B-Mode Failures in Oxides 458
7.5.2.1 Test Structure for Intrinsic Oxide Breakdown Measurements.
7.5.2.2 Damage at the Gate Periphery Resulting from the Polysilicon Etch Process.
7.5.3 Empirical Quantitative Model of Intrinsic Oxide Breakdown 462
7.5.4 Quantitative Models for B-Mode Failures 467
7.5.4.1 Reduced Barrier Height Model.
7.5.4.2 Oxide Thinning Model.
7.5.4.3. The Cumulative Defect Density Function $D(\Delta t_{ox})$.
7.5.4.4. Using $D(\Delta t_{ox})$ to Predict Oxide Breakdown Reliability.
7.5.4.5. Ramp-Voltage Breakdown Test for Obtaining $D(\Delta t_{ox})$ Data.
7.5.4.6. Correlating Ramp-Voltage Test Data and Constant-Voltage TDDB Data.
7.5.4.7 Interface Trap Generation as a Measure of Oxide Damage Due to Charge Injected into the Oxide.
7.5.5 Applications of the Oxide Breakdown Reliability Model 477
7.5.5.1 Projecting the Reliability of Gate Oxides with Various Thicknesses.

7.5.5.2 If a MOS IC is to have a Specific Lifetime (e.g., 10-Years), What is the Minimum Oxide Thickness that can be Used? (Assume the Maximum Cumulative Failure Percentage, Operating Temperature, and Oxide Area are Specified.)

7.6 SCREENING TESTS FOR OXIDE BREAKDOWN (BURN-IN) 482

7.6.1 Modeling the Effect of the Burn-In Stress on the Post-Burn-In Reliability of the Oxides 485
7.6.2 Procedure for Selecting an Optimal Set of Burn-In Conditions for Oxide Breakdown Screening 486
7.6.3 Example of Applying the Procedure for Determining an Optimal Set of Burn-In Conditions 489

7.7 OXIDE BREAKDOWN MODULE IN BERKELEY RELIABILITY TOOL (BERT) 492

7.8 MODELS OF THIN OXIDE GROWTH 492

7.9 TECHNOLOGY OF THIN OXIDE GROWTH 495

7.9.1 Oxidation Furnaces 495
7.9.2 Control of Gate Oxide Growth Rates 495
7.9.3 Factors Impacting Gate-Oxide Thickness Uniformity 496
7.9.4 Gate Oxide Growth Processes for 10-20 nm thick Gate Oxides on 150-200 mm Wafers 499
7.9.5 Alternative Gate Oxide Growth Processes 502
7.9.5.1 Stacked Gate Oxide Films (Thermally Grown Layer and CVD Layer).

7.10 GATE OXIDE DAMAGE FROM PLASMA PROCESSING 504

7.10.1 How Isolated Conductor Regions on a Wafer Surface Can be Charged Up by an RF Plasma 504
7.10.2 Causes of Non-Uniformities in Plasmas 506
7.10.3 Test Structures used to Characterize Plasma-Induced Oxide Damage 507
7.10.4 Electrical Tests for Measuring Oxide Damage 509
7.10.4.1 Wafer Mapping to Determine Positional Dependence of Damage on the Wafer Surface.
7.10.5 Oxide Damage Models in Various Plasma Etching Processes 511
7.10.5.1 Oxide Damage During Polysilicon or Metal Etch Processes.
7.10.5.2 Oxide Damage During Resist Stripping.
7.10.5.3 Radiation Induced Oxide Damage.
7.10.6 Procedures for Reducing Plasma-Induced Oxide Damage 514

REFERENCES 515

Chap. 8 - WELL FORMATION IN CMOS 523

8.1 *p*-WELL, *n*-WELL, and TWIN-WELL TECHNOLOGIES 524

8.1.1 *p*-Well CMOS 525
8.1.2 *n*-Well CMOS 527
8.1.3 CMOS on Epitaxial Substrates 528
8.1.4 Twin-Well CMOS 530
8.1.5 Summary of CMOS Well-Technology Issues 532

8.2 CONVENTIONAL FORMATION OF WELLS AND CHANNEL STOPS 532

8.2.1 Aligning Subsequent Patterns to the Wells 532
8.2.2 Twin-Well and Channel-Stop Formation 535
8.2.3 Split Well-Drive-In Process 538

8.3 RETROGRADE-WELL CMOS 539

8.3.1 Retrograde Well Technologies to Improve Isolation 546

8.4 FUTURE TRENDS IN WELL FORMATION 552

8.5 AFTER-GATE IMPLANTATION (AGI) 554

REFERENCES 555

Chap. 9 - HOT-CARRIER RESISTANT PROCESSING AND DEVICE STRUCTURES 559

9.1 MODEL FOR THE MAXIMUM LATERAL ELECTRIC FIELD $\mathcal{E}_{ymax}$ IN NON-GRADED-DRAIN MOSFETS 561

9.2 SUBSTRATE CURRENTS DUE TO HOT CARRIERS 563

9.3 MODELS OF HOT-CARRIER DEGRADATION PHENOMENA 566

9.3.1. Si-H Bond Breaking Model 569
9.3.2 Hot-Hole and Hot-Electron Trapping Model 569
9.3.3 Location of the Hot-Carrier Damaged Region 571

9.4 CHARACTERIZING MOSFET DEGRADATION DUE TO HOT-CARRIER EFFECTS 572

9.4.1 Device Parameters that are Impacted by Hot-Carrier Damage 573

9.4.2 How to Determine if a MOSFET Can be Operated with out Failing due to Hot-Carrier Effects Under Normal Operating Conditions? 574

9.4.2.1 Selecting a Criterion that Identifies Device Failure Due to Hot-Carrier Damage.

9.4.3 Accelerated Testing of Hot-Carrier Device Lifetime 576

9.4.4 Hot-Carrier Lifetime Under AC Stressing 580

9.4.5 Effect of Device Degradation on Circuit Performance 581

9.4.6 Hot-Carrier Reliability Simulation 582

9.4.6.1 SCALE.

9.4.6.2 Simulators that Relate the Device-Level Degradation Parameters to Circuit Performance (CAS and HOTRON).

9.5 DECREASING HOT CARRIER DEGRADATION BY MODIFYING THE MOSFET STRUCTURE 586

9.5.1 Phosphorus-Drain Structure 588

9.5.2 Double Diffused Drain (DDD) 588

9.6 LIGHTLY DOPED DRAIN (LDD) STRUCTURES 591

9.6.1 Modeling $\mathcal{E}_{ymax}$ in LDD MOSFETs 592

9.6.1.1 Simple Model for Estimating $\mathcal{E}_{ymax}$ in LDD MOSFETs

9.6.1.2 Analytical Model for Calculating $\mathcal{E}_{ymax}$ in LDD MOSFETs Using the Pseudo-2D Approach to Solving the Poisson Equation.

9.6.1.3 Model for the Series Resistance Introduced by the n^- Region of the LDD.

9.6.2 Overview of the Process Sequence to Form LDDs 595

9.7 PERFORMANCE AND HOT-CARRIER LIMITATIONS OF THE CONVENTIONAL LDD 598

9.7.1 Performance Degradation in LDD MOSFETs 599

9.7.1.1 I_D Decrease Due to Added Resistance of n^- Regions.

9.7.1.2 Doping Compensation Effects that can Increase R_{n^-} in Deep Submicron LDD MOSFETs.

9.7.1.3 Increase of R_{sh} by Weak Overlap of the Gate and Drain.

9.7.1.4 The Impact of Poly Re-Ox on Weak Overlap in MOSFETs.

9.7.1.5 Off-Axis Implant Effects in MOSFETs, Leading to Weak Overlap Problems and Asymmetrical MOSFET Device Characteristics.

