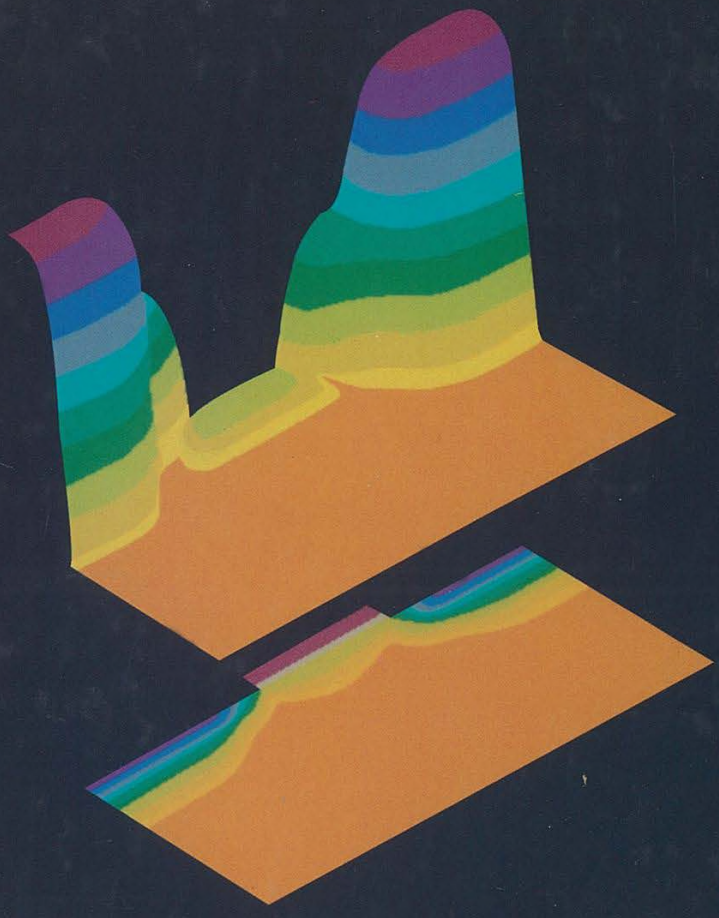
Kirkland & Ellis Library
5 0628 01144179 4

Silicon Processing

for the VLSI Era

Volume 3 - The Submicron MOSFET



S. Wolf

THE COMPANION VOLUMES TO THIS BOOK

Silicon Processing

for the VLSI Era

Volume 1 - Process Technology

By STANLEY WOLF and RICHARD N. TAUBER

TABLE of CONTENTS - Ch.1 Silicon: Single-Crystal Growth and Wafering; **Ch. 2** Crystalline Defects, Thermal Processing, and Gettering; **Ch. 3** Vacuum Technology for VLSI Applications; **Ch. 4** Basics of Thin Films; **Ch. 5** Silicon Epitaxial Film Growth; **Ch. 6** Chemical Vapor Deposition of Amorphous and Polycrystalline Films; **Ch. 7** Thermal Oxidation of Silicon **Ch. 8** Diffusion in Silicon; **Ch. 9** Ion Implantation for VLSI; **Ch. 10** Aluminum Thin Films and Physical Vapor Deposition in VLSI; **Ch. 11** Refractory Metals and Their Silicides in VLSI; **Ch. 12** Lithography I: Optical Photoresists - Material Properties and Process Technology; **Ch. 13** Lithography II: Optical Aligners and Photomasks; **Ch. 14** Advanced Lithography: E-Beam and X-Ray; **Ch. 15** Wet Processing: Cleaning, Etching, and Lift-Off; **Ch. 16** Dry Etching for VLSI; **Ch. 17** Material Characterization Techniques for VLSI Fabrication; **Ch. 18** Structured Approach to Design of Experiments for Process Optimization.

1986 LATTICE PRESS 660 pp. ISBN 0-9616721-3-7

Volume 2 - Process Integration

By STANLEY WOLF

TABLE of CONTENTS - Ch. 1 Process Integration for VLSI and ULSI; **Ch. 2** Isolation Technologies for Integrated Circuits; **Ch. 3** Contact Technology and Local Interconnects for VLSI; **Ch. 4**; Multilevel Interconnect Technology for VLSI and ULSI; **Ch. 5** NMOS Devices and Process Integration; **Ch. 6** CMOS Process Integration; **Ch. 7** Bipolar and BiCMOS Process Integration; **Ch. 8** Semiconductor Memory Process Integration; **Ch. 9** Process Simulation.

1990 LATTICE PRESS 752 pp. ISBN 0-9616721-4-5

on last page of book

0961672153 (249884)
Silicon Processing for the VLSI Era
Vol. 3: The Submicron MOSFET

BOOK

ing

hnology

HARD N. TAUBER

Single-Crystal Growth and Wafering;
essing, and Gettering; **Ch. 3** Vacuum
4 Basics of Thin Films; **Ch. 5** Silicon
Vapor Deposition of Amorphous and
idation of Silicon **Ch. 8** Diffusion in
Ch. 10 Aluminum Thin Films and
11 Refractory Metals and Their
Optical Photoresists - Material
h. 13 Lithography II: Optical Aligners
nography: E-Beam and X-Ray;
ning, and Lift-Off; **Ch. 16** Dry Etching
on Techniques for VLSI Fabrication;
of Experiments for Process

p. ISBN 0-9616721-3-7

tegration

cess Integration for VLSI and ULSI;
rated Circuits; **Ch. 3** Contact Technology
4; Multilevel Interconnect Technology
ices and Process Integration;
h. 7 Bipolar and BiCMOS
uctor Memory Process Integration;

pp. ISBN 0-9616721-4-5

e of book

Silicon Processing
for the VLSI Era

Volume 3

The
Submicron
MOSFET

S. Wolf

TK
7874
.W635
2000
V. 3

Kirkland & Ellis Librar
5 0628 01144

Sili

Volu

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

Dell Ex. 1008C
Page 4

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

PREFACE

SILICON PROCESSING FOR THE VLSI ERA now represents a series of texts designed to provide a comprehensive and up-to-date treatment of this important and rapidly changing field. The volume in hand is the third of this series. Volumes 1 (*Process Technology*) and 2 (*Process Integration*) were published in 1986 and 1990, respectively. Volume 1 deals with the individual processes employed in the fabrication of silicon VLSI circuits, such as epitaxial growth, thermal oxidation of silicon, CVD and PVD of amorphous and polycrystalline films, diffusion and ion implantation of dopants in silicon, microlithography, and patterning technology. Volume 2 describes how the individual processes of Volume 1 are combined in various ways to produce MOS and bipolar VLSI and ULSI circuits. The task of integrating these various fabrication processes together is referred to as *process integration.*

Here, we treat the topics of submicron MOSFET device physics and the relationship between such device physics and submicron MOSFET fabrication. That is, an understanding of device physics has become even more important now that MOSFETs have crossed the long-channel frontier into the submicron realm. Device aspects that could once be ignored (because they cause only second-order effects in large MOS devices) are now significant. Many device effects observed in submicron MOSFETs are impacted by the process technology used to fabricate them. Some of such effects include: short-channel effect on V_T; RSCE (reverse-short-channel effect on V_T); DIBL (drain-induced barrier lowering); narrow-width effect; reverse-narrow-width effect; subsurface punchthrough in NMOSFETs; punchthrough in PMOSFETs; impact of the V_T-adjust implant on subthreshold swing, S_t; boron penetration of gate oxides in p^+-poly-gate PMOSFETs - and the polysilicon depletion effect that this causes; GIDL (gate-induced drain leakage); impact of process technology on the reliability and wearout of thin gate oxides, including damage from plasma processing; hot-carrier degradation and drain-engineered MOSFET structures developed to combat the problem (including LDDs, LATID, "halo" implants, and asymmetrical MOSFETs); well-engineering by high-energy implants (including retrograde-well CMOS); ROXNOX (re-oxidized-nitrided oxides); and very-lightly-nitrided gate oxides. These and many other phenomena associated with submicron MOS fabrication can only be understood in the context of short-channel MOSFET physics. Only with such an understanding can the relationship between circuit behavior, device design, and process technology of submicron MOSFETs be grasped. Consequently, by gaining such an understanding process engineers (and other microelectronic professionals) will be better able to contribute to the task of successfully designing and manufacturing submicron ICs. One of our main purposes is thus to provide a text that treats both the topics of submicron MOSFET device physics and the phenomena associated with fabricating such devices.

Chapter 1 introduces the process and device models employed in the early days of the semiconductor industry. Such models were based on empirical data and/or simplified physical equations (e.g., differential equations in their one-dimensional form). In the

Dell Ex. 1008C
Page 20

submicron era, however, the more general forms of these differential equations must be used to obtain accurate predictive capability of both the fabrication processes and the device physics. Since this generally means that partial differential equations in two- or three-dimensional form must be solved, this can only be carried out by numerical analysis, performed with the aid of high-speed digital computers. Thus, chapter 2 outlines the methodology of this approach.

To facilitate the discussion of submicron MOSFET device physics, we divide the topic up into three chapters: The first of these, chapter 3, deals with basic MOS theory and the MOS capacitor. The next, chapter 4, covers long-channel MOSFETs and the circuit models developed to predict the drain current characteristics of such devices. Finally, chapter 5 describes the characteristics of the short-channel (i.e., submicron) MOSFET. This is the longest chapter of the book, and it includes the following topics: DIBL, subsurface punchthrough, the drain current in saturation, simplified short-channel MOSFET circuit models, MOSFET scaling, and submicron PMOSFETs.

Chapter 6 is concerned with isolation issues in integrated circuits with an emphasis on CMOS technology. Chapter 7 covers thin gate oxides, primarily with respect to their reliability and growth. Chapter 8 deals with well formation in CMOS. Chapter 9 focuses on hot-carrier effects in MOS ICs, as well as on device structures and processing techniques that mitigate the detrimental aspect of such phenomena on MOSFETs. Problems are included at the end of each chapter to assist readers in gauging how well they have assimilated the material in the text.

A book of this length and diversity would not have been possible without the indirect and direct assistance of many other workers. From an indirect perspective, virtually all the information presented here is based on the research efforts of countless numbers of scientists and engineers. Their contributions are recognized to a small degree by citing some of their articles in the references given at the end of each chapter. Likewise, the direct help came in a variety of forms, and was generously provided by many people. The text is a much better work as a result of this aid, and the author expresses heartfelt thanks to those who gave of their time, energy, and intellect.

Roy Montibon of Visionary Art Resources, Inc., Los Angeles, CA designed this cover, as well as those of Volumes 1 and 2.

Stanley Wolf

P.S. Additional copies of the books can be obtained from:

Lattice Press
Post Office Box 340-V
Sunset Beach, CA 90742

For your convenience an order form is provided on the final leaf of the book.

CHAPTER 3

MOS TRANSISTOR DEVICE PHYSICS: PART 1
BASICS MOS PHYSICS
and
THE MOS CAPACITOR

A basic goal of IC manufacture is to produce circuits having the highest performance at lowest cost. For digital CMOS ICs (the largest category of silicon ICs in the mid-1990s), the main performance criteria are speed, power dissipation, and device packing density. Reduction in device size has been the chief vehicle for achieving these goals. Smaller devices also produce the twin benefits of decreased die size (more chips per wafer and higher fractional yield) which exert downward pressure on die cost. However, the mere reduction of all device dimensions by a constant factor, for example, without paying attention to other processing parameters can adversely impact IC performance. That is, power dissipation may worsen because of increased device leakage currents, or circuit speed may be degraded. Reliability problems which afflict submicron MOSFETs might also be exacerbated, including hot-carrier degradation, gate-oxide wearout, and electromigration. As a result, it is necessary to employ more sophisticated scaling guidelines for device design to ensure that the shrunken devices exhibit electrical behavior comparable to that of previous generations of larger devices. In the case of MOSFETs, the aim of such scaling guidelines is to provide a design path to *short-channel devices* which continue to exhibit the ideal characteristics of *long-channel devices*. (These terms are defined later in the chapter.)

Our purpose here is to present the issues of submicron MOS device design for semiconductor process engineers. An understanding of device physics has become even more important now that MOSFETs have crossed the long-channel frontier into the short-channel (submicron) realm. Device aspects which could once be ignored (because they cause only second-order effects in large MOS devices) are now significant - such as source/drain doping profiles, channel doping profiles, and the device geometry around the gate periphery. Furthermore, if the degree of process control is not increased, variations in these parameters will become larger (on a percentage basis) as the devices become even smaller. Since the uniformity of device characteristics will be adversely impacted, process control must be improved. In addition, better processes will need to

Dell Ex. 1008C
Page 22

be developed and devices will have to be designed with less process sensitivity. All of these issues can only be discussed in the context of short-channel MOSFET physics. By understanding the relationship between circuit behavior, device design, and process technology of submicron MOSFETs process engineers will be better able to contribute to the task of successfully designing and manufacturing submicron ICs.

The discussion of MOS device physics also serves to set the stage for some later topics of this book (and others of this series), namely the process technology developed to fabricate silicon ICs. To facilitate the understanding of the submicron MOSFET, we divide the topic of *MOS device physics* into three chapters. The first (chap. 3) deals with basic MOS theory and MOS capacitors The second (chap. 4) covers long-channel MOSFETs and the circuit models developed to predict the drain current characteristics of such devices. The last (chap. 5) describes short-channel (i.e., submicron) MOSFETs and issues of MOSFET device scaling.

In each of these chapters the following general approach is adopted. As each new topic is introduced it is first considered from a *qualitative* perspective, using visualization analogies whenever possible. This is intended to provide an intuitive, rudimentary grasp of the device phenomenon being presented. Then the same device effects are analyzed from a *quantitative* perspective, and mathematical relationships that describe the physical effects are developed. Where applicable, as a final task, quantitative models of the device behavior in terms of the terminal voltages and currents are derived. Where possible complicated mathematics and derivations are minimized to prevent them from standing in the way of conceptual explanations. Where such complications cannot be completely circumvented many of the intermediate steps are given. Hopefully this will make the material somewhat more digestible.

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

As noted in the introduction, the ultimate purpose of this and the next two chapters is to provide an understanding of short-channel MOSFET behavior (extending to devices with submicron channel lengths). Since the behavior of submicron MOSFETs is complex, the subject is best approached in a step-by-step fashion. That is, the device behavior of the simplest MOS structure (i.e., the two-terminal MOS capacitor, to be referred henceforth as the MOS-C) is considered first. The MOS-C exhibits behavior which is the foundation of both long- and short-channel MOSFET operation. Such behavior is easier to describe in the MOS-C since some of the device phenomena exhibited in more complex MOS structures (i.e., multi-dimensional electric fields and drain currents) are absent in the MOS-C. Once the reader acquires a good understanding of MOS-C behavior, these effects can be added, allowing a relatively straightforward explanation of long-channel MOSFET operation. Thereafter, submicron MOSFET behavior can be explored by showing how the long-channel MOSFET behavior is altered as the channel length is reduced. With this rationale in mind, we begin by describing the behavior of the MOS capacitor in this chapter - initially dealing with the *ideal* MOS capacitor, but later incorporating *nonideal* effects. We should also point out

that the MOS-C itself is quite a complex and subtle structure. In fact, a lengthy and rigorous text that deals comprehensively with these subtleties has been written by Nicollian and Brews.[1]

3.1.1 The Structure of the MOS-C and the MOSFET

Although we begin by studying the MOS-C, it is useful to compare its structure with that of the MOSFET. To do this, an even-more-fundamental configuration, the generic MOS structure is first defined. The *MOS* or *metal-oxide-semiconductor* structure is a three-layered object. It consists of a conducting layer (referred to as the *gate*), atop an insulator layer (usually silicon dioxide), and a semiconductor substrate beneath the insulator (termed either the *substrate* or *body*). The substrate can be doped either *n*-type or *p*-type. Note that the acronym MOS stems from the early days of MOS fabrication when all gates were made of metal (aluminum). While heavily doped polysilicon has replaced aluminum as the standard gate material, the name MOS continues to be used.

The structure of the *MOS capacitor* (or *MOS-C*) is shown in Fig. 3-1a. It consists of the generic MOS structure with terminals G and B connected to the gate and substrate, respectively, making it a *two-terminal device*. An external bias can be applied across these terminals, called the *gate-substrate (or gate-body) voltage* V_{GB}. In Fig. 3-1a1 we show a perspective view of a MOS-C having a *p*-type Si substrate. Then in Fig. 3-1a2 we show the cross-section of this device cut along the middle of the gate. This cross-section is shown lying flat Fig. 3-1a3, and in Fig. 3-1a4 the same structure with only the Si beneath the oxide is shown.

The structure of the *MOS field-effect transistor* (or *MOSFET*) is given in Fig. 3-1b. It has two regions of doping opposite that of the substrate, one at each edge of the MOS structure. These regions are called the *source* and *drain*, and a *pn* junction exists between them and the substrate. When terminals are connected to all the various regions of the MOSFET a *four-terminal device* results, with the terminals designated as G (gate), S (source), D (drain), and B (substrate).

We next observe that since one of these terminals can be designated as the common (or reference) terminal, three independent terminal voltages can be applied to the MOSFET (e.g., if B is chosen as common we will have V_{GB}, V_{SB}, and V_{DB}). However, only one significant current exists in an ideal MOSFET. That is, we assume that the gate current is zero ($I_G = 0$), and that the source and drain junctions are always kept under reverse bias during normal MOSFET operation. Since reverse bias current in a *pn* junction can be considered negligible (and $I_G = 0$), substrate current will also be inconsequential ($I_{sub} \approx 0$).* Thus, only the drain current I_D which flows between the source and drain in the MOSFET needs to be considered. In summary, three independent terminal voltages and one current (i.e., I_D) are generally associated with the operation of the ideal MOSFET. In Fig. 3-1b1 we show the MOSFET in perspective, then in cross-section (Fig. 3-1b2), and finally (in Fig. 3-1b3), with the cross-section lying flat.

Figure 3-1 also defines the axes to be used in our discussion of both the MOS-C and the MOSFET structures. That is, the *x*-direction is perpendicular to the Si-SiO_2

* We'll remove these restrictions in chaps. 7 and 9.

interface (*vertical* direction), with $x = 0$ at the Si surface. The y-direction is parallel to the Si surface in the direction from source to drain (*lateral* or *longitudinal* direction). Since the source and drain are separated by a distance L (termed the *channel length*), we usually choose $y = 0$ at the source end of the channel, and $y = L$ at the drain end. The z-direction is the other direction parallel to the Si surface (perpendicular to the y-direction), and defines the channel width.