9.7.1.6 Implanter Issues that Impact the LDD Process Sequence.

9.7.1.7 Process integration issues which arise when an LDD is integrated into a CMOS Technology.

9.7.2 Hot Carrier Reliability Issues in Early LDDs 606

9.8 DISPOSABLE SPACERS 608

9.9 SUMMARY OF THE LIMITATIONS OF EARLY LDD DEVICES 611

9.10 SUMMARY OF THE METHODS TO BE USED IN DESIGNING EFFECTIVE LDD STRUCTURES 612

9.11 IMPROVED CONVENTIONAL (OXIDE-SPACER-BASED) LDD STRUCTURES 613

9.11.1 Increasing the Doping in the n^- Region of the LDD 614
9.11.1.1 Moderate Lightly Doped Drain (M-LDD).
9.11.1.2 Metal-coated LDD (MLD).
9.11.2 Tailoring the Vertical Doping Profile of the n^- Implant to Improve the Conventional LDD 616
9.11.2.1 Buried-LDD (B-LDD).
9.11.2.2 Profiled-LDD (P-LDD).
9.11.3 Improving the Short-Channel Performance of the Conventional NMOS LDD 621

9.12 FULLY OVERLAPPED LDD STRUCTURES 623

9.12.1 Inverse-T Gate LDD (ITLDD) and Gate-Drain Overlapped Device (GOLD) Structures 623
9.12.1.1 Inverse-T Gate LDD Structure.
9.12.1.2 Gate-Drain Overlapped LDD (GOLD).
9.12.1.3 Modified IT-LDD Structures.
9.12.2 Large-Angle-Tilt Implanted Drain (LATID) 630
9.12.3 High Dielectric Spacers 634
9.12.4 Asymmetrical MOSFET Structure: Halo-Source-GOLD (HS-GOLD) 636

9.13 HOT-CARRIER EFFECTS IN PMOS TRANSISTORS 637

9.13.1 Model for Hot-Carrier Damage in PMOSFETs 638
9.13.2 Hot-Electron Induced Punchthrough (HEIP) 638
9.13.3 Hot-Electron Induced Degradation in p^+-Poly PMOSFETs 641

9.14 STRENGTHENING OXIDES AGAINST HOT-CARRIER INJECTION 641

9.14.1 Minimizing Hydrogen and H_2O at the Si-SiO_2 Interface 643
9.14.1.1. The role of hydrogen in hot-carrier degradation.

9.14.1.2 Hydrogen introduced at high temperatures.
9.14.1.3 Moisture in Gate Oxides as a Contributor to Hot-Carrier Degradation.
9.14.2 Gate-Oxide Damage by Metal Etching 647
9.14.3 Nitrided Oxides 648
9.14.3.1 Thermal nitridation of SiO_2.
9.14.3.2 NH_3 Oxides as MOSFET Gate Dielectrics.
9.14.3.3 Low-Pressure and RTP NH_3 Oxides (Lightly Nitrided Reoxidized NH_3-Nitrided Oxides).
9.14.3.4 Oxynitridation of Silicon in N_2O Ambients.
9.14.3.5. Oxynitridation of Oxides in N_2O.
9.14.3.6 Fluorinated Oxides.

REFERENCES 661

***APPENDIX A* - MATRIX METHODS** 675

A.1 INTRODUCTION TO MATRICES 675
A.2 MATRIX MULTIPLICATION 677
A.3 MATRIX INVERSION 678
A.4 MATRIX METHODS FOR SOLVING SIMULTANEOUS LINEAR ALGEBRAIC EQUATIONS 680
A.5 NON-LINEAR ALGEBRAIC EQUATIONS 691

***APPENDIX B* - MAXIMIZING THE CIRCUIT PERFORMANCE OF LONG-CHANNEL MOSFETS THROUGH DEVICE DESIGN AND PROCESSING TECHNOLOGY** 695

B.1 CHARACTERISTICS OF DIGITAL IC MOSFETS 695
B.2 DESIRED CIRCUIT CHARACTERISTICS OF IC INTERCONNECTS 697

***APPENDIX C* - DERIVATION OF THE BSIM I_D MODEL** 705

***APPENDIX D* - DERIVATION OF THE DISCRETIZED 2-D POISSON EQUATIONS USING THE BOX-INTEGRATION METHOD** 708

INDEX 709

CHAPTER 5

MOS TRANSISTOR DEVICE PHYSICS: PART 3

THE SUBMICRON MOSFET

The two major goals of MOSFET scaling are to increase the *density* and *speed* of the digital ICs in which such scaled-down devices are used. Increasing *density* obviously means using smaller channel lengths and widths. To increase the *speed* of digital ICs, the MOSFET saturation drain current I_{Dsat} must be increased (i.e., to allow faster charging and discharging of parasitic capacitances). From the long-channel MOSFET models given in chap. 4, a decrease in either the channel length L or gate oxide thickness t_{ox} will lead to an increase in I_{Dsat}. In fact, the long-channel MOSFET models predict that I_{Dsat} should continue to increase indefinitely as L and t_{ox} are decreased, seeming to imply that only the limitations of process technology (and not device effects) prevent the manufacture of even smaller, higher-performing MOSFETs. In addition, the MOSFET models of chap. 4 also accurately describe many other device characteristics of long-channel MOSFETs, including threshold voltage, subthreshold current, and the I_D-V_{DS} characteristics in strong inversion.

However, as process technology improved to the point where devices could be fabricated with gate lengths smaller than ~2 μm, it turned out that MOSFETs began to exhibit phenomena not predicted by the long-channel MOSFET models. Such phenomena were thus termed *short-channel effects*. One of the more surprising of such short-channel effects (which becomes especially pronounced as L decreases below 0.5 μm) is that the drain current in saturation I_{Dsat} shows far less increase as L is decreased than is predicted by the long-channel models (see Fig. 5-1).[1,47] In fact, in the simple model to be presented in section 5.4, I_{Dsat} is predicted to become independent of L in extremely small MOSFETs, approaching instead, a constant value given by $Q_I v_{sat} Z$ (where v_{sat} is the saturation velocity of the carriers, as described in section 5.4.1.2). In contrast, reducing t_{ox} (which has the effect of increasing the oxide field, and in turn causes Q_I to increase), yields a considerably greater increase in I_{Dsat}. Furthermore, as shown in Fig. 5-1 I_{Dsat} will also increase more rapidly as L is decreased if the MOSFET has a thinner gate oxide. This benefit provides even a greater impetus for making t_{ox} as thin as possible as gate lengths are decreased further in the drive for higher density.

Since the quest for higher density still requires L (and Z) to be further reduced, it will nevertheless be necessary to confront the other short-channel effects (besides the

TSMC Exhibit 1021, Page 20 of 35

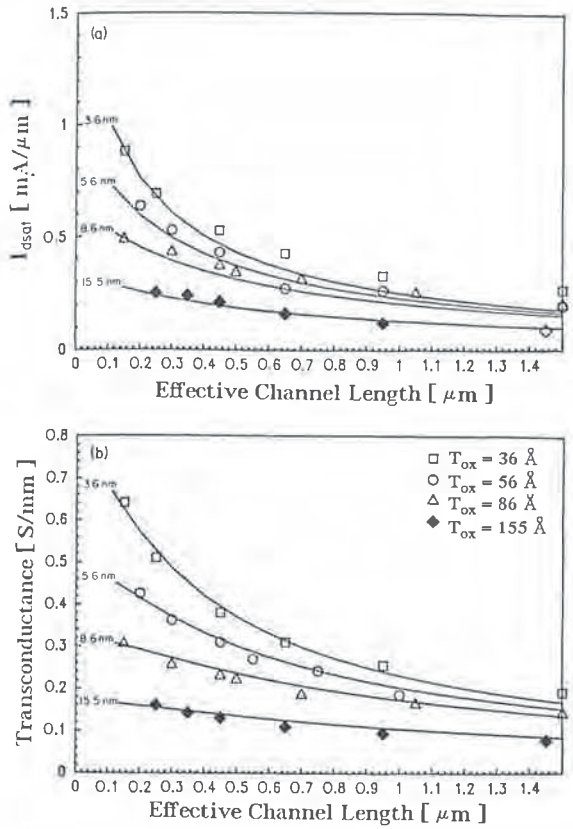


Fig. 5-1 Measured and calculated I_{Dsat} for an array of deep submicron NMOS transistors with four different gate-oxide thicknesses.[47] (© 1987 IEEE).

one of the I_{Dsat} dependence on L mentioned above). These other short-channel effects can be classified into two groups: (1) increased leakage current when the MOSFET is "off," and; (2) reliability problems associated with short-channel device structures.

Increased "off-state" leakage in short-channel MOSFETs is due to several phenomena, including: (a) lowering of the threshold voltage V_T as L is decreased and/or V_{DS} is increased; (b) the onset of punchthrough at smaller drain biases as L is decreased, and; (c) an increase in isolation leakage current as the isolation spacing is decreased. We will discuss V_T lowering and punchthrough in this chapter, and isolation leakage phenomena in chap. 6.

The reliability problems that arise in short-channel MOSFETs include: (a) thin gate-oxide breakdown (to be discussed in chap. 7); (b) device degradation due to hot-carrier effects (to be discussed in chap. 9); and (c) reliability problems associated with the interconnects between MOSFETs (e.g., electromigration failures in the metal lines). The latter problem is treated in Vol. 2, chaps. 3 and 4.

Table 5-1. Comparison of Long-Channel and Short-Channel MOSFET Characteristics

Long-Channel MOSFET Behavior	Short-Channel MOSFET Behavior
1. The threshold voltage V_T is independent of channel length L and width Z.	**1.** V_T decreases as L is decreased. V_T may also be impacted by changes in Z.
2. V_T is independent of drain bias voltage.	**2.** V_T decreases with increasing V_{DS}.
3. V_T depends on V_{BS} according to Eq. 4-12 (with V_{BC} in Eq. 4-12 replaced by V_{BS}).	**3.** V_T increases less rapidly with V_{BS} than predicted by Eq. 4-12.
4. The subthreshold current I_{Dst} increases linearly as L decreases.	**4.** I_{Dst} increases more rapidly than linearly as L decreases.
5. I_{Dst} is independent of drain bias.	**5.** I_{Dst} increases with increasing V_{DS}.
6. The subthreshold swing S_t is independent of gate length.	**6.** S_t increases with decreasing L.
7. The drain current in saturation I_{Dsat} is independent of V_{DS}.	**7.** I_{Dsat} increases as V_{DS} increases.
8. I_{Dsat} is proportional to $(V_{GS} - V_T)^2$	**8.** I_{Dsat} is proportional to $(V_{GS} - V_T)$.
9. I_{Dsat} is proportional to $1/L$.	**9.** As $L \rightarrow 0$, I_{Dsat} becomes independent of L.

In this chapter we will also cover several device design scenarios developed to keep leakage currents sufficiently small as the MOSFET is scaled down. In addition, we describe the issues related to submicron PMOSFET design.

5.1 DEVICE CHARACTERISTICS OF LONG-CHANNEL VERSUS SHORT-CHANNEL MOSFETS

From the perspective of MOSFETs as circuit components, long-channel behavior has been specified in a number of ways. As a result, short-channel device effects can be so designated when a deviation from long-channel MOSFET behavior can be correlated with a reduction in the gate length (and/or gate width) dimension. Table 5-1 gives a list of long-channel device characteristics that undergo variation as the gate dimensions are decreased. As the behavior of the short-channel MOSFET is described later in this chapter, the terms associated with these characteristics will be defined and discussed in more detail.

From Table 5-1 we can see that the short-channel effects can be divided into the following categories: (a) those that impact V_T (Items 1-3): (b) those that impact subthreshold currents (Items 4-6); and (c) those that impact I_D when the MOSFET is operated in saturation [i.e., $V_{DS}>(V_{GS} - V_T)$], namely Items 7-9. In this chapter we

retrograde interstitial concentration profile will cause a net boron diffusion flux toward the surface (i.e., during the source/drain activation anneal) even if a flat boron dopant profile exists in the vertical direction. As a result, boron piles up at the surface, leading to an increase in V_T. Experimental confirmation of such an impurity pile-up is found in SIMS analysis of flat profiles subjected to silicon damage and subsequent anneal.[99] The diffusion of boron toward the channel surface from this effect is a transient phenomenon because the interstitials are eventually annihilated during the anneal. Good agreement between this model and experimental data is also reported if the transient diffusion is modeled using a coupled impurity/point defect simulator.

5.2.3.4 Model which Attributes RSCE to Silicon Interstitial Capture in the Gate Oxide. Jacobs et al. note that V_T can also be shifted by a change in fixed oxide charges Q_f at the Si-SiO_2 interface.[97] Thus, they propose that RSCE in the absence of a vertical gradient in the channel dopant profile can be explained by the trapping of injected silicon interstitials by the gate oxide. These captured interstitials introduce the negatively charged complex in the oxide required for the observed V_T roll-up). The oxide surface is known to be an effective sink for interstitials, and the process steps which supply the excess interstitials can be those mentioned above. The result is a position-dependent oxide charge distribution along the Si-SiO_2 interface, similar to the assumption of an exponentially decaying fixed charge variation made by Arora when deriving the analytical model of RSCE mentioned in section 5.3.2.1.[94]

5.3 PUNCHTHROUGH (Subsurface-DIBL) IN SHORT-CHANNEL MOSFETS

In chap. 4, section 4.7 we described the nature of subthreshold current in long-channel MOSFETs. We noted that in weak inversion I_{Dst} flows at the surface of the channel region, and its value is calculated from Eq. 4-89. We also saw that the subthreshold-swing parameter (S_t) is an important characteristic of subthreshold behavior, and in long-channel, uniformly-doped-substrate devices it can be modeled using Eq. 4-91.

In short-channel MOSFETs, however, if V_{GS} is fixed while V_{DS} is increased, larger values of I_{Dst} values than are predicted by the long-channel device equations are observed. This larger subthreshold current is due in part to increased I_{Dst} flowing at the surface as a result of surface-drain-induced barrier lowering (DIBL), as described in section 5.2). If only surface DIBL is occurring in a MOSFET, S_t values are observed to remain unchanged. However, another short-channel effect known as *subsurface punchthrough* can also give rise to an increase in I_{Dst}, and one manifestation of this effect is an increase in the measured value of S_t. That is, the gate has less control of the subsurface punchthrough current than it has over the surface subthreshold current. Thus, once a MOSFET enters subsurface punchthrough and such current starts to flow, the S_t value of the MOSFET becomes larger than if only normal subthreshold current is flowing. Since the measurable drain current is the sum of both current components,