3.1.1.1. Comparison of MOS-Cs and Conventional Capacitor Structures. It should be noted at the outset that the structure and behavior of the MOS-C differs from that of the more conventional metal-dielectric-metal capacitor structure. That is, in the latter structure both terminals are highly conductive, but in the MOS-C the semiconductor terminal is not of high conductivity. Consequently, if a potential difference is applied across the terminals of the conventional capacitor, an electric field arises within the dielectric, but is zero outside the dielectric. This implies that the "metal-dielectric-metal" capacitor stores all of the charge at the interfaces of the respective metal electrodes and the dielectric.

The situation is different in the MOS-C where one of the terminals of the capacitor is a semiconductor, not a metal. If we consider the case of a p-type semiconductor, then the application of a *negative* bias will induce a negative charge on the gate and a positive charge on the substrate. The high conductivity of the gate requires that the electric field within the dielectric again terminates at the gate-dielectric interface. Here, the charge on which the field lines terminate is the negative charge stored on the gate. At the substrate-dielectric interface, however, the semiconductor must supply positive charge to balance the gate's negative charge. This can readily occur because the majority mobile charge carriers in p-type Si are holes, implying that (1) they exist in great abundance and (2) that they are capable of moving to the interface in response to the negative bias on the gate. In this case, the p-type silicon MOS-C and the conventional capacitor structure behave identically. Furthermore, as we will describe in more detail in the next section, this bias condition is known as "accumulation" because it causes majority carriers to accumulate at the semiconductor surface.

On the other hand, if a *positive* gate bias is applied to this MOS-C, a somewhat different set of phenomena occur. That is, a positive gate potential produces a positive charge on the gate, and a negative charge must arise in the substrate to balance the gate charge. In this case, there are very few mobile negative charge carriers (electrons) in the p-substrate. Thus, their scarcity prevents them from supplying enough of the required negative charge, especially at small positive gate-bias values. Instead, the repulsion of holes (the majority carriers) by the positive bias drives the holes away from the substrate-dielectric interface. This forms a region depleted of mobile carriers (i.e., a *depletion region*) that penetrates from the surface into the substrate. The departure of holes from this surface region leaves it negatively charged because the immobile, ionized acceptor impurities remain within the region. This negative "background" charge thus becomes the negative charge induced by the positive gate potential. As a result, the electric field lines terminate on these negative charges. Unlike in the case of accumulation, however, the terminal charge of the MOS-C in the semiconductor is no

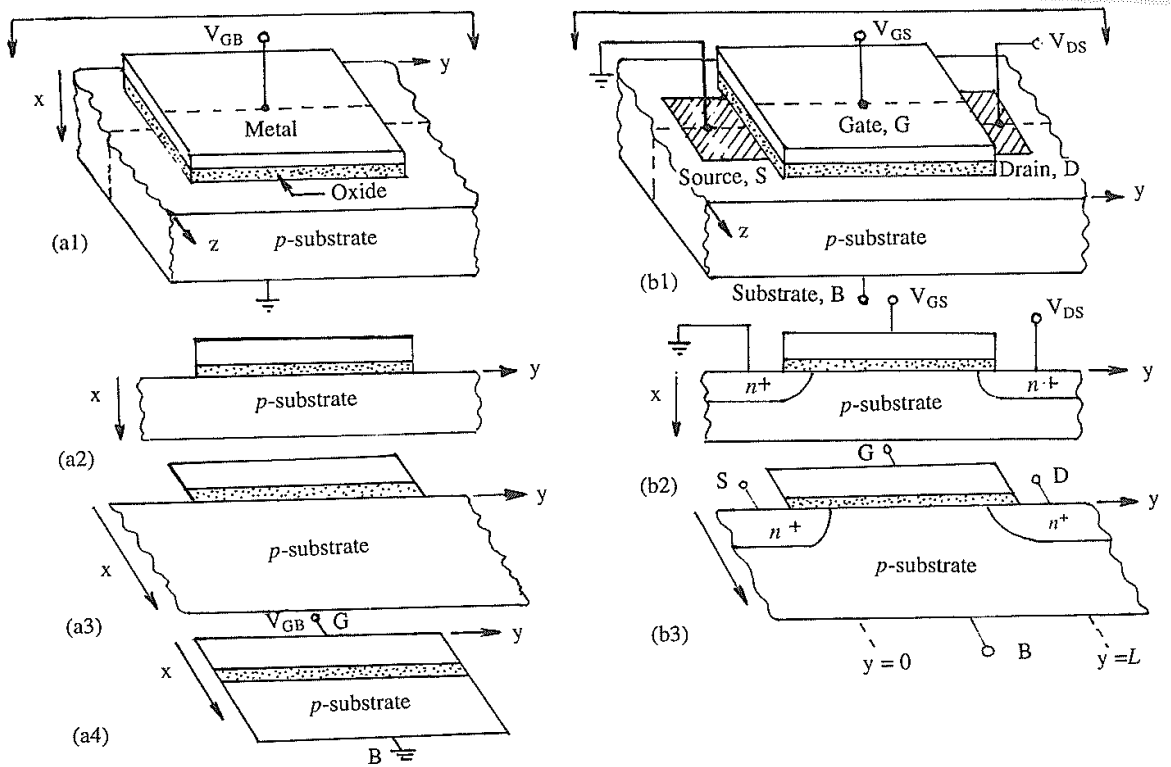


Fig. 3-1 (a1) Perspective view of the MOS-C, (a2) Cross-section of the MOS-C cut along the middle of the structure, (a3) Cross-section shown lying flat, (a4) Cross-section with only the Si beneath the oxide shown. (b1) Perspective view of the MOSFET structure, (b2) Cross-section of the MOSFET cut down the middle of the channel, (b3) Cross-section of the MOSFET lying flat.

longer confined to the substrate-dielectric interface (i.e., to the Si surface). Instead, the charge is spread over some distance extending to the depth of the depletion region. This added distance increases the "effective" dielectric thickness and thus decreases the capacitance of the structure. Thus, unlike a conventional capacitor, whose capacitance is constant regardless of the applied bias, the capacitance of the MOS-C varies with the applied bias under some bias conditions.

Note, however, that if the value of the positive gate bias becomes sufficiently large, a large number of minority carrier electrons will be attracted to the surface. They will collect there and come to dominate the negative charge in the substrate. This induced charge due to the electrons at the surface is called an *inversion layer*. Note also that these phenomena will be described in more detail in the upcoming section, but the purpose of briefly introducing them here is to point out the fact that MOS-Cs indeed behave in a somewhat different fashion than conventional capacitor structures.

3.1.1.2 Why is the MOS-C Studied First? As illustrated above, the MOS-C forms the heart of the MOSFET. Yet it is a fundamental structure in its own right, just as the *pn* junction is a fundamental structure apart from the bipolar transistor. Yet, much in the same way that understanding *pn* junctions is central to understanding the operation of the bipolar transistor, an understanding of the MOS-C is central to understanding of MOSFETs. Hence, it seems logical that the behavior of MOS-Cs should be studied first, before considering the physics of the MOSFET.

We can see this by noting that the basis of MOSFET behavior essentially involves two phenomena: (1) the action of mobile carriers in the semiconductor in response to the application of a gate bias; and (2) the transport of these mobile carriers between the source and drain (in the form of a drain current, I_D) in response to the application of a drain bias. Therefore, to quantitatively describe MOSFET operation, initially we need to establish a quantitative model which allows the first of these two phenomena to be treated (i.e., we need to be able to determine the mobile charge distribution in a MOS structure for a given gate bias). Once this charge distribution is known, it can be used to develop a model for the relationship between I_D and the terminal voltages (i.e., the gate and drain biases). Such a model for mobile-charge versus gate-bias is most easily developed in MOS-C structures for the following reasons:

1. We noted that the MOS capacitor is a two-terminal device. If we assume the oxide is a perfect insulator, then the current through the oxide I_G is zero for dc biases between the gate and substrate. Since this is the only possible current in the MOS-C - and its value in ideal structures is zero - the redistribution of charges in the semiconductor with respect to changes in V_{GB} can be analyzed without the complication of current flow. Furthermore, since no current flows under dc bias, the behavior of MOS-Cs in such cases represents an electrostatic problem. Thus, only one of the five classical device equations introduced in chap. 1 needs to be solved (i.e., the Poisson equation) to completely characterize the potential and mobile charge distribution in the ideal MOS-C (i.e., even if $V_{GB} \neq 0$).

2. The electric field in the MOS-C is perpendicular to the silicon surface, making it essentially uni-directional (vertical). The electrostatic potential in the device can thus be found by solving the Poisson equation in its least complex (e.g., one-dimensional) form. On the other hand, if a bias is also applied between the source and drain of a MOSFET, the direction of the electric field in the semiconductor region of the MOSFET is no longer perpendicular to the silicon surface. That is, it generally contains components parallel as well as vertical to the silicon surface. Hence, to calculate the electrostatic potential in a MOSFET, it is in general necessary to solve the Poisson equation in its two- or three-dimensional form.

3. The absence of current flow in a MOS-C implies that equilibrium is maintained in the semiconductor region even if a dc V_{GB} is applied.* Thus, the equations used to calculate carrier concentrations in terms of equilibrium concepts (e.g., the Fermi level) can be used. This makes the solution of the Poisson equation simpler as well.

4. The MOS-C is widely used as the test structure for characterizing the reliability of the gate oxide film with respect to oxide breakdown and hot-electron degradation (as described in chaps. 7 and 9, respectively). Thus, the study of the MOS-C behavior is intrinsically useful for interpreting data obtained from such structures.

5. The quantitative relationships we develop for the mobile charge concentration in the channel region of the MOS-C under inversion are used in the I_D models of the long-channel MOSFET by invoking the *gradual-channel approximation.* This approximation will be examined in more detail in chap. 4.

3.2 QUALITATIVE DESCRIPTION OF MOS-C DEVICE BEHAVIOR

The behavior of the MOS-C from a qualitative perspective is considered first, namely a verbal description is given of how the mobile charge distribution in the semiconductor depends on the applied dc gate bias. To help visualize these phenomena we make use of energy-band and potential representations, as well as electric-field and charge distribution diagrams.

* The term *equilibrium* should be considered here to mean *thermal equilibrium,* as viewed from the microscopic perspective. That is, the net energy in the semiconductor is stored in the crystal-lattice vibrations (phonons) as well as by free carrier motion. Although external sources of energy (such as incident photoelectric radiation) may also excite a crystal, many situations exist where no such external forces act. In such cases, the total energy is a function of only the crystal temperature, and the semiconductor will spontaneously reach a state referred to as thermal equilibrium. In thermal equilibrium every process is exactly balanced by its inverse process. One such set of competing processes in a semiconductor involves carrier generation and recombination. At thermal equilibrium the rates of these two processes are equal (and are also assumed to be driven only by thermal processes). This makes it valid to use the equilibrium carrier concentration expressions (i.e., Eqs. 3-1a and 3-1b) to calculate the carrier concentrations in the MOS-C even when $V_{GB} \neq 0V$.

3.2.1 Energy-Band Diagram Representation of Semiconductor Devices

One of the most helpful tools for visualizing semiconductor phenomena is the *energy-band diagram.* In such diagrams, the vertical axis represents the energy possessed by electrons when they occupy energy levels at a given location in the device. (In one-dimensional representations, a location in the device is generally specified by a point x, which is designated as the distance along the horizontal axis of the diagram). Such one-dimensional representations are adequate for treating the MOS-C (with the *x*-direction representing the direction perpendicular to the silicon surface).

In the energy-band representation of a semiconductor (Fig. 3-2), E_c and E_v are the conduction and valence band edges, respectively. The intrinsic Fermi level E_i is at the middle of the band gap (i.e., $E_i \cong [E_c+E_v]/2$), corresponding to the Fermi-level of an intrinsic (undoped) semiconductor (Fig. 3-2a). In *n*-type doped silicon the Fermi level E_F is closer to E_c than to E_v (Fig. 3-2b), but in *p*-type it is closer to E_v (Fig. 3-2c) If an electron occupies a state at the bottom edge of the conduction band (E_c), its kinetic energy is defined to be zero (point A in Fig. 3-2d). As the kinetic energy of such an electron increases, it occupies energy levels higher in the conduction band (e.g., point B in Fig. 3-2d). If a semiconductor is in equilibrium, the Fermi level is invariant with position in the material, and this is depicted in a one-dimensional energy-band-versus-distance diagram as E_F being a straight, horizontal line (as in all of Fig. 3-2).

Assuming Maxwell-Boltzmann statistics, the electron and hole concentrations in a semiconductor at equilibrium (n_o and p_o, respectively) can be expressed in terms of quantities found in the energy-band diagram (E_F, E_i), as well as T, and n_i (the intrinsic carrier concentration at T) according to the following expressions:

$$n_o = n_i \exp (E_F - E_i)/kT \qquad (3 - 1a)$$

$$p_o = n_i \exp (E_i - E_F)/kT \qquad (3 - 1b)$$

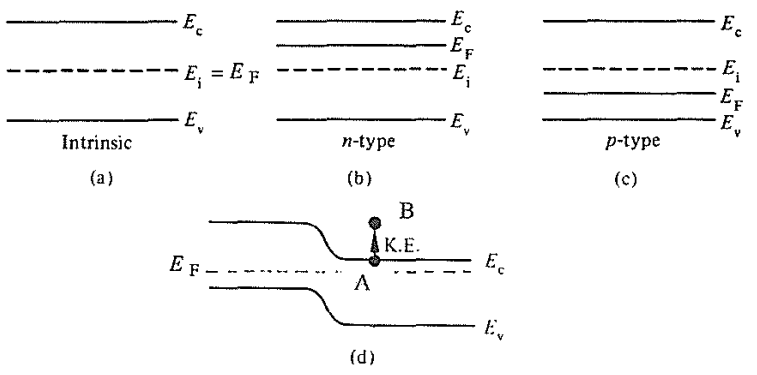


Fig. 3-2 Energy band representation of a semiconductor: (a) intrinsic; (b) *n*-type; (c) *p*-type; (d) Depicting the kinetic energy of an electron in the energy band diagram.

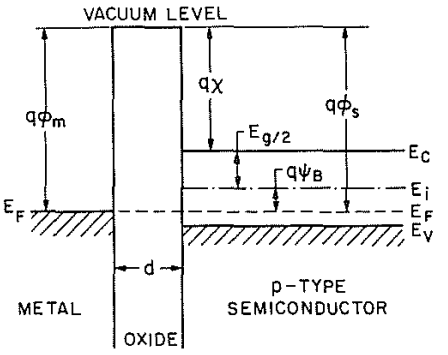


Fig. 3-3 Energy band diagram of an ideal MOS-C when $V_{GB} = 0V$ (*flat-band condition*).

Note also that at equilibrium

$$n_o p_o = n_i^2 \qquad (3 - 1c)$$

3.2.2 Energy Band Diagram of the Ideal MOS Capacitor

Using the energy-band diagram representation we now consider the qualitative behavior of an *ideal* MOS-C subjected to dc bias. The following assumptions are assumed to be valid in an ideal MOS-C:

1. The energy difference between the work function of the gate material (metal), $q\varphi_m$ and that of the semiconductor, $q\varphi_s$ is zero (i.e., $q\varphi_{ms} = 0$);
2. No charges exist in the MOS-C under any biasing conditions, other than those existing in the semiconductor and those with equal but opposite sign on the gate, i.e., within the oxide there are no mobile charge carriers nor any charge centers;
3. No current can flow through the oxide under dc-biasing conditions ($I_G = 0$);
4. The semiconductor region is uniformly doped;
5. The gate layer is considered to be an equipotential region.

For our example we will consider an ideal MOS-C with a *p*-type semiconductor region.

First consider the case when no bias is applied ($V_{GB} = 0V$). The energy-band diagram for this case is shown in Fig. 3-3. We note that since equilibrium exists in the semiconductor and $q\varphi_{ms} = 0$ (Assumption #1), the Fermi level throughout the semiconductor must be invariant with position, and it must also line up with the Fermi level in the gate. This also implies that E_c, E_v, and E_i in the semiconductor are horizontal and straight (i.e., flat). An energy-band diagram in which E_c, E_v, and E_i are flat is said to be in a *flat-band condition. Thus, when $V_{GB} = 0$ in an ideal MOS capacitor, it's energy-band diagram is in a flat-band condition.* Note also that in *p*-type Si, E_F is below E_i, by an amount expressed as $|E_F - E_i|$.

Next let a dc bias be applied between the gate and substrate. Since $I_G = 0$ (Assumption #3), the silicon region remains in equilibrium, implying that E_F remains flat. This aspect of MOS-C behavior is unique, since the application of a dc bias across the electrodes of other semiconductor devices will cause at least some current to flow. Hence, MOS-C analysis remains simple under dc bias because the Fermi-level concept is still valid. Although the semiconductor region remains in equilibrium under dc bias (and E_F remains flat), the other bands are altered from that of the $V_{GB} = 0V$ case. This is seen in Fig. 3-4 where four different gate-bias conditions are depicted. Here in Fig. 3-4a we depict the MOS-C in cross-section, as was also shown in Fig. 3-1b2. This structure is then rotated as shown in Fig. 3-4b, and the energy-band and charge distribution versus x diagrams we show next refer to the MOS-C cross-section oriented in the manner of Fig. 3-4b.

3.2.2.1 Accumulation. In the first biasing case a *negative* bias is applied ($V_{GB}<0V$, Fig. 3-4c). The negative bias causes holes (the majority carriers in p-type Si) to be attracted to the silicon surface adjacent to the oxide. More holes accumulate near the Si surface than if $V_{GB} = 0V$. The effect of attracting additional majority carriers to the semiconductor surface is called *accumulation*. The conductivity of the region near the Si surface is increased by the presence of the extra holes.

Note that in the energy-band representation, extra holes at one location of a semiconductor (e.g., in this case, at the Si surface) implies that E_F must be closer to E_V than in a region containing fewer holes (e.g., the region deeper below the surface [bulk region]). However, since the *Fermi level* in the semiconductor must remain flat, the energy bands in the semiconductor (E_C, E_V, and E_i) near the surface must bend upward to reflect the presence of the extra holes at that location compared to the *bulk* (Fig. 3-4c1)

The applied gate bias also causes the Fermi levels in the gate and semiconductor regions to be separated by an amount equal to qV_{GB}, or

$$E_F(\text{gate}) - E_F(\text{semiconductor}) = -qV_{GB} \qquad (3 - 2)$$

Although the capacitance of the MOS-C varies with applied dc voltage under some bias conditions, in the accumulation mode it does not. That is, in this mode the capacitance per unit area remains constant as V_{GB} is changed, and is given by $C_{ox} = \kappa_{ox}\varepsilon_o/t_{ox}$, where κ_{ox} and t_{ox} are the dielectric constant and thickness of the oxide, and ε_o is the permittivity of free space.