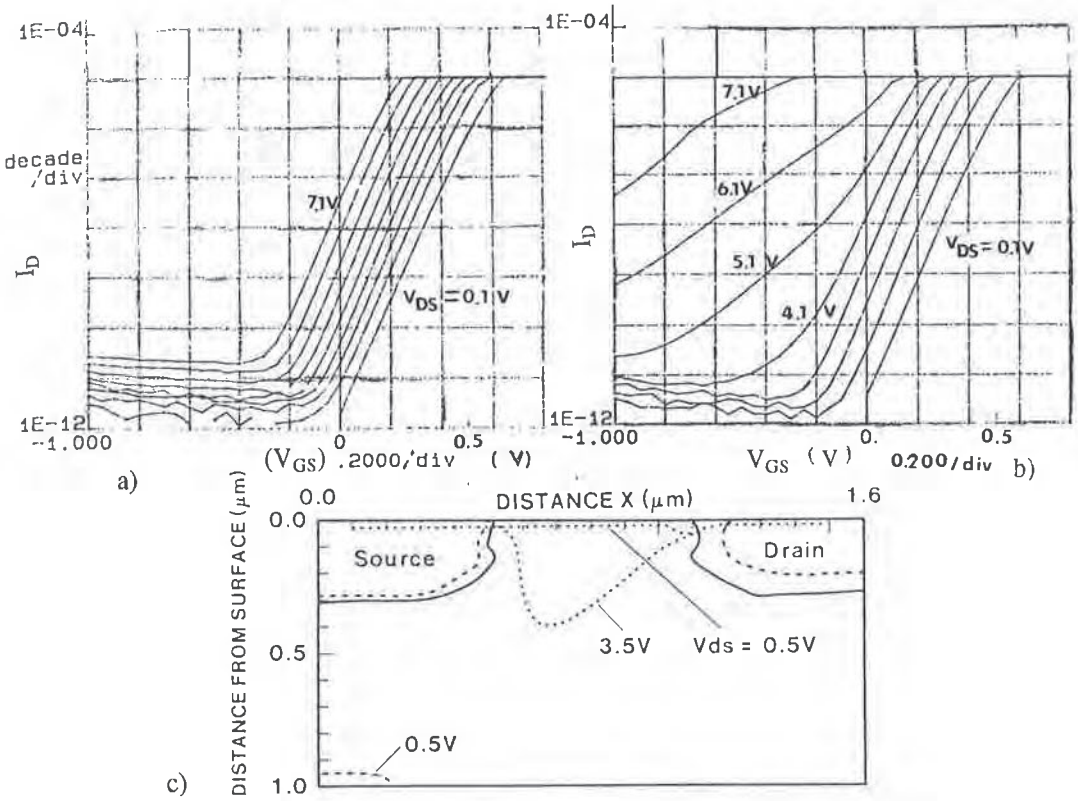


Fig. 5-21 Two sets of log I_D versus V_{GS} curves for MOSFETs, with V_{DS} varying in 1V increments. (a) No indication of punchthrough is observed, as S_t remains constant for all values of V_{DS}. (b) Onset of punchthrough occurs as V_{DS} reaches ~6V, as evidenced by an increase in S_t that point. (c) Punchthrough current paths (dotted) in a surface-channel NMOSFET with V_{DS} = 0.5 and 3.5V. Solid line shows the depletion region edge. The current path transfers from the surface to bulk as V_{DS} is increased.[74] (© 1988 IEEE).

one technique for detecting the onset of punchthrough in a MOSFET is to monitor S_t as V_{DS} is increased. .

For example, in Fig. 5-21 two sets of I_{Dst} versus V_{GS} curves with V_{DS} as the parameter are shown. In the MOSFET characterized by Fig. 5-21a S_t remains constant as V_{DS} is increased, indicating that no punchthrough is observed up to the maximum applied V_{DS}. In Fig. 5-21b S_t increases as V_{DS} is increased beyond ~4V, indicating that punchthrough sets in at about this value.

Punchthrough is a phenomenon associated with the merging of the source and drain depletion regions in the MOSFET. That is, as the channel gets shorter, these depletion region edges get closer (assuming that the channel-region doping is kept constant as L is decreased). When the channel length is decreased to roughly the sum of the two junction depletion widths, punchthrough is established. (Note that punchthrough may

even exist in a short-channel device at $V_{DS} = 0$, depending on L and the substrate doping.) Nevertheless, since the depletion regions of a *pn*-junction widen as reverse-bias is increased, all MOSFETs would eventually enter punchthrough if a high enough V_{DS} could be applied. However, in MOSFETs with $L>2.0$ μm, breakdown of the drain-substrate junction generally sets in before this punchthrough voltage is reached. As a result, in practice, punchthrough is not a limiting factor in long-channel digital MOSFET design. In shorter-channel devices (i.e., $L\leq2.0$ μm), however, punchthrough does represent a serious limitation.

Another factor also impacts punchthrough behavior in *n*-channel MOSFETs. That is, in submicron NMOSFETs a V_T-adjust implant is used to raise the doping of the surface-channel region above the doping of the substrate. (This can also be viewed as an increase in the electrostatic potential at that point near the source.) Thus, the source/drain depletion-region widening is greater beneath the surface due to the lighter doping there. A punchthrough condition would therefore be first established below the surface in such devices, and hence this effect is referred to as *subsurface punchthrough* (see Figs. 5-22 and 5-23). Note that in PMOSFETs fabricated with n^+ polysilicon gates the punchthrough problem occurs in a somewhat different manner, which we will describe in a later section dealing with this issue (i.e., section 5.8.2.3).

Once subsurface punchthrough occurs, any increase in the applied reverse-bias drain voltage (beyond that required to establish punchthrough) will lower the potential energy barrier for majority carriers in the source.[6] With a diminished barrier height, a larger number of source-region majority carriers (e.g., electrons in an NMOSFET) have enough energy to be injected from source to substrate. Some of these injected electrons are collected by the drain, causing I_{Dst} to increase. The component of I_{Dst} which flows as a result of subsurface punchthrough is known as *punchthrough current.* When designing MOSFETs one criterion of a successful device design is that the punchthrough current should be insignificant under normal bias conditions. That is, if sufficient punchthrough current is observed, the device structure must be redesigned to effectively suppress such current.

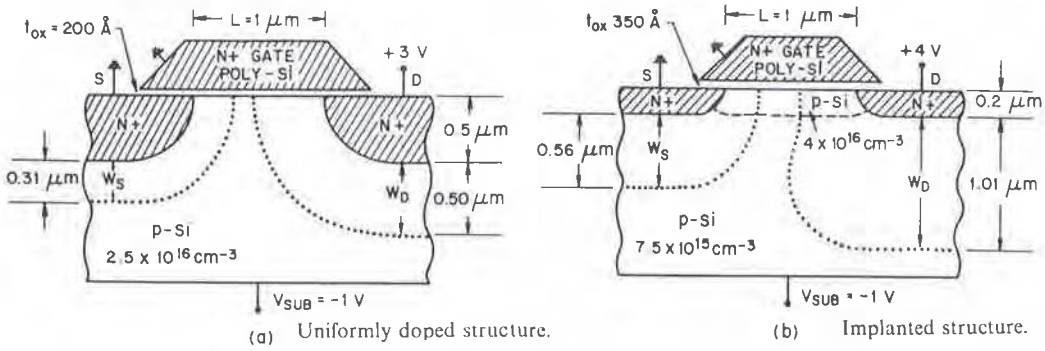


Fig. 5-22 (a) NMOSFET without V_T-adjust implant. (b) NMOSFET with V_T-adjust implant, showing the bulging of the drain depletion region beneath the surface as V_{DS} is increased, due to the lower doping concentration in the substrate below the depth of the V_T-adjust implant.

Two-dimensional simulations of the potential in the channel regions of short channel NMOSFETs with V_T-adjust implants support the view that depletion-region widening is greater beneath the surface. In addition, they indicate that the point of maximum source-substrate potential energy barrier lowering is also beneath the Si-SiO_2 interface (generally at approximately the same depth as the source/drain junction depths). Figures 5-23a, c, e and g show a "bird's eye" or perspective view of the electrostatic potential variation in a short-channel MOSFET computed using such 2-D numerical analyses, as V_{DS} is increased from V_{DS} = 3V to V_{DS} = 9V. Figures 5-23b, d, f and h show contour plots of these same results. We can see that when the device enters punch-through the location of the electrostatic potential saddle point is beneath the surface, and the value of the potential at this saddle point may be higher than the potential minimum at the surface. Since electrons follow the potential peak (see Figs. 5-23g and i), in such cases the component of current due to subsurface punchthrough will be larger than the channel-current component, and the subsurface current will flow along the path indicated by the subsurface potential ridge connecting the source and drain. Figure 5-21c shows a simulation of the subthreshold current paths in an NMOSFET with L = 0.6 μm for surface-DIBL (i.e., V_{DS} = 0.5V) and for subsurface-DIBL (punchthrough, V_{DS} = 3.5V). This shows that the subthreshold current path transfers from the surface to the bulk as V_{DS} increases, which supports the view that at low V_{DS} surface-DIBL is responsible for the increase in I_{Dst} in short-channel NMOSFETs, but that at high V_{DS} punchthrough becomes the primary cause of subthreshold leakage.

Subsurface punchthrough exerts the biggest impact on circuit operation when the MOSFET is OFF, since a tiny punchthrough current in this case can still represent intolerable leakage. However, the device parameter most commonly used to characterize punchthrough behavior is the *punchthrough voltage* V_{PT} defined as the value of V_{DS} at which I_{Dst} reaches some specific magnitude (e.g., 1 nA/μm) when V_{GS} = 0. The goal in designing a MOSFET is to have V_{PT} exceed any possible operating V_{DS}. Thus, V_{PT} is measured when the device is in subthreshold, as will be discussed in upcoming paragraphs. Note that V_{PT} can be roughly estimated by determining the value of V_{DS} at which the sum of the depletion widths W_S under the source and W_D under the drain (using the value of the doping concentration in the substrate as N_{SUB}) equals the channel length L, or from the expression[24]

$$V_{PT} \propto N_{SUB} (L - r_j)^3 \qquad (5 - 17)$$

Nevertheless, no accurate analytical model exists for estimating V_{PT} as a function of implant dose and depth. A full numerical solution of the 2-D Poisson equation must thus be obtained to predict the variation of V_{PT} for each device structure.

Several measures can be taken to prevent short-channel MOSFETs from entering subsurface punchthrough. First, the substrate doping can be increased to decrease the depletion-layer widths. A simple rule-of-thumb proposed by Klaasen[25] is to select a minimum substrate doping value according to $N_{SUB}>N_{ch}/10$, where N_{SUB} is the uniform bulk-substrate doping/cm^3 and N_{ch} is the average doping/cm^3 in the implanted region of the channel (e.g., D_I/x_c). This approach, together with a single V_T-adjust implant can produce appropriate threshold and punchthrough voltages for NMOSFETs

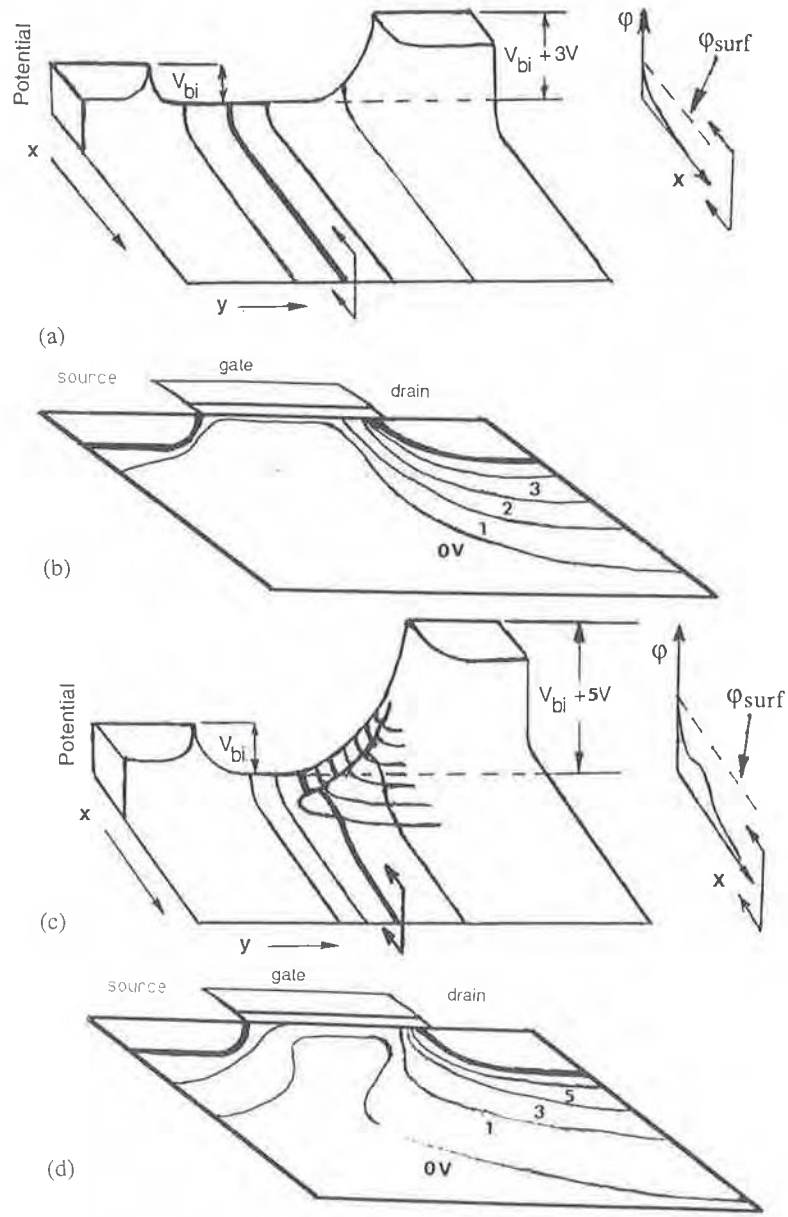


Fig. 5-23 (a and c) 2-D perspective plots of the potential in an *n*-channel MOSFET with a drain bias of (a) 3 V and (c) 5 V, respectively. Channel length = 1 μm. (b and d) 2-D contour plots of the same cases. From K. M. Cham, et al., *Computer Aided Design & VLSI Device Development*. Copyright 1986 Kluwer Academic Publishers. Reprinted with permission.

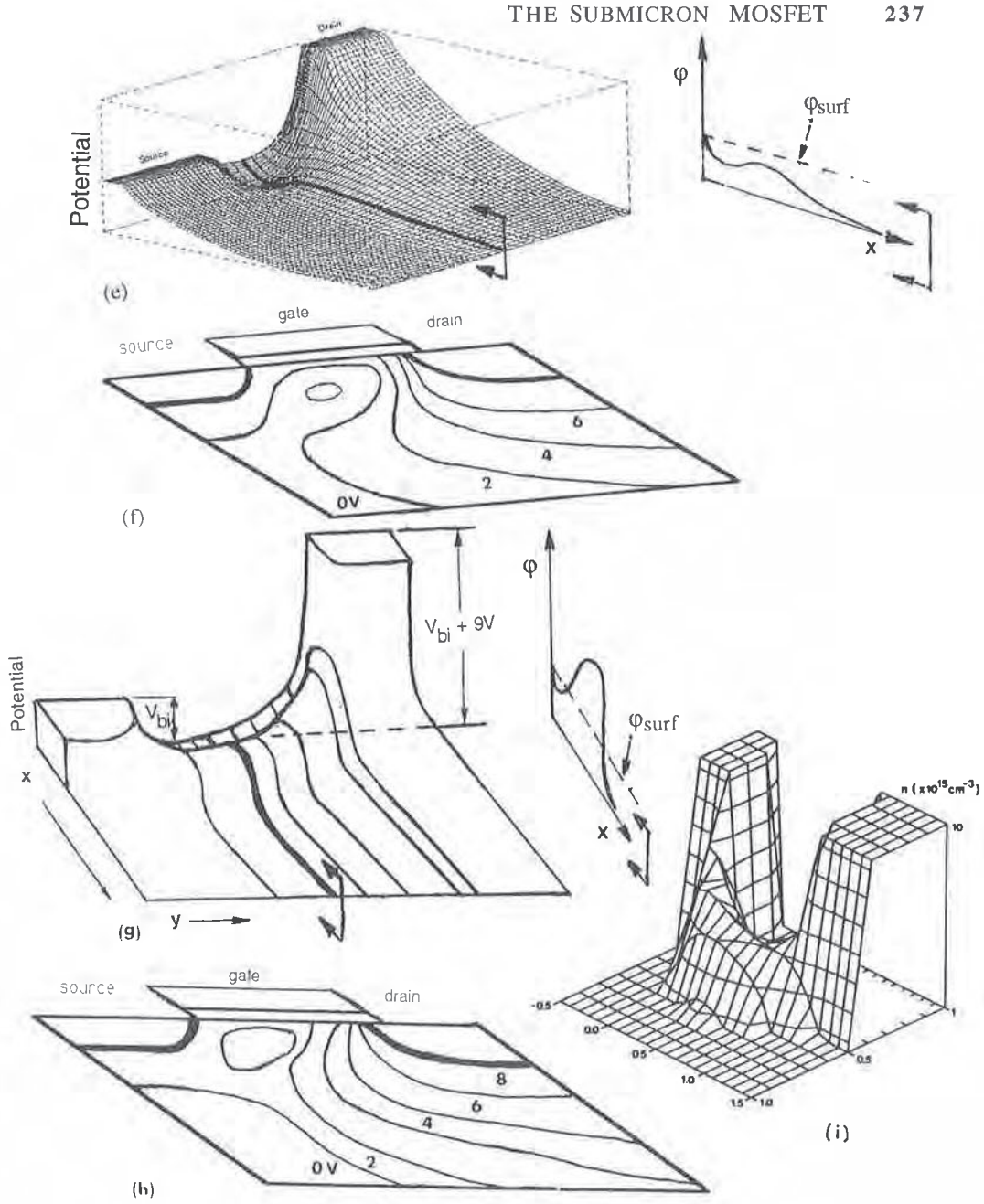


Fig. 5-23 cont. (e and g) 2-D perspective plots of the potential in an *n*-channel MOSFET with a drain bias of (e) 7 V and (g) 9 V, respectively. Channel length = 1 μm. (f and h) 2-D contour plots of the same cases. From K. M. Cham, et al., *Computer Aided Design & VLSI Device Development*. Copyright 1986 Kluwer Academic Publishers. Reprinted with permission. (i) Electron concentration of the MOSFET in punchthrough, i.e., case g & h.