We now introduce the *charge-distribution versus distance* diagram as an additional visualization tool. This allows us to depict the approximate charge distribution inside the MOS structure. For the case of accumulation, applying a negative bias causes more electrons to move onto the gate. Holes are attracted toward the silicon surface in response. Thus, the charge inside the device as a function of position can be approximated as shown in Fig. 3-4c2. Note that a squared-off (or block) representation of the charge is used in this simple plot. Such block approximations are only intended to be a qualitative representation of the actual charge distribution, which, in general, is not distributed so simply. When interpreting the magnitude and spatial extent of the charge using such simplified diagrams this fact should be kept in mind.

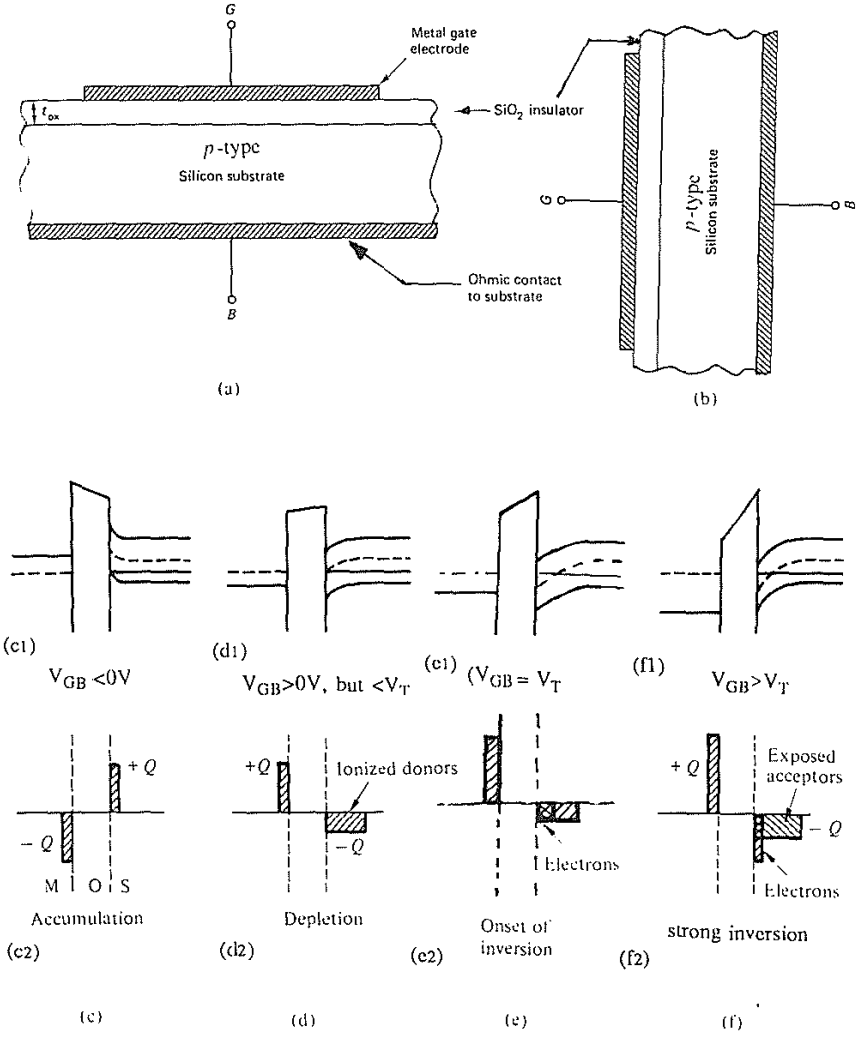


Fig. 3-4 (a) Cross-section of a MOS-C with p-substrate; (b) Cross-section of the same MOS-C, but rotated 90°; (c1) and (c2) Energy band and block charge diagrams when the MOS-C is biased to accumulation ($V_{GB}<0V$); (d1) and (d2) Same diagrams for depletion biasing ($V_{GB}>0V$, but $<V_T$); (e1) and (e2) Same diagrams at onset of inversion ($V_{GB} = V_T$); (f1) and (f2) Same diagrams at strong inversion ($V_{GB}>V_T$).

3.2.2.2 Depletion. In the second bias condition (Fig. 3-4d), a *small positive* voltage is applied to the gate of the *p*-type MOS capacitor. This causes holes in the semiconductor to be pushed away from the surface, creating a depletion region of width d, consisting primarily of negatively charged acceptor ions. In the energy-band representation of this phenomenon the bands bend downward near the surface (Fig. 3-4d1). In this mode of operation, known as *depletion,* the MOS capacitor exhibits two capacitances in series - the oxide capacitance per unit area C_{ox} and the depletion capacitance per unit area $C_d = \kappa_{Si}\varepsilon_o/d$. While C_{ox} is constant, C_d is not. That is, C_d decreases as V_{GB} is made larger (more positive). The approximate charge distribution for this bias condition is shown in Fig. 3-4d2.

3.2.2.3 Inversion. If the positive voltage applied to the gate is increased, the bands are bent further downward, eventually causing E_F to cross over E_i and to be closer to E_c than to E_v at the surface. At this point the semiconductor surface is *inverted* (i.e., it has been changed from *p*-type to *n*-type) because there are now more electrons than holes at the surface.*

Another way to view this situation is to remember that the kinetic energy of an electron is zero when it occupies an energy state at the conduction band edge (corresponding to an energy E_c). Thus, under depletion or inversion biasing electrons near the Si surface having zero kinetic energy occupy energy states of lower *total* energy than electrons deeper in the Si having zero kinetic energy. Hence, when a positive bias is applied to the gate, electrons are more likely to fill such states near the surface. Consequently, electrons will be found at higher concentrations closer to the surface when $V_{GB}>0V$.

3.2.2.4 Strong Inversion. If V_{GB} is made even more positive, the bands will continue to bend further downward, until E_i at the surface is eventually below E_F by an energy equal to $|E_F - E_i|$. At this point the electron density per unit volume at the surface n_{surf} (located in a thin layer known as the *inversion layer*) is equal to the hole density in the bulk. This is known as the point of *onset of strong inversion*, and in this condition

$$E_i(\text{bulk}) - E_i(\text{surface}) = 2[E_i(\text{bulk}) - E_F] \qquad (3 - 3)$$

Solving this equation for E_i(surface), and using the result in Eq. 3-1a, the concentration of electrons at the Si surface at the onset of strong inversion, n_{surf}(inv) is

$$n_{surf}(\text{inv}) = n_i \exp[E_F - E_i(\text{surface})]/kT = n_i \exp[E_i(\text{bulk}) - E_F]/kT = p_{obulk} \approx N_A \quad (3 - 4)$$

Figure 3-4e1 is the energy-band diagram depicting the condition of onset of strong

* The surface is actually inverted as soon as E_F crosses over E_i. However, the density of electrons at the surface remains small until E_F is considerably greater than E_i. The mode of operation when the electron density is smaller than the hole density in the substrate bulk is referred to as *weak inversion.* Weak inversion is important in submicron MOSFET behavior and will be discussed in detail in section 4.7.

inversion. In this diagram the width of the depletion region is denoted by d_{max}. The block charge diagram at the onset of strong inversion is shown in Fig. 3-4e2. Note that the voltage applied between the gate and substrate at the onset of strong inversion is known as the threshold voltage, V_T.

For further increases in positive bias, the electron concentration in the inversion layer at the surface exceeds the bulk equilibrium hole concentration, $p_o = N_A$. This is referred to as *strong inversion.* The energy-band and block-charge diagrams depicting the strong inversion condition are shown in Figs. 3-4f1 and 3-4f2, respectively.

In general, the charges on the gate Q_G of a MOS-C must be balanced by the charges in the semiconductor Q_C

$$Q_G\ (C/cm^2) = -Q_C\ (C/cm^2) \qquad (3 - 5)$$

In strong inversion, the charge per unit area Q_C in a *p*-type semiconductor is the sum of the electrons in the inversion layer (Q_I) and the negative charges in the depletion region (Q_{Bmax}). Thus, in strong inversion

$$Q_G = -Q_C = -(Q_I + Q_{Bmax}) \qquad (3 - 6)$$

where $Q_{Bmax} = -qN_A d_{max}$. (Note that Q's in this chapter are expressed in units of charge/unit area - specifically, C/cm^2.) The total capacitance per unit area at strong inversion is the oxide capacitance C_{ox} in series with the depletion capacitance at the maximum depletion width in equilibrium $C_{dmax} = \kappa_{Si}\varepsilon_o/d_{max}$. While the above discussion applies to a MOS-C with a *p*-type semiconductor, similar results can be obtained for an *n*-type semiconductor, with the polarity of the voltages and charges being reversed. A typical value of Q_I in a MOS-C or MOSFET in strong inversion is on the order of 10^{12} cm^{-2}.

3.2.2.5 Microscopic Perspective of the Mobile Charge Behavior in a MOS-C. It is also useful to consider from a (qualitative) microscopic perspective the phenomena involving charge carriers near the silicon surface of the MOS-C under conditions of depletion and inversion biasing. That is, when the MOS-C is thus biased, a potential difference between the surface and the bulk exists, giving rise to a potential energy barrier (PEB) with respect to the majority carriers. The majority carriers (holes in our example) are thereby repelled from the surface, and minority carriers (i.e., electrons in this example) are attracted toward it. In either inversion or depletion a region largely devoid of carriers is established (depletion region), and in this region thermal generation of carriers continuously produces holes and electrons in an attempt to re-establish the mobile charge concentration values present in the charge-neutral regions at equilibrium. The holes so produced, however, are continuously repelled from the surface, while the generated electrons are likewise drifted toward the surface.

In a state of thermal equilibrium, some holes in the bulk nevertheless still possess sufficient energy to overcome the PEB at the surface, and they are able to penetrate the depletion region so that recombination with the electrons attracted to the surface can occur. At thermal equilibrium, the rate of such recombination at the surface exactly equals the rate of electron generation in the depletion region, and the electron and hole

concentrations are thus maintained at a fixed, specific value, dependent on the temperature, doping concentration, and gate bias.

If the gate bias is instantaneously changed from one fixed value to another, the carrier concentrations must also change to reflect this new bias value. However, since an increase in the electron concentration relies essentially on their generation in the depletion region by thermal processes, some time must elapse before the new equilibrium electron concentration can be established. In fact, the response of a silicon MOS-C due to instantaneous changes in V_{GB} is relatively slow because thermal generation and recombination processes in the silicon substrate occur at low rates at room temperature. The situation that exists in a MOSFET, as we will describe in chap. 4, is quite different. That is, the presence of diffused regions of doping type opposite that of the substrate (i.e., the source/drain regions) permit electrons (under favorable bias conditions) to be injected into (or removed from) this surface region much more rapidly than if only thermal processes are responsible for contributing carriers (see section 4.2).

3.3 QUANTITATIVE DESCRIPTION OF MOS-C BEHAVIOR

We now consider the behavior of the ideal MOS-C from a quantitative perspective. Specifically, we seek to obtain mathematical relationships that permit us to calculate the following parameters in terms of the applied dc gate bias V_{GB}, the device structure parameters (i.e., t_{ox} and the doping concentration in the silicon), and the material parameters (i.e., ε_{si} and ε_{ox}):

1. The electrostatic potential as a function of position $\varphi(x)$;
2. The surface potential, φ_{surf}, which is the potential at $x = 0$, or $\varphi(0)$
3. The electric field as a function of position $\mathcal{E}(x)$;
4. The mobile charge concentrations $n(x)$ and $p(x)$ in the silicon;
5. The bulk charge per unit area Q_B in the silicon;
6. The inversion layer charge per unit area Q_I

Later, we will treat the nonideal MOS-C in which we will also include the effects due to the properties of the gate material and the charges in the oxide and at the $Si\text{-}SiO_2$ interface. We will also describe how to calculate the threshold voltage of MOS structures.

Three quantitative models of MOS-C behavior will be developed, differing in their complexity, accuracy, and generality. The simplest MOS-C model, based on the *depletion approximation* will be derived first. It is also the least general and least accurate of the three models, and provides accurate values of the mobile charge concentrations in the silicon only beyond strong inversion. Nevertheless, this model offers easy-to-interpret analytical expressions of MOS-C behavior, and is the model developed in most basic device physics texts. It is also the basis of the simpler long-channel MOSFET models described in chap. 4.

The second MOS-C model we derive is the most general and accurate model (in that it is valid for the accumulation, depletion, and inversion regimes of MOS-C operation), but it provides these benefits only at the expense of much greater mathematical complexity. That is, the second model does not provide analytical expressions for the quantities listed above, but requires numerical analysis for specific values to be obtained. Despite its mathematical complexity, this model is presented for several important reasons:

1. The MOS-C parameters listed above when calculated using this model are so accurate compared to experimental measurements that the model represents a benchmark to which the results of the other models can be compared to establish their accuracy.

2. It is the basis for the most general long-channel MOSFET model we present in chap. 4 (i.e., the Pao-Sah model).

3. It illustrates the principle that numerical analysis must be relied upon to obtain the most accurate quantitative representations even for the cases of the MOS-C and the long-channel MOSFET.

4. It implies that numerical analysis must certainly be employed to obtain accurate predictions of submicron MOSFET behavior.

However, the complexity of this model is such that it is not feasible to use it for constructing models of MOSFET circuit behavior, nor does it provide analytical expressions for the MOS-C quantities listed above. Thus, quantitative insights about MOS-C behavior are difficult to extract from this model, making it a poor tool for learning about MOS-C behavior. As a result, the previous simple model and a third model (which we describe last), find much wider use than this one.

The third MOS-C model is intermediate in accuracy, generality, and complexity between the other two models. It is referred to as the *charge sheet model,* and it calculates the characteristics of the MOS-C quite accurately over the entire range of inversion (i.e., from weak to strong inversion). It requires more computational effort than the simple model, but far less than the general model. Unlike the general model it provides analytical expressions for the mobile charge density which also extend beyond the operating regimes treated by the simple model. These expressions provide quantitative insight about the behavior of device operating regions that become more important in short-channel MOSFETs.

Note that we will designate the externally applied (i.e., terminal) voltages with the symbol V (e.g., V_{GB}), while the voltage drops internal to the MOS structure will be expressed with the symbol φ (e.g., the voltage drop across the oxide φ_{ox} and the voltage drop across the semiconductor from the Si surface to the bulk, φ_{surf}). Note, however, that both the external and internal voltage drops have units of volts.

3.3.1 The Electrostatic Potential in MOS Structures

Before setting out to derive the expressions listed above, let us make some introductory remarks about the *electrostatic potential* φ. First, we note that since the bias voltages applied to MOS devices are expressed in units of volts, it would be preferable if the quantitative MOS-C relationships we seek were also expressed in the same units (rather than in terms of units of energy, which would result if we used the energy-band representation of MOS behavior). In addition, one of the five classical semiconductor device equations is the Poisson equation, and its solution yields the electrostatic potential in the device. Hence, two of the relationships we obtain when solving the 1-D Poisson equation in the semiconductor region of a MOS-C involve the electrostatic potential, i.e.,

a. The relationship between V_{GB} and the *surface electrostatic potential* φ_{surf} (henceforth referred to as the *surface potential*), and;

b. An expression for the electrostatic potential versus position $\varphi(x)$ [which uses φ_{surf} obtained from **a** as one of the boundary values for calculating this function].

It turns out that the electrostatic potential at any location in a MOS device is related to the potential energy an electron possesses at that location. That is, the electrostatic potential-difference $\Delta\varphi$ observed in moving from one position to another in the semiconductor (assuming the kinetic energy of the electron is the same at both positions) can be computed (in terms of an electron) from

$$\Delta\varphi = \Delta E/(-q) \qquad (3 - 7)$$

where q is the magnitude of the electron's charge. It is evident from Eq. 3-7 that if an electron undergoes a *potential energy* decrease, it means that a *potential* increase exists between two locations, because of the algebraic sign of the electron charge. Thus, for the case of an electron, an energy-band diagram can be converted to a potential diagram by merely inverting one of the bands E_c, E_i, or E_v (as shown in Fig. 3-7, which depicts both representations).* Note that *downward* band bending while going, for example, from the bulk to the surface in a MOS structure, implies an *increase* in the electrostatic potential when moving between those same locations.

Since the potential is a relative quantity, the zero value of φ can be arbitrarily chosen. In our discussion we select the zero value as the electrostatic potential in the bulk (i.e., at $x = \infty$). We also choose the energy level of the intrinsic Fermi level E_i in the bulk [E_i(bulk)] to designate this level of zero on an energy-band diagram. Thus, elsewhere, the value of $\varphi(x)$ is measured with respect to E_i(bulk) according to

$$\varphi(x) = [E_i(\text{bulk}) - E_i(x)]/q \qquad (3 - 8)$$

* Note that since an electron at E_c will have the same kinetic energy (zero) at any location, the bending of the E_c band directly indicates the change in potential energy versus location.

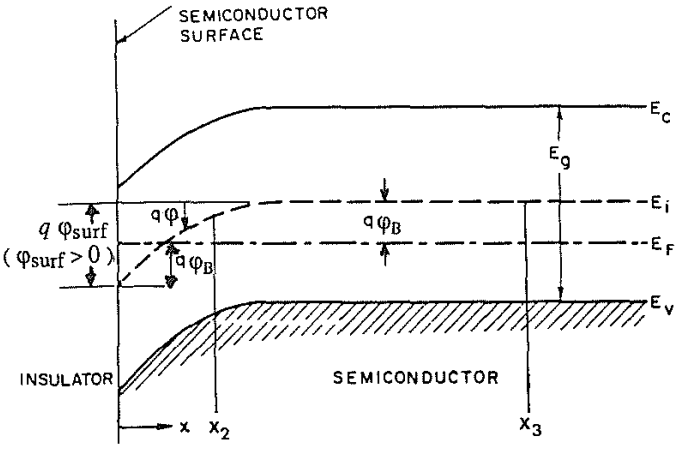


Fig. 3-5 The energy-band and potential diagram representations of the MOS-C (in this figure a MOS-C with a *p*-type substrate and biased to inversion is depicted).

as depicted in the figure. For example, in the bulk of the semiconductor (position x_3 in Fig. 3-5), $E_i(x_3) = E_i(\text{bulk})$. Hence, from Eq. 3-8 the potential at x_3 is 0. At x_2, however, $\varphi(x_2) \neq 0$, but instead is $\varphi(x_2) = [E_i(\text{bulk}) - E_i(x_2)]/q$.

Another consequence of choosing $\varphi\ (x = \infty) = 0$ is this: If we apply a bias to the gate of a MOS-C, part of the voltage will be dropped across the oxide φ_{ox} and part across the silicon φ_{si} , or,

$$V_{GB} = \varphi_{ox} + \varphi_{si} \qquad (3\text{-}9)$$

If $\varphi(x = \infty) = 0$, then the potential at the surface of the silicon (φ_{surf}) represents the voltage drop from the bulk ($x = \infty$) to the silicon surface ($x = 0$) in a MOS structure. In that case we can write Eq. 3-9 as

$$V_{GB} = \varphi_{ox} + \varphi_{surf} \qquad (3\text{-}10)$$

Now we see that if we can derive an expression for φ_{ox} in terms of φ_{surf}, an equation relating V_{GB} and φ_{surf} results. We show how this is done in the next section.

Let us also introduce another dimensionless parameter called the *doping parameter* φ_B which relates the potential of an electron at the Fermi level (in the bulk of the silicon in a MOS structure) to the doping concentration. The doping parameter is defined as

$$\varphi_B = [E_i(\text{bulk}) - E_F]/q \approx (kT/q)\ln(N_A/n_i)_{p\text{-type Si}} \approx (kT/q)\ln(n_i/N_D)_{n\text{-type Si}} \qquad (3\text{-}11)$$

in the bulk region of the Si in the MOS capacitor, see Fig. 3-5). Note that $\varphi_B>0$ for a *p*-type semiconductor and $\varphi_B<0$ for an *n*-type semiconductor. For substrate doping concentrations usually encountered in MOSFETs (10^{14}-10^{17} cm^{-3}), the magnitude of φ_B is between $9kT/q$ and $16kT/q$ (where kT/q at 300K is ~0.025 V); that is, the magnitude of φ_B is between 225 and 400 mV at 300K for this range of doping concentrations.

Also, the value of the surface potential at the onset of weak and strong inversion can be expressed in terms of φ_B. That is, in section 3.2 we defined the onset of *weak inversion* as the condition when the electron concentration at the surface of *p*-type Si equals n_i. At that point, $\varphi_{surf} = \varphi_B$. Since the onset of *strong inversion* is the condition when the surface electron concentration equals N_A, at that point E_i (surf) - E_i (bulk) = 2 [E_F - E_i (bulk)]. By combining this fact with Eqs. 3-4 and 3-11, we see that the surface potential at the onset of strong inversion $\varphi_{surf}(inv) = 2\varphi_B$.

Note that it is also possible to express the equilibrium concentrations of electrons n_p (x) and p_p (x) in terms of $\varphi(x)$. That is, when the zero value of φ is defined as above

$$n_p(x) = n_{po} \exp[q\varphi(x)/kT] \qquad (3 - 12a)$$

$$p_p(x) = p_{po} \exp[-q\varphi(x)/kT] \qquad (3 - 12b)$$

where n_{po} and p_{po} are the equilibrium concentrations of electrons and holes in the bulk of the silicon (in this case in *p*-type Si). At the surface of the silicon, the concentrations n_{surf} and p_{surf} are given by

$$n_{surf} = n_{po} \exp(q\varphi_{surf}/kT) \qquad (3 - 13a)$$

$$p_{surf} = p_{po} \exp(-q\varphi_{surf}/kT) \qquad (3 - 13b)$$

Using Eq. 3-11 in Eq. 3-13a, we can express the electron concentration at the surface in terms of N_A and φ_B as:

$$n_{surf} = N_A \exp[q(\varphi_{surf} - 2\varphi_B)/kT] \qquad (3 - 13c)$$

This expression is plotted versus φ_{surf} in Fig. 3-6. This figure reiterates several aspects of MOS behavior established with the energy-band representation of MOS characteristics. First, using Eq. 3-11 we see that when $\varphi_{surf} = \varphi_B$, n_{surf} becomes equal to the intrinsic concentration, n_i. Thus, $\varphi_{surf} = \varphi_B$ is defined as the limit point between the regions of depletion and weak inversion. Second, as φ_{surf} increases above φ_B, n_{surf} grows very rapidly. Eventually, when $\varphi_{surf} = 2\varphi_B$, $n_{surf} = N_A$. This is commonly defined as the demarcation point between weak and strong inversion. Finally, beyond this point (i.e., $\varphi_{surf}>2\varphi_B$), n_{surf} continues to increase rapidly, becoming much greater than N_A for small increases of φ_{surf} beyond $2\varphi_B$. This implies that in strong inversion the inversion layer charge density Q_I is generally much larger than the ionic charge density in the depletion region Q_B. An additional important point can also be inferred from this figure. That is, if the potential as a function of x is known we can calculate $n_p(x)$ from Eq. 3-12a.

3.3.2 General Approach to Deriving MOS-C Characteristics

Now we can set out to derive the expressions listed at the beginning of section 3.3. This can be done in four stages, assuming we are given V_{GB}, the oxide thickness t_{ox}, and the doping concentration N_A.

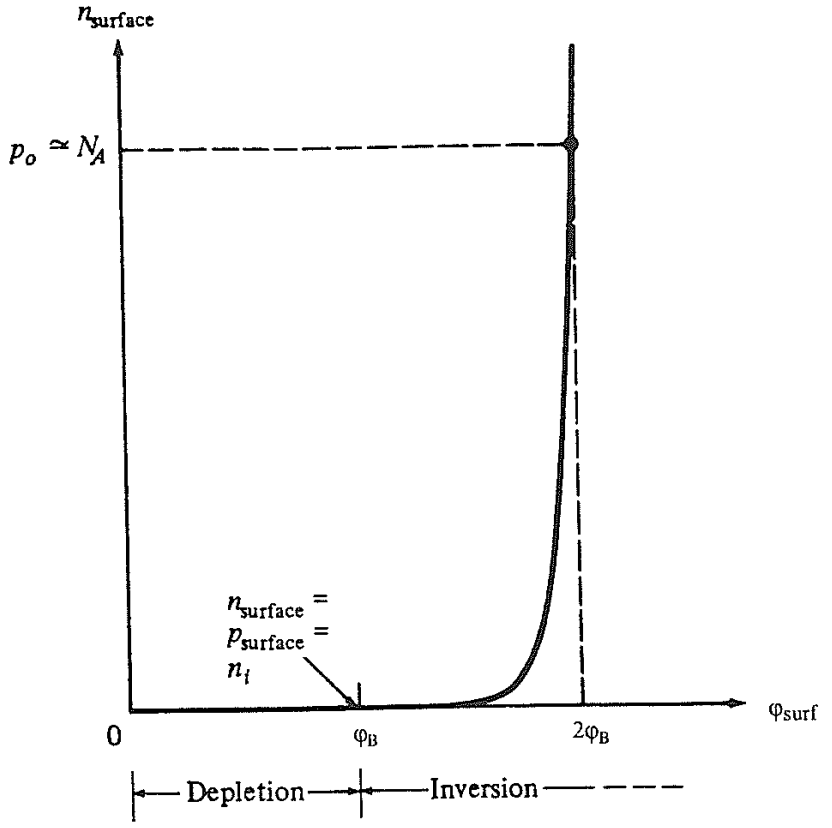


Fig. 3-6 Electron concentration at the surface of a MOS-C (with a *p*-type substrate) as a function of φ_{surf}.[6] Linear axes are used. Reprinted with permission of McGraw-Hill Book Co.

1. We first find a relationship between V_{GB} and φ_{surf} in the MOS structure. This is done by recalling that if we have an expression for φ_{ox} in terms of φ_{surf}, it can be substituted in Eq. 3-10 to yield an equation for φ_{surf} in terms of V_{GB}.

We begin the derivation of such an expression by noting that the oxide is considered to be an ideal insulator, with no mobile carriers or charge centers. Hence, the electric field in the oxide must be constant, or $\mathcal{E}_{ox} = - d\varphi_{ox}/dx =$ constant. Thus, the voltage across the oxide φ_{ox} is found from:

$$\varphi_{ox} = -\int_{-t_{ox}}^{0} \mathcal{E}_{ox}\, dx = t_{ox}\mathcal{E}_{ox} \qquad (3 - 14)$$

The next step is to relate $\mathcal{E}_{ox}$ to the electric field in the semiconductor. Since the charge per unit area located at the Si-SiO_2 interface of an ideal MOS-C is assumed to be zero, the electric displacement, $D = \varepsilon\mathcal{E}$ perpendicular to the interface, is continuous [i.e., $D_{ox} = D_{si}(x = 0)$]. The field in the oxide is therefore related to the field at the surface of the silicon $\mathcal{E}_{si}$ (x = 0) by the equation

$$\mathcal{E}_{ox} = (\kappa_{si}/\kappa_{ox})\, \mathcal{E}_{si}\ (x = 0) \qquad (3 - 15)$$

Now, using Eq. 3-15 in Eq. 3-14 we can write φ_{ox} as

$$\varphi_{ox} = t_{ox}\, (\kappa_{si}/\kappa_{ox})\, \mathcal{E}_{si}\ (x = 0) \qquad (3 - 16)$$

and this result can be substituted for φ_{ox} in Eq. 3-10. In order to complete our goal of finding an expression for φ_{surf} in terms of V_{GB} we see that an expression for $\mathcal{E}_{si}$ (x = 0) in terms of φ_{surf} is needed. Since $\mathcal{E}(x) = -\, d\varphi/dx$, such expression can be obtained by integrating the one-dimensional form of the Poisson equation once, using the appropriate total charge density $\rho(x)$, and the boundary conditions $\mathcal{E}\ (x = \infty) = 0$, and $\varphi\ (x = 0) = \varphi_{surf}$. We will carry out this integration for three cases in the upcoming sections.

2. We next solve the Poisson equation to find $\varphi(x)$. Since we have noted that the electric field is normal to the oxide surface in a MOS-C (neglecting fringing fields at the edges of the capacitor), we can consider the variation in electrostatic potential φ in a MOS-C to be one-dimensional (i.e., we assume $\partial\varphi/\partial y$ and $\partial\varphi/\partial z = 0$). Thus, it is only necessary to solve the Poisson equation in its one-dimensional form. We use the value of φ_{surf} from step **1** as one boundary condition [$\varphi\ (x = 0) = \varphi_{surf}$], and [$\varphi\ (x = \infty) = 0$] as the other, when we solve the Poisson equation to find $\varphi\ (x)$.

3. To calculate the mobile charge concentrations $n_p(x)$ and $p_p(x)$, we can use the values of $\varphi\ (x)$ found from step (2) in Eqs. 3-12.

4. To find expressions for Q_B and Q_I as a function of gate bias we proceed as follows:

a. First, we recall Gauss' law, which states that the charge contained in a volume equals the permittivity times the electric field emanating from the volume. Applying Gauss' law to a volume extending from the silicon surface to the field-free bulk region, which contains a total charge per unit area Q_C, we can write*

$$Q_C = -\, \kappa_{si}\varepsilon_o\, \mathcal{E}_{si}\ (x = 0) \qquad (3 - 17)$$

* Note that only the $\mathcal{E}$-field at the Si-SiO_2 interface (which represents the top surface of the volume in which the charge in the semiconductor exists) must be included in Eq. 3-17 because the $\mathcal{E}$-field does not emanate from any other surfaces of the volume. This is because: (1) the $\mathcal{E}$–field emanating from the bottom of the volume under consideration (i.e., at x = d) is assumed to be zero, and; (2) since the $\mathcal{E}$-field in the MOS-C is assumed to be unidirectional, $\mathcal{E}_y = \mathcal{E}_z = 0$. Thus, the $\mathcal{E}$-field emanating from the sides of the volume is also zero.

We also note from Eq. 3-16 and from the fact that $C_{ox} = \kappa_{ox}\varepsilon_o/t_{ox}$ that

$$\varphi_{ox} = t_{ox}\,(\kappa_{si}/\kappa_{ox})\,\mathcal{E}_{si}\,(x = 0) = \{\kappa_{si}\varepsilon_o\,\mathcal{E}_{si}\,(x = 0)\}/C_{ox} \quad (3 - 18)$$

Thus, we can also write

$$Q_C = -\varphi_{ox}\,C_{ox} \quad (3 - 19)$$

b. Next, we note that the charge per unit area in the semiconductor Q_C is the sum of the inversion layer charge per unit area Q_I and the ionic charge per unit area Q_B. So we can write

$$Q_C = Q_I + Q_B \quad (3 - 20)$$

In general, Q_I can be found by calculating the total mobile charge in the inversion layer (in our example, the electrons). In other words, to get Q_I, we could integrate Eq. 3-12a from x = 0 to the depth x_c at which the electron concentration becomes negligible, or

$$Q_I(x) = -q\int_0^{x_c} n(x)\,dx \quad (3 - 21)$$

Then, Q_B could be found by subtracting Q_I from Q_C.

However, this approach to calculating Q_I and Q_B is quite complex, in fact, it requires that $\varphi(x)$ be computed first, and then this result be used to integrate $n(\varphi)$. Both of these calculations, in general, require numerical integration. Although we shall outline this approach in more detail, we will also illustrate simpler methods for calculating Q_I and Q_B which still provide values accurate enough for many applications.

3.3.3 Finding the Quantitative Characteristics of a MOS-C when the Depletion Approximation is Invoked

Let us derive the quantitative characteristics of the ideal MOS-C under some assumptions that simplify the mathematical treatment outlined above, but still provide expressions that are accurate enough for many applications. The simplest of such quantitative analyses assumes that we deal with the ideal uniformly doped MOS-C, and this analysis is also the one carried out in most basic device textbooks. Here we invoke the following approximations: (1) the *depletion approximation* is assumed to be valid; (2) the *one-sided abrupt junction approximation* is assumed; and (3) other simplifying approximations are assumed which allow the depletion region width d, the surface potential φ_{surf}, and the inversion layer charge Q_I to be simply computed beyond the point of strong inversion.

The *depletion approximation* invoked in this case is based on the following simplifying assumptions:

1. The mobile charges in the depletion region are totally absent in the depletion region for values of V_{GB} smaller than those needed to cause inversion (i.e., $n = p = 0$ in the depletion region over this range of gate-bias values).

2. For values of gate-bias voltages beyond inversion, the depletion region is still totally devoid of mobile carriers, except in the inversion layer, which is assumed to be a very thin layer near the silicon surface.

The *one-sided abrupt-junction approximation* assumes that the depletion region extends to a depth d, beyond which the silicon returns abruptly to a condition in which the mobile carrier densities are again equal to their equilibrium values [i.e., at $x = d$, $p = N_A = p_{po}$, and $n = n_{po}$].

The simplifying assumptions dealing with the d (depletion region width), φ_{surf} (surface potential), and Q_I (mobile charges in the inversion layer) are the following:

1. Beyond strong inversion, additional positive bias on the gate simply draws more electrons into the inversion layer. This negative electron charge balances the additional positive gate charge, and the depletion-region width d no longer increases. That is, d remains fixed at the value attained at the onset of strong inversion, and this value is denoted as d_{max}.

2. Since the depletion region no longer increases once the inversion layer is formed, the bands of the energy band diagram are not bent any further with increasing positive bias. Hence, φ_{surf} also remains fixed (or is *clamped*) at the value it has at the onset of inversion, namely at $\varphi_{surf}(\text{inv}) = 2\varphi_B$.

3. Beyond strong inversion, the ionic charge per unit area Q_B contained in the depletion region remains unchanged, since $d = d_{max}$. Hence, we can compute Q_B beyond inversion simply from a knowledge of N_A and d_{max}. Also if V_{GB} is given, and we know φ_{surf} beyond strong inversion, we can compute φ_{ox} from Eq. 3-10, and from this we can obtain Q_C. Then, Q_I can be found by subtracting Q_B from Q_C.

Using these assumptions, we set out to find the desired expressions. We initially solve the Poisson equation in the semiconductor region for $\varphi(x)$. That is, under depletion biasing conditions we have assumed $n = p = 0$, and so the charge density $\rho(x)$ is merely a constant term ($q\, N_A/\kappa_{si}\varepsilon_o$). In this case, the Poisson equation has the form:

$$\begin{aligned} d^2\varphi/dx^2 &= q\, N_A/\kappa_{si}\varepsilon_o \qquad & 0 < x < d \\ &= 0 & x > d \end{aligned} \qquad (3\text{-}22)$$

This is an ordinary, linear differential equation that can easily be integrated twice to yield a closed-form solution of $\varphi(x)$ in the semiconductor. That is, assuming the boundary conditions $\varphi(x = d) = 0$, and $\varphi(x = 0) = \varphi_{surf}$ we get,

$$\varphi(x) = \varphi_{surf}(1 - x/d)^2 \qquad (3\text{-}23)$$

where d is the depletion-region depth. The electric field $\mathcal{E}$ (x) is given by $-d\varphi/dx$ or

$$\mathcal{E}(x) = -d\varphi/dx = 2\varphi_{surf}(1 - x/d)(-1/d) \qquad (3-24)$$

and

$$d^2\varphi/dx^2 = 2\varphi_{surf}/d^2 \qquad \text{[from (3 - 23)]} \qquad (3-25)$$

$$= q N_A/\kappa_{si}\varepsilon_o \qquad \text{[from (3 - 22)]} \qquad (3-26)$$

Then from (3 - 25) and (3 - 26) we can find d in terms of φ_{surf} and N_A

$$d = [\kappa_{si}\varepsilon_o(2\varphi_{surf})/qN_A]^{1/2}. \qquad (3-27)$$

Beyond the point of strong inversion $\varphi_{surf}(inv) = 2\varphi_B$, and so d_{max} is then

$$d_{max} = (4\kappa_{si}\varepsilon_o\varphi_B/qN_A)^{1/2} \qquad (3-28)$$

Next, we solve Eq. 3-10 for φ_{surf}. In this case, let us initially assume that the MOS-C is biased into depletion, but below inversion. In that case, no inversion layer exists ($Q_I = 0$), thus, $Q_C = Q_B$ and Q_B is given by

$$Q_B = q N_A d \qquad (3-29)$$

Since $\varphi_{ox} = -Q_C/C_{ox} = -Q_B/C_{ox}$, by using Eq. 3-29 in this result, we get

$$\varphi_{ox} = -q N_A d/C_{ox} = \{[\kappa_{si}\varepsilon_o(2\varphi_{surf})qN_A]^{1/2}\}/C_{ox} \qquad (3-30)$$

Now if we let $\eta = \{[2\kappa_{si}\varepsilon_o qN_A]^{1/2}\}/C_{ox}$, we can write $\varphi_{ox} = \eta\sqrt{\varphi_{surf}}$. Substituting this into Eq. 3-10 we get

$$V_{GB} = \varphi_{surf} + \eta\sqrt{\varphi_{surf}} \qquad (3-31)$$

If we let $\alpha = \sqrt{\varphi_{surf}}$, we can write Eq. 3-31 as a quadratic equation

$$0 = \alpha^2 + \eta\alpha - V_{GB} \qquad (3-32)$$

We can now easily solve this for α, using the quadratic formula. Once we have α we can find φ_{surf}, since $\varphi_{surf} = \alpha^2$.