with drawn gate lengths of 1.3 μm nominally operated at 2.5V power supply (e.g., a single 20 keV B^+ implant at ~$9x10^{11}$ cm^{-2}). A similar process has also been applied to 1.25 μm CMOS devices using a 5V supply. Although some short-channel effects are observed, V_T and V_{PT} voltages are reasonable. However, for devices with drawn gate lengths of ~1 μm (L_{eff} ~0.55 μm), a single V_T/punchthrough implant cannot adequately satisfy all requirements, especially if high resistivity (>10 Ω-cm) substrates are used.

As a result, another approach using an additional implant known as a *punchthrough stopper implant* (or *punchthrough implant,* PTI) must be employed. The second implant peak concentration is located at a depth near the bottom of the source-drain regions. Fig. 5-24a qualitatively depicts such a MOSFET structure. The additional doping introduced by such an implant (e.g., boron in NMOSFETs) reduces the lateral widening of the drain-depletion region below the surface without increasing the doping under the junction regions. With such implants, the component of the punchthrough current can be suppressed to well below the long-channel I_{Dst} value in MOSFETs down to 0.5 μm, while still maintaining a lower value of N_{SUB} (for reduced values of C_j). However, precise control of the placement and dose of this implant within the channel region is required to achieve the intended enhancements. For example, Fig. 5-24b shows the as-implanted doping profile of the channel with (dashed line) and without (solid line) a punchthrough implant (i.e., prior to annealing the implants). In addition, it is important to keep the punchthrough implant profiles from spreading out appreciably during annealing (see Fig. 5-24c), since the additional doping above the well's value adds to the source/drain-to-well capacitance and the transistor's threshold voltage sensitivity to back biasing of the*p*-well. Thus, the post channel implant thermal cycles should be kept to a minimum. An example of such a punchthrough implant for a 1 μm NMOSFETs is a 70 keV B^+ implant at $6x10^{11}$ cm^{-2}.

A third technique is to locally implant *p*-type dopants (in NMOSFETs) under the lightly doped tip region of the LDD as shown in Fig. 5-25a, and such implants have been termed a "halo" implants. (The use of halo implants as punchthrough stoppers in PMOSFETs is discussed in section 5.8.1.1.) Because the implanted dopant raises the doping concentration only on the inside walls of the source/drain junctions, the channel length can be decreased without needing to use a higher-doped substrate. That is, punchthrough does not set in until a shorter channel length because of the "halo." A penalty in increased sidewall junction capacitance is paid, but this cost is smaller than that incurred should the bottom area of the junctions lie above heavily doped substrate material. The maximum halo doping is determined by the onset of low-level avalanche breakdown of the pn^+ junction as is shown in Fig. 5-25b.

A halo NMOSFET with L = 0.35 μm, t_{ox} = 15 nm, and operated at 3V has been reported.[26] The junction depth r_j and *p*-halo peak doping were ~0.14 μm and ~$3x10^{17}$ cm^{-3}, respectively. A more recent report applied a large-angle tilt (LAT) implant of boron ions in NMOS and phosphorus in PMOS (Fig. 5-25c) to form a halo-like structure in each device (called a *self-aligned pocket implantation*). A tilt angle of 25-30° was used to implant the pocket dopants after gate formation, so that the gate structure acted as a self-aligned implant mask. Devices with channel lengths as small as 0.2 μm that still exhibited long-channel subthreshold V_T behavior were fabricated.[27]

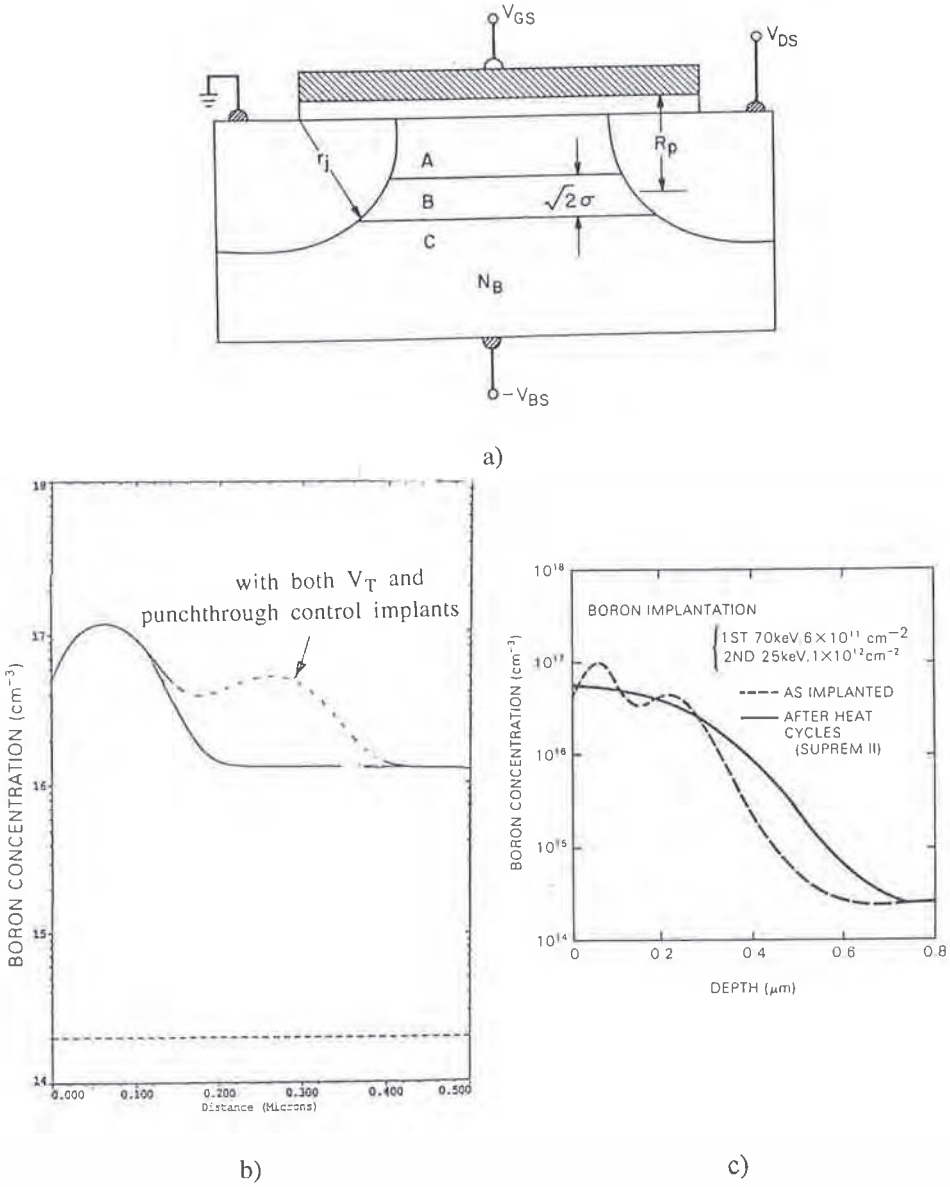


Fig. 5-24 (a) An implanted NMOS structure with three regions of different doping. Level A contains the dopant from the V_T implant. Level B contains the dopant from the punchthrough control implant, considered to be a Gaussian centered at a distance R_p from the gate, with a straggle σ. Level C is the original substrate. (b) As-implanted channel doping profiles with (dashed line) and without (solid line) punchthrough implant. (c) Example of simulated channel impurity profile of an NMOSFET with both V_T and punchthrough control implants - as-implanted (dashed lines), and after anneal (solid line).[113] (© 1982 IEEE).