Now that φ_{surf} has been obtained, we can find Q_B (from Eq. 3-29) and $\varphi(x)$ (from Eq. 3-23). Below inversion we do not need to find n (x) and p (x), since for this case we have assumed them to be zero for x<d.

Beyond inversion we can calculate d_{max} from Eq. 3-28, and Q_{Bmax} from Eq. 3-29, (using $d = d_{max}$). Then, since $\varphi_{ox} = V_{GB} - 2\varphi_B$, we can find Q_C from

$$Q_C = -(V_{GB} - 2\varphi_B) C_{ox} \qquad (3-33)$$

Now, since $Q_I = Q_C - Q_{Bmax}$

$$Q_I = -[(V_{GB} - 2\varphi_B) C_{ox}] - q N_A d_{max} \qquad (3-34)$$

$$Q_I = -[(V_{GB} - 2\varphi_B)\, C_{ox}] + \sqrt{4\kappa_{si}\varepsilon_o\varphi_B q N_A} \qquad (3-35)$$

The results of this analysis are summarized in Fig. 3-7. Here the energy band diagrams of an ideal MOS capacitor (*p*-type Si substrate) are shown for the case of $V_{GB} = 0V$ (Fig. 3-7a). In Fig. 3-7b the case when sufficient gate voltage is applied to strongly invert the silicon surface is shown. Also depicted in these figures are the charge (Fig. 3-7c), $\mathcal{E}$-field (Fig. 3-7d) and electrostatic potential versus distance diagrams (Fig. 3-7e) for these two conditions. We note that since there are no charges in the oxide, the electric field is constant in the oxide, and that the slope of φ is thus constant in the oxide. In Fig. 3-7f we show the electrostatic potential versus distance diagram for an ideal MOS-C with an *n*-type substrate biased into strong inversion. It is useful to note that since a positive bias is needed to invert the *p*-type MOS-C, the φ versus x diagram curves upward, with the maximum value of φ at the surface. However, for the *n*-type MOS-C a negative gate bias is needed. Thus, in such cases the φ versus x diagram curves downward, with its minimum at the surface.

Another important point to which we will refer when discussing the so-called *gradual-channel approximation* (GCA) involves the gradient of the vertical electric field in the silicon, i.e., $\partial\mathcal{E}/\partial x$. In strong inversion $\partial\mathcal{E}/\partial x$ in the silicon is greatest in the inversion layer thickness, since $Q_I >> Q_B$ in strong inversion and the inversion layer is quite thin (~30-300Å thick). Hence, $\partial\mathcal{E}/\partial x$ can be estimated by approximating the vertical electric field value to be zero at the bottom edge of the inversion layer thickness t_{inv}. Then, $\partial\mathcal{E}/\partial x \sim [\mathcal{E}_{si}(x = 0) - 0]/t_{inv} = \mathcal{E}_{si}(x = 0)/t_{inv}$. We will make use of this result when examining the GCA in chap. 4.

3.3.4 More Accurate and General Model for MOS-C Behavior

The assumption that no mobile charges exist in the depletion region of the MOS-C (depletion approximation) does not accurately reflect the actual charge distribution condition in the silicon, especially below inversion. Thus, the expressions for $\varphi(x)$, φ_{surf}, $\mathcal{E}(x)$, and Q_I, obtained using this assumption are not accurate. In order to obtain a more accurate quantitative model of the MOS-C, we need to use the Poisson equation in which *p* and *n* have non-zero values. This will allow us to obtain more accurate expressions of the above quantities. As noted in the introduction of this section, this analysis is complicated. The complexity arises in part because it provides a model that is valid in all regimes (accumulation, depletion, and all ranges of inversion). Although the accumulation and depletion regions are not very important regimes of operation in most MOSFET applications, we nevertheless present this complex MOS-C model here for the several reasons listed in the introduction.

In this more accurate and general representation of the MOS-C, we treat the Poisson equation in the following form:

$$\frac{d^2\phi}{dx^2} = -\left(\frac{q}{\kappa_{Si}\varepsilon_o}\right)[-N_A + N_D + p - n] \qquad (3-36)$$

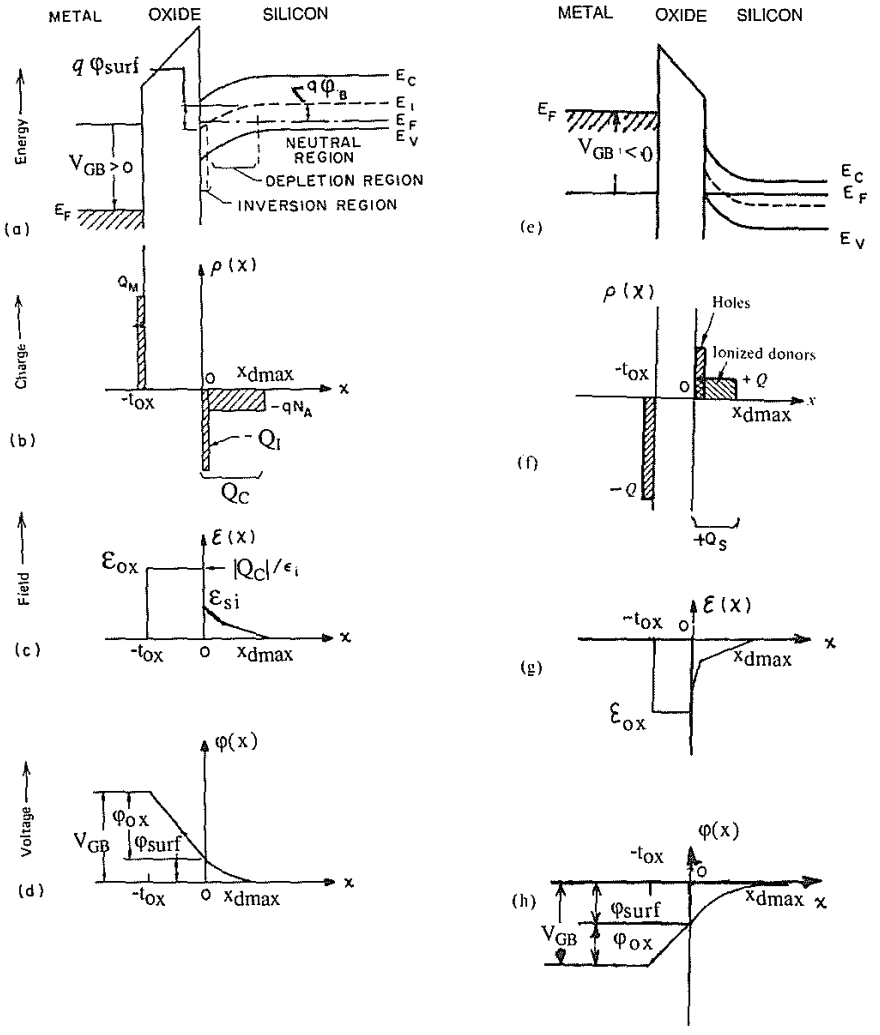


Fig. 3-7 Ideal *p*-type Si MOS-C when biased into inversion [$V_{GB}>V_T$]: (a) Energy-band diagram; (b) Charge distribution; (c) Electric field; and (d) Potential distribution; (e) - (h) Comparable diagrams for an ideal *n*-type Si MOS-C.

Equations 3-12a and 3-12b for $n(\varphi)$ and $p(\varphi)$ are substituted for n and p to give

$$\frac{d^2\phi}{dx^2} = -\left(\frac{q}{\kappa_{Si}\varepsilon_o}\right)\left[-N_A + p_{po}\, e^{-q\varphi/kT} + N_D - n_{po}\, e^{q\varphi/kT}\right] \qquad (3 - 37)$$

In general, Eq. 3-37 can be integrated once to give an analytic expression for the electric field, $\mathcal{E} = -\, d\varphi/dx$. This is accomplished by first multiplying both sides of Eq. 3-37 by $2(d\varphi/dx)$. The left-hand side is then recognized as $(d/dx)(d\varphi/dx)^2$, and we can write the result as

$$\frac{1}{2}\frac{d}{dx}\left(\frac{d\phi}{dx}\right)^2 = -\left(\frac{q}{\kappa_{Si}\varepsilon_o}\right)\left[-N_A + p_{po}\, e^{-q\varphi/kT} + N_D - n_{po}\, e^{q\varphi/kT}\right]\left(\frac{d\varphi}{dx}\right) \qquad (3 - 38)$$

Integrating this expression from the bulk ($x = \infty$, where $\varphi = 0$) to an arbitrary point x, we get

$$\left(\frac{d\phi}{dx}\right)^2_{\infty} - \left(\frac{d\phi}{dx}\right)^2_{x} = -\left(\frac{2q}{\kappa_{Si}\varepsilon_o}\right)\int_0^{\varphi}\left[-N_A + p_{po}\, e^{-q\varphi/kT} + N_D - n_{po}\, e^{q\varphi/kT}\right]d\varphi \qquad (3 - 39)$$

If we note that $p_{po} \approx N_A$, then $n_{po} = n_i^2/N_A$. In addition, since $\varphi_t = q/kT$, and $\mathcal{E} = -\, d\varphi/dx$, and $\mathcal{E}_{si}(x = \infty) = 0$, the first term on the left hand side of Eq. 3-39 is zero. Thus, after integrating the right-hand side we get

$$-\frac{d\varphi}{dx} = \mathcal{E}_{si}(x) = \sqrt{\frac{2q\,N_A}{\kappa_{Si}\varepsilon_o}}\sqrt{\varphi_t e^{-\varphi/\varphi_t} + \varphi - \varphi_t + e^{-2\varphi_B/\varphi_t}\left(\varphi_t e^{-\varphi/\varphi_t} - \varphi - \varphi_t\right)} \qquad (3 - 40)$$

This result allows us to obtain the electric field at the surface ($x = 0$) if φ_{surf} is known

$$\mathcal{E}_{si}(x = 0) = \sqrt{\frac{2q\,N_A}{\kappa_{Si}\varepsilon_o}}\sqrt{\varphi_t e^{-\varphi_{surf}/\varphi_t} + \varphi_{surf} - \varphi_t + e^{-2\varphi_B/\varphi_t}\left(\varphi_t e^{-\varphi_{surf}/\varphi_t} - \varphi_{surf} - \varphi_t\right)} \qquad (3 - 41)$$

In addition, by applying Gauss' law we can use Eq. 3-41 to find Q_C - i.e., from Eq. 3-17 [$Q_C = -\,\kappa_{Si}\varepsilon_o\mathcal{E}_{si}(x = 0)$] we get:

$$Q_C = \sqrt{2q\,N_A\kappa_{Si}\varepsilon_o}\sqrt{\varphi_t e^{-\varphi_{surf}/\varphi_t} + \varphi_{surf} - \varphi_t + e^{-2\varphi_B/\varphi_t}\left(\varphi_t e^{-\varphi_{surf}/\varphi_t} - \varphi_{surf} - \varphi_t\right)} \qquad (3 - 42)$$

This equation is valid for all regimes (accumulation, depletion, and inversion). Figure 3-8 depicts the variation of total charge concentration Q_C with φ_{surf} as calculated using Eq. 3-42.[6] It provides a quantitative representation of the total charge dependence on φ_{surf} that was qualitatively described in section 3.2.* We see that for negative values of φ_{surf}, Q_C is positive, corresponding to the accumulation regime. For $\varphi_{surf} = 0$, we

* Note that in section 3.3.1 we also examined the variation of the mobile charge concentration at the silicon surface n_{surf} versus φ_{surf} for depletion and inversion ($\varphi_{surf} > 0V$).

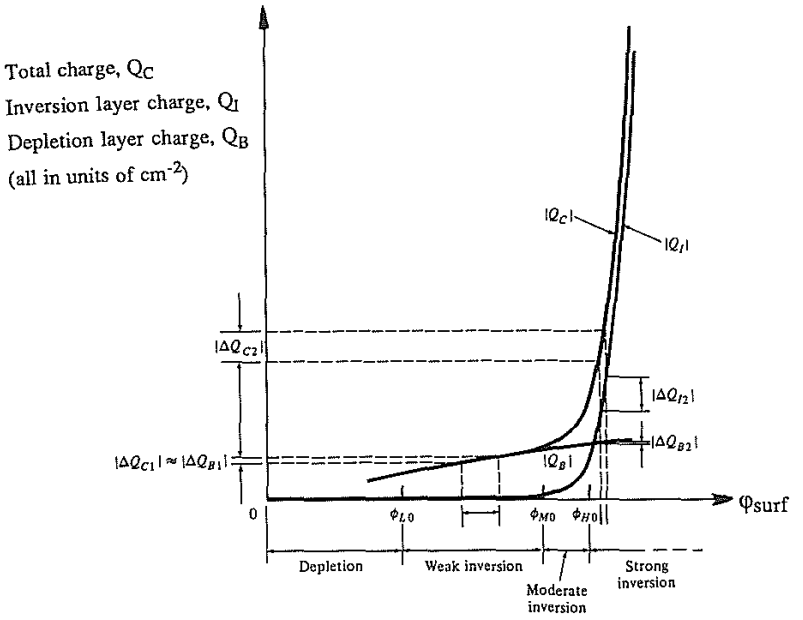


Fig. 3-8 Magnitude of inversion layer charge, depletion layer charge, and their sum (all per unit area) versus φ_{surf}. [6] Reprinted with permission of McGraw-Hill Book Company.

have the flat-band condition, in which $Q_C = 0$. When φ_{surf} is positive, and its value lies between 0 and $2\varphi_B$, Q_C is negative and is $\sim\sqrt{\varphi_{surf}}$. This means that in depletion and weak inversion the mobile charge density is much smaller than the ionic charge density. For $\varphi_{surf}>2\varphi_B$, strong inversion occurs, and $Q_C \cong -\exp(q\varphi_{surf}/2kT$, indicating that the mobile charge density is now much greater than the ionic charge density.

To calculate φ_{surf}, assuming we are given V_{GB}, C_{ox} and N_A, we use

$$V_{GB} = \varphi_{ox} + \varphi_{surf} \tag{3 - 43}$$

and

$$\varphi_{ox} = \{\kappa_{si}\varepsilon_o\, \mathcal{E}_{si}\,(x = 0)\}/C_{ox} \tag{3 - 44}$$

$$V_{GB} = \varphi_{surf} + \frac{\sqrt{2q\, N_A \kappa_{si}\varepsilon_o}}{C_{ox}}\sqrt{\varphi_t e^{-\varphi_{surf}/\varphi_t} + \varphi_{surf} - \varphi_t + e^{-2\varphi_B/\varphi_t}\left(\varphi_t e^{-\varphi_{surf}/\varphi_t} - \varphi_{surf} - \varphi_t\right)} \tag{3 - 45}$$

* (cont.) The conclusions we drew concerning the mobile charge concentration in the silicon in that case for these regions are reiterated here. However, in Fig. 3-8 the contribution of Q_B to the change in Q_C as φ_{surf} is varied is also included.

Note that to find φ_{surf} from this equation we must perform a numerical iteration.