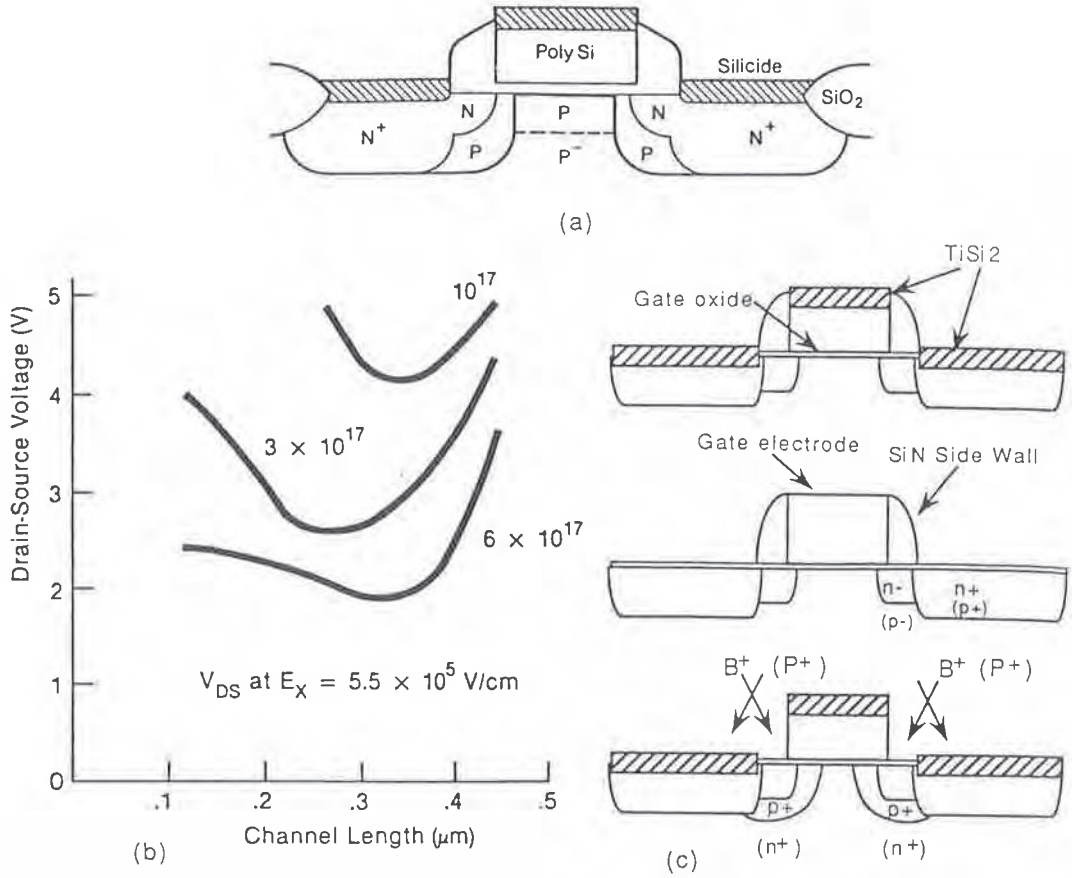


Fig. 5-25 (a) "Halo" NMOS structure.[26] (© 1985 IEEE). (b) V_{DSmax} limited by avalanche breakdown versus channel length for various peak halo doses.[26] (© 1985 IEEE). Process for producing "halo" NMOS structure using a Self-Aligned Pocket Implantation [SPI] technology is shown.[27] (© 1991 IEEE).

5.3.1 Experimentally Characterizing Punchthrough

The onset of punchthrough in a device can be measured in several ways. As mentioned at the outset of the discussion, for a specific device (i.e., one that has a single, specific channel length and channel doping profile), the onset of punchthrough over the operating range of V_{DS} can be monitored by measuring log I_{Dst} vs. V_{GS} for different V_{DS} values (Figs. 5-21a and b). If the curves look like those of Fig. 5-21a, the device does not exhibit subsurface punchthrough up to the maximum V_{DS} applied (since the slope of the log I_{Dst} curves remain constant as V_{DS} is increased). The curves of Fig. 5-21b however, indicate that punchthrough is setting in, with its onset marked by the V_{DS}-curve having a larger S_t value than that exhibited at small V_{DS}. (Note that since punchthrough current flows in parallel with ordinary I_{Dst}, S_t will noticeably increase

While this model is useful in providing insight into the general directions that LDD device-design should take, it does not provide enough information to be able to predict the actual hot-carrier lifetimes of LDD devices designed only from the data provided by the model. That is, the assumption is still made that an LDD MOSFET with the smallest value of $\mathcal{E}_{ymax}$ will yield a device structure with the highest hot-carrier reliability. Unfortunately, this is not always the case. Instead, other factors enter into the picture, as will be discussed in the section that examines the issues of hot-carrier reliability in LDD devices (section 9.7.2). We will also describe methods that are better able to predict hot-carrier hardness in LDD devices (section 9.10). These are based on more rigorous simulation procedures than those used in the above model.

9.6.1.3 Model for the Series Resistance Introduced by the n^- Region of the LDD. The value of the resistance R_{n^-} introduced by the n^- region of an LDD can be simply modeled if we assume that the doping concentration in the n^- region is uniform (with a value N_D) along the distance L_{n^-} and to the depth of the region r_j. In this case, we can model R_{n^-} in an NMOSFET according to:

$$R_{n^-} = L_{n^-} / (q\, \mu_n\, N_D\, r_j)\, Z \qquad (9 - 14)$$

where Z is the width of the device and μ_n is the mobility of the electrons in the n^- region (e.g., 800 cm^2/Vsec). The impact of R_{n^-} on I_D is described in section 9.7.1.1.

9.6.2 Overview of the Process Sequence to Form LDDs

The basic process sequence used to form early LDD MOSFETs in simple NMOS technologies is shown in Fig. 9-31. The poly gate is patterned with an RIE etch process designed to produce a poly feature with vertical sidewalls (Fig. 9-31b). In some processes the thermally grown gate oxide beyond the edges of the gate is kept in place after the poly etch, in an attempt to maintain a high quality interface under the sidewall oxide. This requires a polysilicon etch with extremely high selectivity to oxide, since the gate oxide is normally quite thin (e.g., 12.5-20 nm thick). Next, the n^- implant (also called the *graded-drain,* or *tip* implant) is carried out with phosphorus to form the lightly doped region of the drain. In early LDDs the doses ranged from 5×10^{12} to 1×10^{13} cm^{-2} at energies of 40-60 keV (Fig. 9-31c). Note that the n^- region at this point is essentially self-aligned to the gate electrode. The n^- implant is carried out using a medium-current implanter with the ion beam tilted at 7° (to minimize implant channeling effects - see vol. 1, chap. 9).

Following the n^- implant step a thin oxide is thermally grown on the patterned poly gate using a short dry-oxidation process (this step is called *poly re-ox*). A conformal CVD oxide film is then deposited. A popular method is to decompose TEOS at 725°C, as this produces a film with excellent conformality, and if the poly profile is vertical, no re-entrant profile. A conformal deposition process produces a thicker film at the edge of the gate-to-active area step than on flat areas. Thus, an anisotropic dry-etch process will clear the oxide in the flat areas while leaving spacers at the sidewalls of the poly

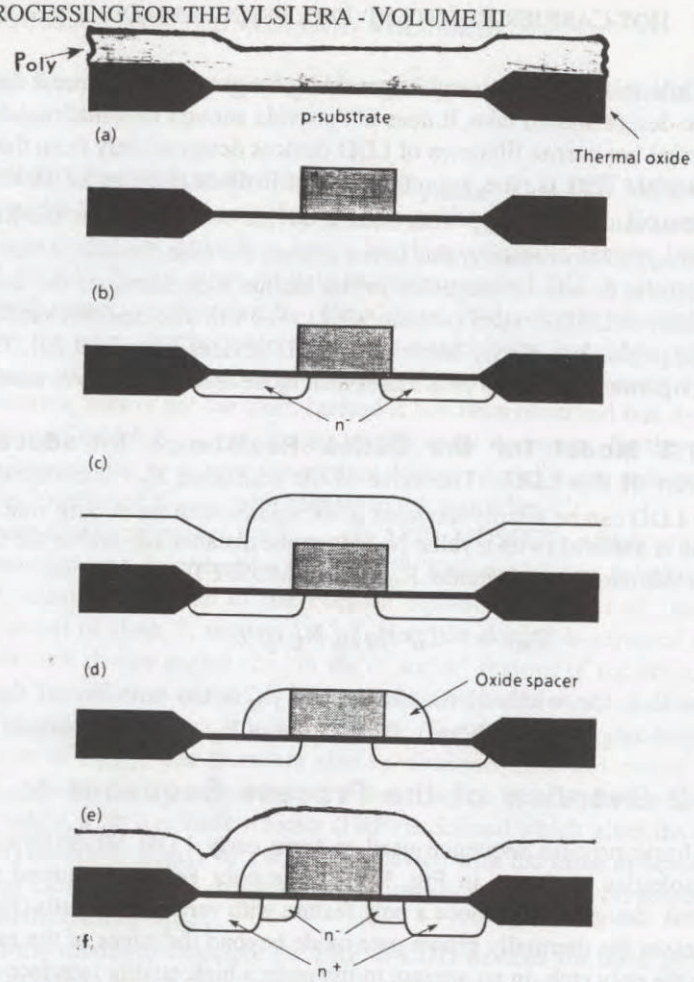


Fig. 9-31 Basic process sequence to form an LDD NMOSFET. (a) After poly deposition; (b) after poly etch; (c) LDD implant; (d) spacer deposition; (e) spacer etch; (f) n^+ implant and passivation.

gate (Fig. 9-31d). If the spacers are etched to endpoint with minimal overetch, the spacer height will be close to the same height as the poly. For poly gate thicknesses of 0.3 μm and a deposited oxide film thickness of ~0.3 μm, spacer widths of ~0.2 μm are typically obtained. A discussion of spacer thickness and shape dependence, with respect to variations in the deposited film thickness, conformality, overetch time etc. is given in ref. 51.

It should also be noted that to account for variations in the spacer oxide layer thickness, some overetch is needed in the etch step that creates the spacer. During the overetch time the spacer shape is altered (i.e., its thickness and height from the top of the poly gate will change). In addition, the field oxide and the silicon in the source/drain regions are also etched. Excessive overetching of these regions is undesirable from

device isolation and junction leakage viewpoints, respectively. Figure 9-32 shows a SEM photograph of an NMOS transistor with an oxide spacer at the conclusion of the spacer etch process, where the n^+ gate and drain regions have been decorated with a selective etch (the n^- region is not visible). Note that the silicon in the drain has been slightly etched during the spacer overetch as shown by the depression in the silicon surface to the right of the gate. Leakage of the source/drain junctions, localized in the junction directly beneath the spacer edge has been reported in oxide-spacer LDDs. This problem has been attributed to the silicon loss and etch damage in the substrate arising from the spacer etch,[52,53] and to defects formed by the subsequent n^+ implant and anneal. (The shape of the oxide spacer produced by the spacer etch process has also been reported to impact the formation of these defects.[54])

After the oxide spacers are formed, a high dose n^+ source/drain implant is carried out with a high-current implanter. Typically, arsenic is implanted at a dose of about 5×10^{15} cm^{-2} and at energies of 40-80 keV. This forms the low-resistivity drain region, which is merged with the lightly doped n^- region. Because the spacer serves as a mask for the As implant, the heavily doped n^+ region is self-aligned to the sidewall spacer edges and is thus offset from the gate edge. Since the edge of the n^+ region is further away from the channel than would be the case in a conventional drain structure, the depth of the heavily-doped region of the drain can be made somewhat greater without adversely impacting the device operation (e.g., 0.3 μm deep versus 0.18 μm deep). The increased junction depth lowers both the sheet resistance and the contact resistance of the drain. Deeper junctions also provide better protection against junction spiking. An anneal is then performed to cause the n^- dopant to sufficiently diffuse laterally under the gate edge and to activate the implanted n^+ dopants. (Actually, the entire thermal budget after n^- implant determines the degree of lateral diffusion, and this is often dominated by the interlevel dielectric densification and reflow step, e.g., 950°C for 20 min.)

By introducing the sidewall spacer after the n^- implant to laterally offset the n^+ implant, the LDD structure offers more freedom in optimizing the n^- length than the DDD structure. That is, the n^- length is controlled by the lateral spacer dimension and thermal drive cycle, and is made independent from the n^+ junction depth. If a salicide

Fig. 9-32 SEM photograph of an NMOS transistor with an oxide spacer at the conclusion of the spacer etch process, where the n^+ gate and drain regions have been decorated with a selective etch (the n^- region is not visible).[67] Reprinted with permission of Semiconductor International.

process is being implemented, an oxide spacer is also needed to prevent shorting of the gate and the source/drain regions, and this structure is intrinsic to the LDD approach. However, as we shall describe in the next section, the conventional LDD has other problems and disadvantages, and improvement in hot carrier resistance in the basic LDD structure is obtained only by trading off device performance.

9.7 PERFORMANCE AND HOT-CARRIER LIMITATIONS OF THE CONVENTIONAL LDD

As emphasized earlier, the reason LDDs are incorporated into submicron MOSFETs is to increase their hot-carrier reliability. It would be ideal if this benefit could be obtained without degrading the other performance characteristics of the device. However, the introduction of an LDD generally does cause some performance degradation, primarily by causing a reduction in the drive current of the MOSFET. An example of this effect is illustrated in Fig. 9-33,[55] which compares the drain current in conventional and LDD MOSFETs as their gate length and n^- dose is varied. The LDD MOSFETs exhibit smaller I_D values. To make matters worse, if LDD MOSFETs are poorly designed they may even be more susceptible to hot-carrier degradation than comparable non-graded-drain MOSFETs.

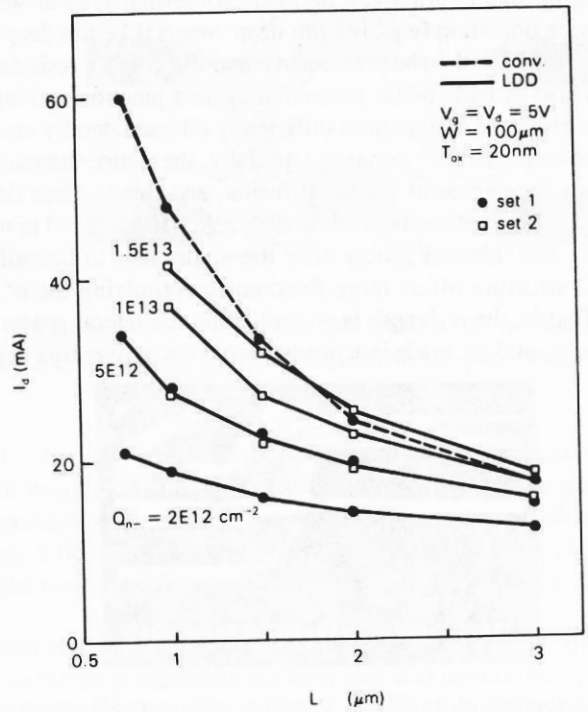


Fig. 9-33 Drain current in LDD MOSFETs with various gate lengths.[55] (© 1984 IEEE).