To calculate Q_I accurately we must follow the procedure outlined in section 3.3.2. That is, we must use Eq. 3-21. Here we also change the variable of integration to φ

$$Q_I = -q\int_0^{x_c} n(x)\,dx = -q\int_{\varphi_{surf}}^{\varphi_c} \frac{n(\varphi)\,d\varphi}{d\varphi/dx} \qquad (3-46)$$

By using Eqs. 3-12a and 3-40 in Eq. 3-46 (since $\mathcal{E} = -d\varphi/dx$), we get

$$Q_I = q\int_{\varphi_{surf}}^{\varphi_c} \frac{n_{po}\, e^{\varphi/\varphi_t}}{\sqrt{\frac{2q\,N_A}{\kappa_{Si}\varepsilon_o}}\sqrt{\varphi_t e^{-\varphi/\varphi_t} + \varphi - \varphi_t + e^{-2\varphi_B/\varphi_t}\left(\varphi_t e^{-\varphi/\varphi_t} - \varphi - \varphi_t\right)}}\,d\varphi \qquad (3-47)$$

A similar procedure gives Q_B (consisting of ionized acceptor atoms and holes)

$$Q_B = q\,p_{po}\int_{\varphi_{surf}}^{\varphi_c} \frac{e^{-\varphi/\varphi_t} - 1}{\sqrt{\frac{2q\,N_A}{\kappa_{Si}\varepsilon_o}}\sqrt{\varphi_t e^{-\varphi/\varphi_t} + \varphi - \varphi_t + e^{-2\varphi_B/\varphi_t}\left(\varphi_t e^{-\varphi/\varphi_t} - \varphi - \varphi_t\right)}}\,d\varphi \qquad (3-48)$$

To get $\varphi(x)$ we must integrate the Poisson equation a second time. That is, to find φ at some point x we start with Eq. 3-40, separate variables and integrate from x to 0 on the right hand side, and from φ to φ_{surf} on the left:

$$-\int_{\varphi}^{\varphi_{surf}} \frac{d\varphi}{\sqrt{\frac{2q\,N_A}{\kappa_{Si}\varepsilon_o}}\sqrt{\varphi_t e^{-\varphi/\varphi_t} + \varphi - \varphi_t + e^{-2\varphi_B/\varphi_t}\left(\varphi_t e^{-\varphi/\varphi_t} - \varphi - \varphi_t\right)}} = x \qquad (3-49)$$

However, unlike the first integration which yielded the analytical expression for $\mathcal{E}(x)$ given by Eq. 3-40, this second integration cannot be carried out analytically. That is, we must use numerical procedures to calculate $\varphi(x)$ in this case. This was first done by Young in 1961.[2] Numerical integration must also be used to find the electron and hole densities using Eqs. 3-12 a and 3-12b. Sample plots of $[\rho/qN_A]$ versus x and $\varphi(x)$ versus x calculated using such numerical analysis techniques are shown for accumulation (Fig. 3-9a), depletion (Fig. 3-9b), onset of strong inversion, $\varphi_{surf} = 2\varphi_B$ (Fig. 3-9c), and strong inversion, $\varphi_{surf} = 2\varphi_B + 0.15V$ (Fig. 3-9d) for a MOS-C with $\varphi_B = 0.3V$.[3]

We observe from these curves that: (1) the mobile charge in the inversion layer is confined to a very thin layer near the surface; and (2) the depletion-region edge is fairly abrupt, with the charge concentration dropping to negligible values over a distance of ~0.5 μm. The first observation supports the use of the following two approximations

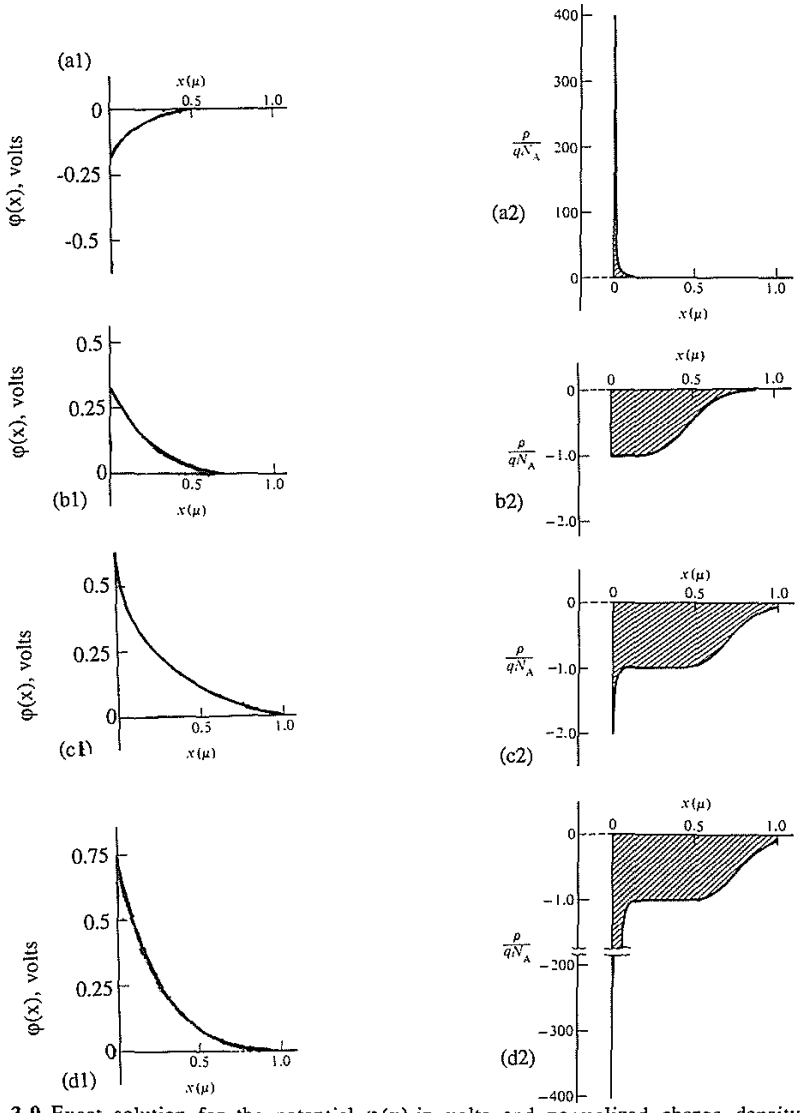


Fig. 3-9 Exact solution for the potential $\varphi(x)$ in volts and normalized charge density (ρ/qN_A), respectively, in the semiconductor region of a MOS-C with a *p*-type Si substrate, as a function of depth, x. Here we assume $\varphi_B = 0.3V$ and T = 300K: (a1 and a2) Accumulation ($\varphi_{surf} = -0.15V$); (b1 and b2) middle of depletion ($\varphi_{surf} = \varphi_B = 0.3V$); (c1 and c2) Onset of inversion ($\varphi_{surf} = 2\varphi_B = 0.6V$); and (d1 and d2) strong inversion ($\varphi_{surf} = 2\varphi_B + 0.15V = 0.75V$). The charge concentration ρ/N_A diagrams are drawn on a linear scale.[3] (Compare Figs. a2 - d2 to the simpler representation of charge in the silicon shown in Figs. 3-4c2 - f2.)

invoked in both the simplified and charge sheet models of the MOS-C: (a) the inversion layer can be treated as an infinitesimally-thin sheet of charge at the interface; and (b) the presence of mobile charge in the depletion region can be ignored (the depletion approximation). The second observation indicates it is reasonable to assume that transition from the bulk to the depletion region occurs abruptly.

In summary, the complicated procedure outlined above involves the following steps to get exact values of MOS-C parameters. For an ideal, uniformly doped *p*-type MOS-C, if we are given N_A, t_{ox}, and V_{GB}

1. Calculate φ_{surf} using Eq. 3-45.

2. With this value of φ_{surf} we can get Q_I from Eq. 3-47, $\mathcal{E}(x)$ from Eq. 3-40, and $\varphi(x)$ from Eq. 3-49.

3. Once $\varphi(x)$ is known we can calculate $n(x)$ and $p(x)$ using Eqs. 3-12a and b.

From the sample plots of the charge distribution we see that in accumulation and inversion the mobile charge resides in an extremely narrow portion of the semiconductor immediately adjacent to the oxide-semiconductor surface. This substantiates the validity of the assumption used in the depletion approximation analysis that the inversion layer charge is confined to a very thin layer near the silicon surface. Thus, a third model based on this approximation (that is intermediate in complexity between the two models just described) has been developed by Baccarani et al.,[4] and Brews.[5] It is called the *charge-sheet model* and will be presented next.

3.3.5 Charge-Sheet Model for Determining the Characteristics of the Ideal MOS-C

The third quantitative model of the MOS-C is the *charge-sheet model.* It is more accurate than the simple depletion approximation model, but does not require the numerical analysis of the general model to obtain Q_I. Since it includes mobile charge carriers in the Poisson equation it provides information about the charge distribution in the MOS-C below strong inversion (unlike the depletion approximation model, which assumes that no mobile charges are present below strong inversion). Since the charge-sheet model assumes that the mobile charge concentrations are zero in the depletion region outside of the thin layer at the surface (that would contain the inversion layer - see assumption 2 below), this model is not as accurate for characterizing the MOS-C in depletion or accumulation as the general model. However, since inversion is responsible for current conduction in MOSFETs, accurate quantitative characteristics of the depletion and accumulation regimes in MOS-Cs and MOSFETs are usually not of much concern. Hence, this does not represent a significant limitation for most applications.

To begin the derivation, we list the assumptions upon which the charge-sheet model is based:

1. Mobile charge can exist in the silicon beyond the onset of *weak* inversion i.e., when $\varphi_{surf}>\varphi_B$ (unlike the previous model that assumed that no mobile charge exists in the channel until strong inversion, i.e., until $\varphi_{surf}>2\varphi_B$);

2. All this mobile charge is confined to a negligibly thin inversion layer at the silicon surface, so that the remainder of the depletion region contains no mobile charges ($n = p = 0$, for $0 < x \leq d$);

3. The depletion region is defined by a sharp boundary at a depth d (i.e., an abrupt transition between the depletion region and the bulk exists), in the same manner as was invoked the depletion approximation model;

4. The voltage drop across the inversion layer in the x-direction is zero, hence V_{GB} (in an ideal MOS-C) is still the sum of φ_{surf} and φ_{ox};

5. The value of φ_{surf} beyond strong inversion is not clamped at exactly $2\varphi_B$. Instead its exact value can be calculated from the model equations;

6. The depletion width at strong inversion is calculated using Eq. 3-28, so that the ionic charge in the depletion region can be calculated using Eq. 3-29;

7. The value of Q_I can be found by subtracting Q_B from Q_C.

Since both holes and electrons are included in the charge density term of the Poisson equation, we start the development of the charge sheet model again with Eq. 3-37. The first integration of the Poisson equation is also carried out in the same way as in the general model analysis, yielding the identical value for $\mathcal{E}_{si}(x = 0)$. Q_C is also expressed by Eq. 3-42. However, since the goal of the charge sheet model is to provide accurate MOS-C expressions only for the case of inversion (including weak inversion), Q_C can be simplified for this operating range. That is, if $\varphi_{surf} \geq \varphi_B$ (where, as we noted in section 3.3.1, φ_B is $9\varphi_t$ to $16\varphi_t$ for the range of doping concentrations used in MOSFETs), several terms in Eq. 3-42 become insignificant and can be dropped, yielding

$$Q_C \approx \sqrt{2q\, N_A \kappa_S \varepsilon_o} \sqrt{\varphi_{surf} + \varphi_t\, e^{(\varphi_{surf} - 2\varphi_B)/\varphi_t}} \qquad (3 - 50)$$

Now, since we assume that in the depletion region there are no mobile charges present (i.e., all the mobile charge is in the inversion layer, which is infinitesimally thin), the depth of the depletion region is again given by Eq. 3-27

$$d = [\kappa_{si}\varepsilon_o(2\varphi_{surf})/qN_A]^{1/2} \qquad (3 - 51)$$

Then, Q_B will consist only of the uncovered acceptor atoms in the depletion region, and can again be expressed by Eq. 3-29, repeated here:

$$Q_B = q\, N_A\, d = (2\kappa_{si}\varepsilon_o q N_A \varphi_{surf})^{1/2} \qquad (3 - 52)$$

To get Q_I we can now subtract Q_B from Q_C as given by Eq. 3-50 to get

$$Q_I = \sqrt{2q\, N_A \kappa_S \varepsilon_o}\left(\sqrt{\varphi_{surf} + \varphi_t\, e^{(\varphi_{surf} - 2\varphi_B)/\varphi_t}} - \sqrt{\varphi_{surf}}\right) \qquad (3 - 53)$$

Also note that if we use Eq. 3-50 to approximate Q_C in Eq. 3-19 for φ_{ox}, and insert this into Eq. 3-10, the expression relating φ_{surf} and V_{GB} is now

$$V_{GB} = \varphi_{surf} - Q_C/C_{ox} \tag{3 - 54}$$

or

$$V_{GB} = \varphi_{surf} + \frac{\sqrt{2q\, N_A K_S \varepsilon_0}}{C_{ox}} \sqrt{\varphi_{surf} + \varphi_t e^{(\varphi_{surf} - 2\varphi_B)/\varphi_t}} \tag{3 - 55}$$

We note, however, that this equation cannot be solved explicitly for φ_{surf}, but requires a numerical iteration approach to solve for φ_{surf}.

Nevertheless, the charge sheet model provides analytical expressions for Q_I and Q_B in a MOS-C that are quite accurate over the entire range of inversion (i.e., $\varphi_{surf}>\varphi_B$). This makes it more useful in many applications than the more general MOS-C model.

It should also be noted that when the equation for Q_I from the charge sheet model is plotted versus the surface potential (i.e., using a log-linear plot, as shown in Fig. 3-10),[6] two inversion regimes can be identified: (1) weak inversion; and (2) strong inversion. In each of these regimes we can simplify the expression for Q_I given by Eq. 3-53.

3.3.5.1 Approximation of Q_I in Weak Inversion. First, we recall from section 3.3.4 that Q_I is much smaller than Q_B in weak inversion. Therefore, we can simplify Q_I by treating the term $\varphi_t \exp(\varphi_{surf} - 2\varphi_B)/\varphi_t$ in Eq. 3-53 as small compared to

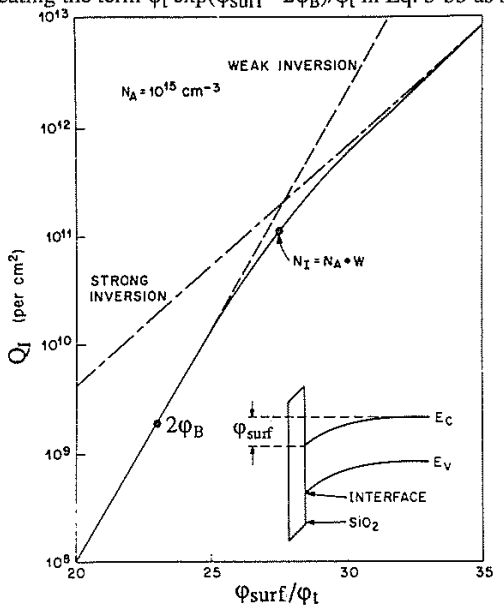


Fig. 3-10 Inversion layer carrier density per unit area versus φ_{surf}/φ_t as computed from Eq. 3-53 and from approximations to this equation for the case of weak inversion (Eq. 3-58), and strong inversion (Eq. 3-62).[5] (© IEEE 1978).

the term φ_{surf}. Then, if we let

$$\zeta = \varphi_t \exp(\varphi_{surf} - 2\varphi_B)/\varphi_t \qquad (3-56)$$

the first square root term in the parentheses in Eq. 3-53 becomes $\sqrt{\varphi_{surf}} + \zeta$. Since ζ is smaller than φ_{surf}, we can approximate the function $\sqrt{\varphi_{surf}} + \zeta$ by the first two terms of its Taylor expansion about $\zeta = 0$:

$$\sqrt{\varphi_{surf} + \xi} \approx \sqrt{\varphi_{surf}} + \frac{1}{2\sqrt{\varphi_{surf}}}\xi \qquad (3-57)$$

Using Eq. 3-56 and Eq. 3-57 in Eq. 3-53 we get

$$Q_I\,(\text{weak inv}) \approx \frac{\sqrt{2q\,N_A\kappa_{Si}\varepsilon_o}}{2\sqrt{\varphi_{surf}}}\varphi_t\, e^{(\varphi_{surf} - 2\varphi_B)/\varphi_t} \qquad (3-58)$$

This expression implies that in weak inversion the inversion layer charge is essentially an exponential function of φ_{surf}. Note from Fig. 3-10 that the weak inversion approximation of Q_I from Eq. 3-53 is quite accurate up to $2\varphi_B$, but becomes increasingly inaccurate beyond strong inversion.

We can easily solve for φ_{surf} in weak inversion as well. That is, since $Q_I \ll Q_B$, we can write Eq. 3-10 as

$$V_{GB} = \varphi_{surf} - Q_B/C_{ox} \qquad (3-59)$$

or using Eq. 3-52

$$V_{GB} = \varphi_{surf} + [(\sqrt{2\kappa_{Si}\varepsilon_o q N_A})/C_{ox}]\sqrt{\varphi_{surf}} \qquad (3-60)$$

$$= \varphi_{surf} + \gamma\sqrt{\varphi_{surf}} \qquad (3-60a)$$

where $\gamma = (\sqrt{2\kappa_{Si}\varepsilon_o q N_A})/C_{ox}$. Equation 3-60a can be solved for φ_{surf} to yield

$$\varphi_{surf} = \left[-\frac{\gamma}{2} + \left(\frac{\gamma^2}{4} + V_{GB}\right)^{1/2}\right]^2 \qquad (3-61)$$

To relate Q_I to V_{GB}, we can substitute the value of φ_{surf} from Eq. 3-61 into Eq. 3-58.

3.3.5.2 Approximation of Q_I in Strong Inversion. In strong inversion $Q_I \gg Q_B$. As a result, the exponential term in Eq. 3-53 is the only significant term within the parentheses. Thus, Q_I in strong inversion can be well approximated as

$$Q_I\,(\text{strong inv}) \approx (\sqrt{2\kappa_{Si}\varepsilon_o q}\; n_i \exp(\varphi_{surf}/2\varphi_t) \qquad (3-62)$$

The inversion-layer charge in strong inversion is thus seen to be an exponential function of the surface potential with a slope of $1/(2\varphi_t)$. Therefore, a small increase in the surface potential induces a large change in Q_I (strong inv). This implies that φ_{surf} will remain almost unchanged even in the face of a large increase in V_{GB} or Q_I beyond strong inversion. The approximation that φ_{surf} (inv) is clamped at $2\varphi_B$ made in the depletion approximation model is thus seen to be reasonably accurate for actual MOSFETs. If the gate voltage corresponding to value needed to produce φ_{surf}(inv) is

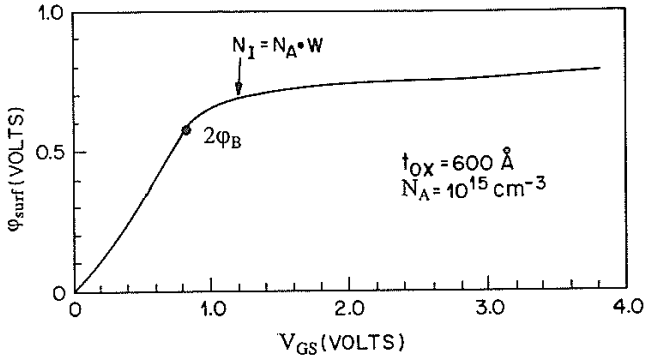


Fig. 3-11 φ_{surf} versus V_{GB} calculated from the charge sheet model for an oxide thickness of 60 nm, and a substrate doping level of $N_A = 10^{15}$ cm^{-3}. The dot indicates $\varphi_{surf} = 2\varphi_B$.[5] (© IEEE 1978).

termed the threshold voltage V_T (see next section for how to compute this value), then the Q_I (strong inv) dependence on V_{GB} can be approximated by

$$Q_I \approx - C_{ox} (V_{GB} - V_T) \qquad (3 - 63)$$

We should mention that φ_{surf}(inv) does change slightly as V_{GB} is increased beyond strong inversion, as shown in Fig. 3-11, where φ_{surf} from Eq. 3-55 is plotted versus V_{GB}. We see that φ_{surf} in strong inversion actually increases beyond $2\varphi_B$ by ~150 mV before leveling off. Thus, models of Q_I that use $\varphi_{surf}(inv) = 2\varphi_B$ underestimate the actual value. This has an impact on the accuracy of predicted drain current values in MOSFET circuit models based on this approximation. We will discuss this further in chap. 4.

3.4 THRESHOLD VOLTAGE of IDEAL MOS CAPACITORS

Strong inversion in MOSFETs occurs in essentially the same manner as in MOS-Cs. In fact, one of the most important parameters of a MOSFET is the gate voltage at the onset of strong inversion. This parameter is known as the *threshold voltage* V_T. We can obtain an expression for V_T using an ideal MOS-C in which we assume the depletion approximation is valid. Then we follow the same analysis as in section 3.3.3.

* Note that a third inversion regime, intermediate to the two mentioned above, namely *moderate inversion*, has been defined by Tsividis (see Fig. 3-6).[6] This allows an even more careful representation of MOS behavior to be developed. Interested readers can consult ref. 6 for more details.

That is, we begin with Eq. 3-10:

$$V_{GB} = \varphi_{ox} + \varphi_{surf} \qquad (3 - 64)$$

At the *onset* of strong inversion (i.e., implying that the inversion layer is just at the verge of forming, so Q_I is still zero), the magnitude of the charge on the gate is balanced by just the charge in depletion region of the semiconductor Q_B due to the ionized dopants, or:

$$Q_G = - Q_B = q\, N_{SUB}\, d_{max} \qquad (3 - 65)$$

where N_{SUB} is the substrate doping (i.e., N_A in *p*-type and N_D in *n*-type substrates), respectively. Then, φ_{surf} at the onset of inversion = $2\varphi_B$. It can also be expressed in terms of d_{max} using Eq. 3-27. Setting these two equal, we have for a *p*-type substrate

$$\varphi_{surf}(inv) = 2\varphi_B = qN_A\,(d_{max})^2/2\kappa_{si}\varepsilon_o) \qquad (3 - 66)$$

Solving this for d_{max}, and substituting into Eq. 3-65 we get Q_B in terms of N_A and φ_B. Finally, using this expression for Q_B in Eqs. 3-19 and 3-20 ($\varphi_{ox} = -Q_B/C_{ox}$), and putting this into Eq. 3-64 we obtain the gate voltage value at the onset of strong inversion in an ideal MOS capacitor (V_T) expressed as a single equation in terms of t_{ox}, T, n_i, and N_A:

$$V_T \text{ (of an ideal MOS capacitor)} = 2\sqrt{\kappa_{si}\varepsilon_o q N_A \varphi_B}/C_{ox} + 2\varphi_B \qquad (3 - 67)$$

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

Although the ideal MOS capacitor is a useful structure for introducing MOS behavior, from the perspective of this text the most important MOS capacitor structure is the metal (Al, or heavily-doped polysilicon)-SiO_2-Si structure. In the metal-SiO_2-Si structure (and other "real" MOS capacitors), the work function difference φ_{ms} is virtually never zero, and various charges are also always present inside the oxide or at the Si-SiO_2 interface. These phenomena exhibited by "real" MOS capacitors significantly alter ideal MOS behavior. They are also inherent in "real" MOSFETs, and must therefore be included to accurately describe actual device behavior.

3.5.1 Workfunction Difference Between Gate and Semiconductor

We noted earlier that the energy-band diagram of an *ideal* MOS capacitor is in a flat band condition when $V_{GB} = 0$. In real MOS capacitors, however, a flat band condition with $V_{GB} = 0$ is essentially never observed. Instead, when $V_{GB} = 0$ the bands are almost always bent as a result of the work function difference between the gate and semiconductor regions, as well as the charges that exist in the MOS capacitor.

The *work function* of a material is defined as the minimum energy required to bring an electron from the Fermi level to the vacuum level. In MOS structures, however, it

is more useful to deal with the *modified* work function. This is defined as the minimum energy required to bring an electron from the Fermi level in the gate E_{Fm} or the semiconductor E_{Fs} to the oxide conduction-band edge ($q\varphi_m'$ or $q\varphi_s'$, respectively), as shown in Fig. 3-12. Note also that $q\varphi_s' = q[\chi' + (E_c - E_F)_\infty] = q(\chi' + E_G/2 + \varphi_B)$ for a *p*-type semiconductor, where χ' is the *modified electron affinity* (defined as the energy difference between the conduction band edges of the semiconductor and the oxide). In the Si-SiO_2 case, $q\chi' = 3.23$ eV.

Figure 3-12a shows the energy-band diagram for an Al-SiO_2-*p*-type-Si structure before contact. For Al-SiO_2, $q\varphi_m' = 3.2$ eV, while for *p*-type Si-SiO_2 $q\varphi_s' = (3.23 + 0.55 + q\varphi_B) \sim 4.1$ eV. Obviously, $q\varphi_m'$ and $q\varphi_s'$ are not equal. Hence, if a MOS structure could be formed by bringing the 2 regions in contact, the effect of this difference on the energy-band diagram of the MOS capacitor would be as shown in Fig. 3-12b. That is, if such a MOS capacitor was unbiased ($V_{GB} = 0$), the Fermi levels would have to "line up" (be at the same energy level) everywhere in the capacitor, Fig.

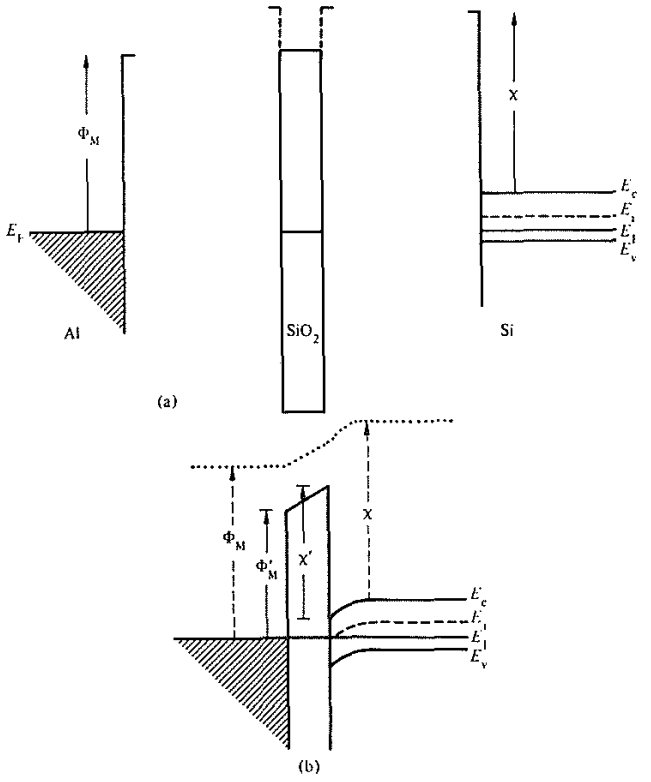


Fig. 3-12 The MOS capacitor structure and idealized energy-band diagram (a) before contact and (b) after contact.[12] Reprinted with permission of Addison-Wesley.

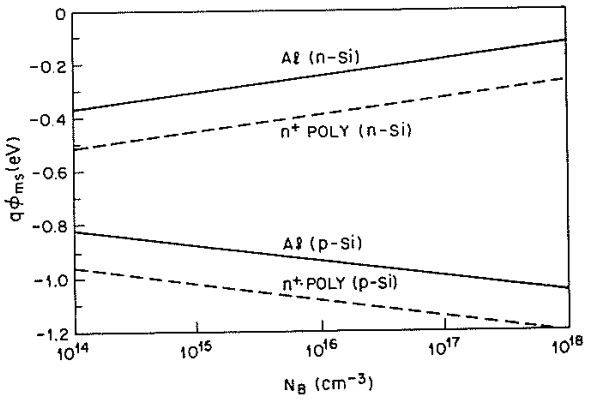


Fig. 3-13 Work-function difference ϕ_{ms} (V) versus doping for n^+ and p^+ polysilicon, and Al gate electrodes.[7] Reprinted with permission of Solid-State Electronics.

3-12). Since the modified work functions in this structure are different, the E_c, E_i, and E_v bands would be bent downward when $V_{GB} = 0$ to satisfy the requirement of a constant Fermi level. To achieve a flat-band condition in this structure, a negative voltage [equal to $(\varphi_m - \varphi_s)$], would have to be applied to the gate (in this example $\cong$ - 0.9 V).

In submicron ICs, however, the gate material is normally heavily-doped polysilicon. The Fermi level in heavily doped n-type polysilicon resides at the conduction band edge (and in heavily doped p-type poly, at the valence band edge). Hence, for an n-channel device with n^+ poly on a p-type substrate $\varphi_{ms}(n\text{-ch}) = (\varphi_m - \varphi_s) = (-E_G/2q) - \varphi_B$ (with its value between -0.9 to -0.95V for $10^{16}\,cm^{-3}<N_A<10^{17}\,cm^{-3}$ - the appropriate doping range for submicron technologies if $L>0.5\ \mu m$). Likewise, the p-channel φ_{ms} for an n^+ poly on an n-type substrate is $\varphi_{ms}(p\text{-ch}) = (\varphi_m - \varphi_s) = (-E_G/2q) + |\varphi_B|$, with its value between -0.15 to -0.2V for the same doping range. Finally, for a p-channel device having a p^+ poly on an n-substrate, $\varphi_{ms}(p\text{-ch}) = (\varphi_m - \varphi_s) = (E_G/2q) + |\varphi_B|$. Figure 3-13 shows the value of φ_{ms} for various substrate-doping values if an aluminum, n^+-poly, or p^+-poly gate electrode is used.

3.5.2 Oxide Charges and Traps

Four different types of charges are associated with the oxide and Si-SiO_2 interface as shown schematically in Fig. 3-14. The effective net (i.e., total) charge per unit area at the Si-SiO_2 interface (in C/cm^2) is denoted by the symbol Q_{tot}, whereas the net (total) number of charges per unit area (Q_{tot}/q) is given by N_{tot}. The standardized terminology and symbols used to differentiate the various individual charge types are as follows:

1. Mobile ionic charge per unit area Q_m (C/cm^2), and N_m (no. of mobile charges/cm^2);

2. Fixed oxide charge per unit area Q_f (C/cm^2), and N_f (no. of fixed oxide charges/cm^2);

3. Interface trap charge per unit area Q_{it} (C/cm^2), and N_{it} (no. of interface trapped charges/cm^2); D_{it} - number of interface trapped charges per unit area *and* energy (no. of interface trapped charges/cm^2-eV);

4. Oxide trapped charge Q_{ot} (C/cm^2), and N_{ot} (no. of oxide trapped charges/cm^2).

Since issues relating to the reliability of submicron MOSFETs also involve these charges, a brief review concerning the origin of the charges, their effect on device behavior, and techniques to minimize their presence is given here. A more detailed description of their characteristics can be found in Vol. 1 and refs. 1 and 7.

3.5.2.1 Mobile Ionic Charges. Mobile ionic charges Q_m in the oxide layer arise mostly from the presence of sodium or potassium ions, which have very high diffusivities in the oxide (even at temperatures of less than 200°C). Their presence leads to threshold instabilities and to a deterioration of the oxide reliability. After it was confirmed that sodium in even very small concentrations is ruinous to MOSFET operation, stringent procedures were developed to keep the level below $N_m = 10^9$ cm^{-2}. At such levels, the instabilities in V_T are negligibly small. Procedures to counteract Na contamination include: using clean processing techniques; employing gettering techniques; neutralization of the sodium in the oxide by growing the oxide in ambients containing chlorine compounds; and using SiN (or oxynitride) passivation films. (Such films are impervious to the transport of Na.) Tests for Na contamination are also routinely performed (using MOS-C monitor structures) to detect any accidental contamination. Steps can then be taken to find and eliminate the contamination source. More information on the details of these techniques can be found in Vol. 1, Chap. 7.

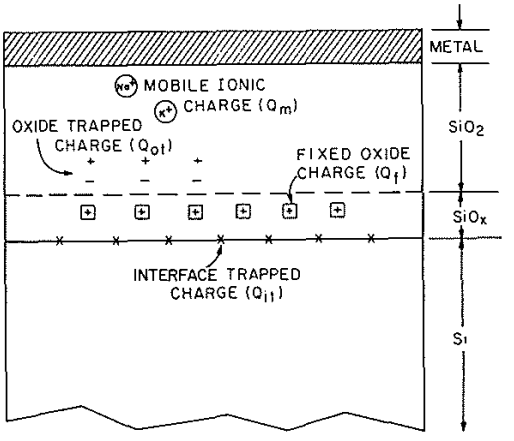


Fig. 3-14 Standardized terminology for oxide charges associated with thermally oxidized silicon.[13] (© IEEE 1980).

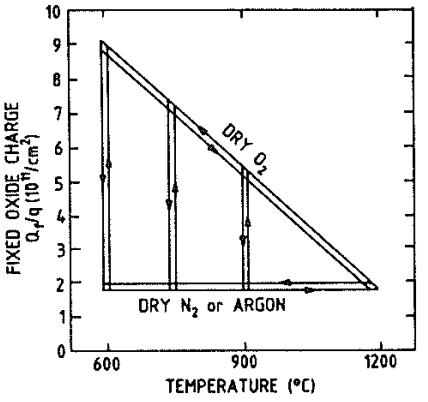


Fig. 3-15 The *Deal triangle,* which specifies the expected Q_f/q after dry O_2 oxidation and after inert ambient annealing.[9] Reprinted by permission of the publisher, The Electrochemical Society, Inc.

3.5.2.2 Fixed Oxide Charge. The fixed oxide charge Q_f (usually positive) is located in the oxide within approximately 30Å of the Si-SiO_2 interface. Since its density does not vary with surface potential (unlike interface trapped charge Q_{it}), it is termed *fixed* oxide charge. It is generally accepted that Q_f is associated with the structure of the interfacial transition region between the silicon and SiO_2. One hypothesis is that Q_f is due to *excess ionic silicon* which has broken away from the silicon proper and is waiting to react in the vicinity of the Si-SiO_2 interface when the oxidation process is abruptly terminated.

Regardless of the origin, it has been experimentally confirmed that the magnitude of Q_f is influenced by substrate orientation, oxidation temperature, and anneal conditions after oxide growth. These factors were carefully studied by Deal et al., and their results are summarized by the well-known "Deal-triangle" shown in Fig. 3-15. In this figure N_f (Q_f/q) is plotted for (111) Si against final oxidation or anneal temperature. It can be seen that the minimum N_f value for (111) material is ~2×10^{11} cm^{-2}. The time required to reach a minimum value of N_f is less than 5 min at 1200°C and about an hour at 550°C. In addition, after an oxide growth step, wafers are annealed in an inert ambient (part of the annealing step occurring during the slow withdrawal from the furnace). It should be noted, however, that continued annealing in inert ambients (e.g., N_2) at high temperatures leads to a marked increase of N_f. Fig. 3-16 shows how N_f changes as a function of time during a nitrogen anneal at 950, 1100 or 1200°C.

The value of N_f also depends on the orientation of the Si surface, and is lowest on (100) orientations (in which the minimum N_f value is ~1.1×10^{10} cm^{-2}). As the lowest values of N_f are desired, (100) material is used to build MOS ICs.

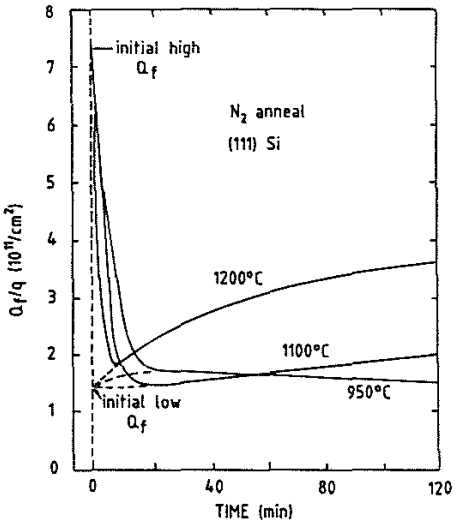


Fig. 3-16 Variation of fixed oxide charge as a function of time during nitrogen anneal at 950, 1100 or 1200°C.[10] Reprinted by permission of the publisher, The Electrochemical Society, Inc.

Since Q_f is located at the Si-SiO_2 interface, the change in V_T that it causes (ΔV_T) is simply calculated as the shift in V_T that would be produced by a sheet of positive charge placed at the interface, or

$$\Delta V_T = -Q_f/C_{ox} = q\, N_f / C_{ox} \qquad (3\text{-}68)$$

3.5.2.3 Interface Trapped Charge. Interface trapped charge Q_{it} arises from allowed energy states (also referred to as *surface states*, *interface states*, or *interface traps*) that somehow exist in the forbidden gap of the Si in the region very close to the Si-SiO_2 interface. Although models exist that detail the *electrical* behavior of interface traps, their *physical origin* has not been totally clarified. The weight of experimental evidence supports the view that the interface traps arise primarily from unsatisfied chemical bonds or so-called "dangling bonds" at the surface of the semiconductor. When a silicon lattice is abruptly terminated along a given plane to form a surface, one of the four surface-atom bonds will be left dangling as pictured in Fig. 3-17. It seems plausible that most, but maybe not all, of such Si-surface bonds would be tied up as a result of thermally growing an SiO_2 layer (Fig. 3-17b). The remaining dangling bonds might become interfacial traps. Other possible interfacial trap-formation effects include Si-Si bond stretching - due to the absence of an oxygen atom (Fig. 3-17c), Si-O bond stretching (Fig. 3-17d), and the presence of a metallic impurity at the Si surface (Fig. 3-17e).[11] These interface traps are also considered further in chaps. 7 and 8.

Based on the wide-ranging and degrading effects they cause in the operational behavior of MOSFETs (and the fact that Q_m and Q_f can generally be acceptably

minimized), interface trapped charges must be considered to be the most perfidious of the oxide charges. Because they are not usually caused by the presence of foreign impurities, but instead are due to a disruption of the periodicity of the crystal lattice, they are not associated with well-defined energy levels. Instead, interface trapped charges are distributed throughout the entire band gap at closely spaced levels. (Interface levels

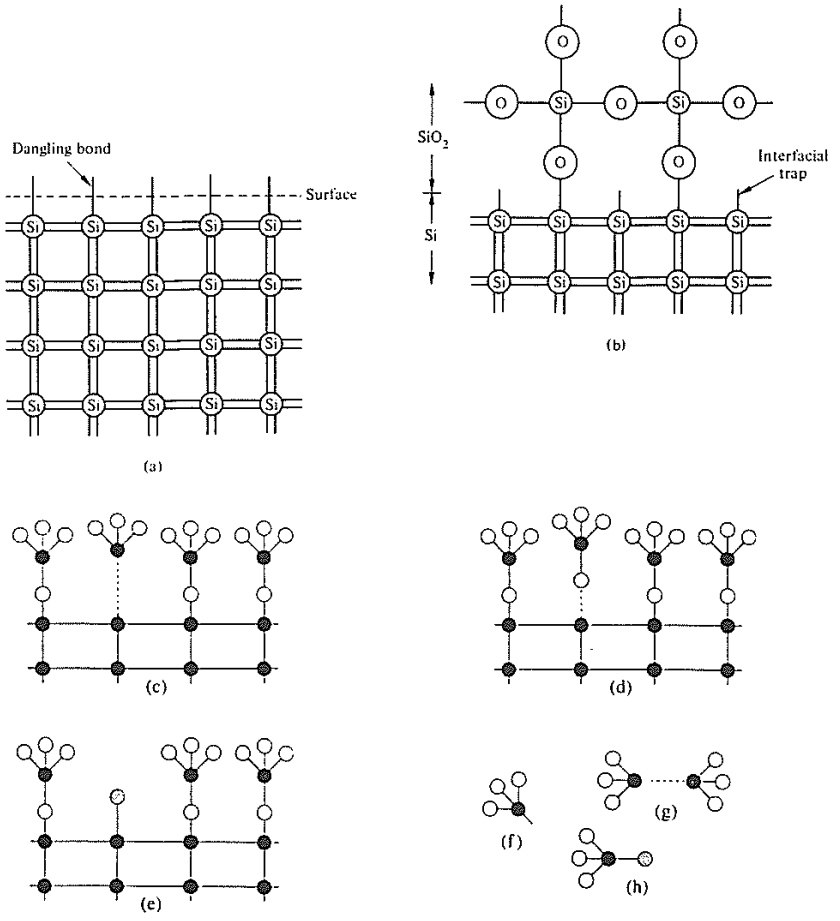


Fig. 3-17 Physical model for the interfacial traps. (a) "Dangling bonds" which occur when the Si lattice is abruptly terminated along a given plane to form a surface. (b) Post oxidation dangling bonds (relative number greatly exaggerated) that become the interfacial traps. (c) Si-Si bond stretching. (d) Si-O bond stretching. (e) presence of a metallic impurity at the Si surface. (f) - (h) possible defects in the SiO_2 bulk that may give rise to oxide trapped charge.

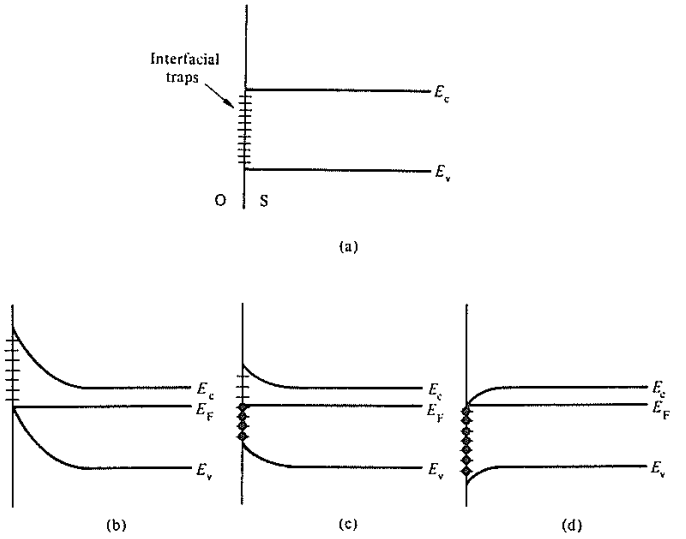


Fig. 3-18 (a) Electrical modeling of interfacial traps as allowed electronic levels localized in space at the oxide-semiconductor interface. (b) - (d) Filling of the interface levels under (b) inversion, (c) depletion, and (d) accumulation biasing in an *n*-type device. From R.F. Pierret, *Modular Series on Solid-State Devices*; *Vol. IV - Field Effect Devices.* Copyright, 1983, Addison-Wesley. Reprinted with permission.

can also occur at energies greater than E_C or less than E_V, but such levels are usually obscured by the much larger density of conduction- or valence-band states.) By using an energy-band diagram to show how these energy levels are distributed throughout the forbidden gap (Fig. 3-18a), we can more easily visualize their impact on a MOS-C.

That is, if an *n*-doped Si MOSFET (PMOS) is biased into inversion (Fig. 3-18b), the Fermi level at the surface will be close to E_V. In this case, the interfacial traps will mostly be empty, as they lie in energies above E_F. Furthermore, if these empty trap sites are assumed to be donorlike (that is, positively charged when empty and neutral when filled with an electron), the net charge per unit area associated with the interfacial traps (Q_{it}) will be positive. Changing the gate bias to achieve depletion conditions (Fig. 3-18c) positions E_F near the middle of the band gap at the surface. Since the interface levels always remain fixed in energy relative to E_C and E_V at the surface, depletion biasing obviously draws electrons into the lower interface state levels, and Q_{it} reflects the added negative charge. If the MOS capacitor is biased into the accumulation mode, most of the traps will be filled with electrons (Fig. 3-18d), and Q_{it} will approach its minimum value. The purpose of the above discussion is to show that the interface traps charge and discharge as a function of bias, thereby affecting the charge distribution inside the device, as well as the device characteristics in a complex but somewhat predictable manner.

Once again, however, since Q_{it} is located at the Si-SiO_2 interface, the change in V_T due to Q_{it} can be simply calculated from

$$\Delta V_T = -Q_{it}(\varphi_{surf})/C_{ox} \tag{3-69}$$

Since Eq. 3-69 shows that Q_{it} depends upon the bias condition (and is therefore a function of φ_{surf}), its value will change as V_{GB} changes. Under conditions of inversion Q_{it} takes on its largest positive value, and decreases as the bias progresses from depletion to accumulation. The distribution of trap levels as a function of energy in the band gap has a U-shaped distribution (see Fig. 3-19). The minimum level of D_{it} at midgap is usually the density used to characterize their presence.

The overall value of Q_{it}, as well as the precise density of states as a function of energy, makes the value of D_{it} sensitive to a variety of process conditions used to fabricate the MOSFET. Nevertheless, certain definite trends concerning these values have been uncovered. First, Q_{it}, like Q_f, is greatest on (111) surfaces, and smallest on (100) surfaces, and the ratio of midgap-D_{it} values on the two surfaces is ~3:1. Second,

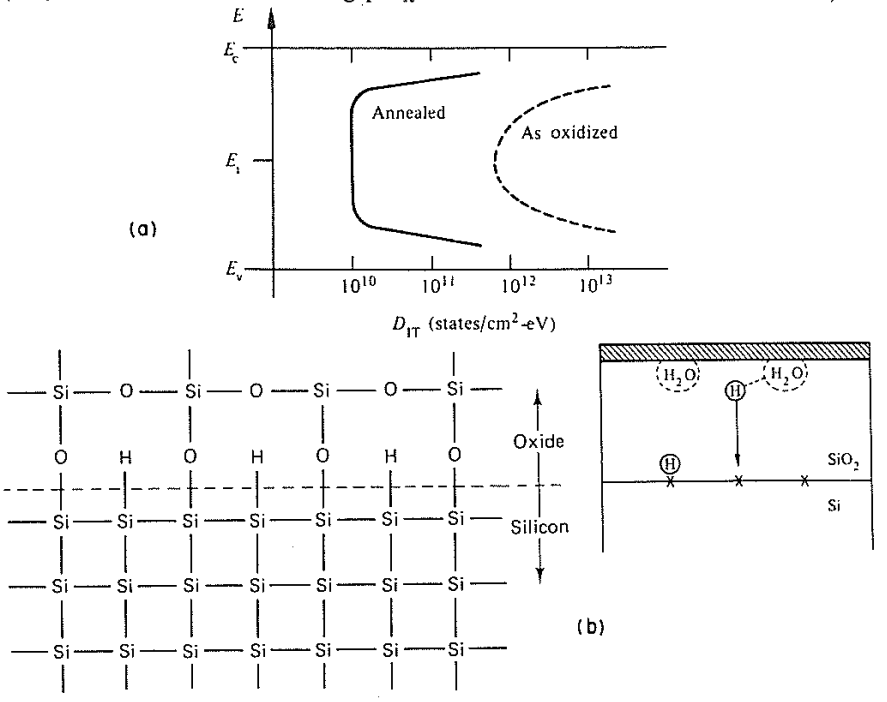


Fig. 3-19 (a) Energy distribution of interface states within the band gap. General form and magnitude of the interfacial trap density observed before and after interface state anneal. (b) Conceptual picture of how hydrogen is thought to passivate dangling bonds.

after oxidation in a dry O_2 ambient, the midgap value of D_{it} is relatively high ($\sim 10^{11}$-10^{12} states/cm^2-eV), with the values decreasing as the oxide temperature is increased (this is also the case with Q_f). However, annealing at lower temperatures in inert ambients *does not* reduce the value of D_{it}. Instead, annealing in the presence of hydrogen at relatively low temperatures (≤500°C) for 5-10 minutes minimizes D_{it}. D_{it} at midgap under an optimum set of processing conditions is $\sim 10^{10}$/cm^2-eV. See Fig. 3-19b for a conceptual picture of how hydrogen is thought to passivate the dangling bonds at the silicon surface and how this thus reduces the interface trap density.

3.5.2.4 Oxide Trapped Charge. The oxide trapped charge Q_{ot} is due to holes or electrons trapped in the *bulk* of the oxide, and hence can be positive or negative. Q_{ot} is associated with defects in the SiO_2, which may arise from ionizing radiation, hot carrier injection, or high currents through the oxide (e.g., due to Fowler-Nordheim tunneling - as described in chap. 7), or other reactions which either tend to break Si-O bonds in the oxide or otherwise lead to carrier trapping on sites or traps somehow already present in the oxide. Figure 3-17 f - h shows possible defects in the SiO_2 bulk that may give rise to such traps. For example, a trivalent Si atom in the SiO_2 will introduce a deep trap level in the SiO_2 bandgap ($\bullet Si{\equiv}O_3$), and stretched Si-O or Si-Si bonds (in the latter a missing bridging oxygen atom leads to a defect, $(O_3{\equiv}Si\bullet\bullet S{\equiv}O_3)$ introduce a continuum of shallow or deep trap levels in the SiO_2 bandgap when located in the bulk of the SiO_2.

Q_{ot} can be annealed out by low-temperature treatments (450-550°C) although some neutral traps may remain. Therefore, oxide trapped charge has been considered relatively less important than the other oxide charges in just-completed MOSFETs. In early generations of MOSFET ICs, exposure of the devices to ionizing radiation encountered in space flights was the main concern involving Q_{ot}. However, more recently-introduced processing techniques such as ion-implantation, electron-beam evaporation of metal films, reactive-ion etching, and x-ray lithography are all potential sources for creating charged and neutral traps in the gate dielectric (all of which may not be completely annealed out at the end of the fabrication process). This may result in long-term reliability problems, especially if hot carriers are injected in these oxides and get trapped (see chap. 9). Another reliability concern involves gate oxide degradation due to high oxide electric fields (and subsequent Fowler-Nordheim tunneling currents in the oxide). This has become more of a concern as oxide thickness has been scaled down.

3.5.2.5 Combined Effect of Oxide Charges on V_T of Submicron MOSFETs. It is now possible to control the level of Q_m in modern IC fabrication processes so that it has an insignificant effect on V_T stability. Thus, Q_f and Q_{it} are the remaining important charges that must be controlled during fabrication. For this discussion, let the sum of Q_f and $Q_{it} = Q_{tot}$. Then, combining Eqs. 3-68 and 3-69,

$$\Delta V_T = -Q_{tot}/C_{ox} = -(Q_f + Q_{it})/C_{ox}. \qquad (3-70)$$

If $N_{tot} = Q_{tot}/q$ is maintained at $<5\times10^{10}$/cm^2, for oxides <200Å thick:

$$\Delta V_T = Q_{tot}/C_{ox} = qN_{tot}\, t_{ox}/3.9\, \varepsilon_o$$

$$= (1.6\text{x}10^{-19}\ \text{C})(5\text{x}10^{10}\ \text{cm}^{-2})(2\text{x}10^{-6}\ \text{cm})\ /(3.5\text{x}10^{-13}\ \text{F/cm})$$

$$= 0.05\ \text{V}$$

Thus, for routine submicron process technology, it may seem that these charges are no longer a major problem. However, this is not true as far as the problems of gate oxide reliability and MOSFET degradation due to hot-carrier injection effects are concerned. That is, due the high electric fields present in submicron MOSFETs, phenomena occur that give rise to the generation of additional Q_{ot} and Q_{it} during normal device operation. Thus, the value of these charges can rise during the lifetime of the device, and cause performance degradation or failure. This topic will be revisited in more detail in chaps. 7 and 9.

3.6 THE THRESHOLD VOLTAGE EQUATION OF Si-SiO$_2$ MOS CAPACITORS

To obtain a flat-band condition in a real (non-ideal) MOS capacitor, a voltage must be applied to the gate to overcome the effects of $\varphi_{ms} \neq 0$ and $Q_{tot} \neq 0$. The total voltage needed to offset these effects is referred to as the *flat-band voltage* V_{FB} and is given as

$$V_{FB} = \varphi_{ms} - Q_{tot}/C_{ox} \qquad (3\text{-}71)$$

The gate voltage that must then be applied to a real MOS capacitor structure to achieve the onset of strong inversion (i.e., threshold voltage, V_T) must include the flat band voltage as well. Thus, the V_T equation for real MOS capacitors (and long-channel MOSFETs when the source-to-substrate voltage $V_{SB} = 0$V) is given as

$$V_{T\ (\text{uniform doping, long-channel})} = V_{T(ud,lc)} = V_{FB} + 2\sqrt{\kappa_{si}\varepsilon_o q N_{sub}\varphi_B}/C_{ox} + 2\varphi_B$$

$$= \varphi_{ms} - Q_{tot}/C_{ox} + 2\sqrt{\kappa_{si}\varepsilon_o q N_{sub}\varphi_B}/C_{ox} + 2\varphi_B \qquad (3\text{-}72)$$

Figure 3-20 plots Eq. 3-72 for n^+ silicon-gate MOSFETs (both PMOS and NMOS) versus substrate doping concentration for several values of t_{ox} (i.e., 65, 25, and 15 nm thick) as the parameter, and assuming also that $Q_{tot} = 0$.

Note that Eq. 3-72 also assumes that the Si substrate is uniformly doped. In practical MOSFETs, however, V_T is normally adjusted by implanting a thin layer of impurity atoms (boron or arsenic) at the Si surface. Thus, the substrate becomes non-uniformly doped, and the V_T equation must be modified to take this into account. To a first order this is done by treating the implanted impurity atoms as if they are an additional layer of charge very close to the Si-SiO$_2$ interface. Their presence is accounted for by adding a term to the V_T equation similar to the term used to describe

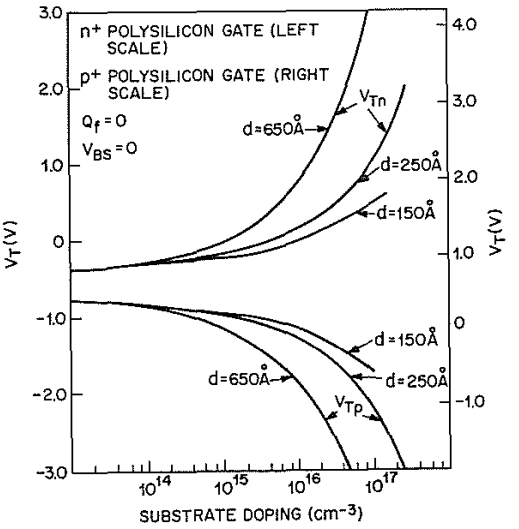


Fig. 3-20 Calculated threshold voltages of n-channel (V_{Tn}) and p-channel (V_{Tp}) transistors as a function of their substrate's doping, assuming n^+ polysilicon gate (left scale) and p^+ polysilicon gate (right scale). Curves for gate oxide thicknesses d of 150 Å, 250 Å, and 650Å are shown. From S.M. Sze Ed. *VLSI Technology,* 2nd Ed., Chap. 11, "VLSI Process Integration." Copyright, 1988 Bell Telephone Laboratories. Reprinted with permission.

the presence of oxide charge Q_{tot}. Details of V_T adjustment and the term used to modify Eq. 3-72 are discussed in chap. 4.

Example 3-1: Find $V_{T(ud,lc)}$ of an ideal MOS-C (φ_{ms} = 0V, Q_{tot} = 0) with a uniform substrate doping $N_A = 10^{15}/cm^2$. The gate-oxide thickness is t_{ox} = 15 nm.

Solution: $\varphi_B(\text{sub}) = (E_i - E_F)/q = 0.29$ V

$$\varepsilon_{ox} = 3.9\varepsilon_o = 3.5\text{x}10^{-13}\ \text{F/cm};\quad C_{ox} = \varepsilon_{ox}/1.5\text{x}10^{-6} = 2.3\text{x}10^{-7}\ \text{F/cm}^2$$

$$2\sqrt{\kappa_{si}\varepsilon_o q N_A \varphi_B} = 2\ [1.02\ \text{x}10^{-12}\ \text{x}\ 1.6\text{x}10^{-19}\ \text{x}\ 10^{15}\ \text{x}\ 0.29]^{1/2} = 1.4\text{x}10^{-8}\ \text{C/cm}^2$$

$$2\sqrt{\kappa_{si}\varepsilon_o q N_A \varphi_B}/C_{ox} = 1.4\ \text{x}\ 10^{-8}/2.3\ \text{x}\ 10^{-7} = 0.06\ \text{V}$$

$$V_{T(ud,lc,\ ideal)} = 0.58\text{V} + 0.06\text{V} = 0.64\ \text{V}$$

Example 3-2: Find $V_{T(ud,lc)}$ of a 'real' MOS-C just like the one of Example 3-1, except that it has a heavily-doped silicon gate (doped to $N_D = 10^{20}/cm^3$), and $Q_{tot} = q\ (5\ \text{x}\ 10^{10}/cm^2)$.

Solution: According to Eq. 3-72, the threshold voltage of the 'real' MOS-C is shifted from that of the 'ideal' MOS-C by the value of V_{FB}. This term is the sum of φ_{ms} + Q_{tot}/C_{ox}. Thus, since

$$\varphi_{ms} \text{ (from Fig. 3-13)} = -1.02 \text{ V}$$

and

$$Q_{tot}/C_{ox} = -1.6\text{x}10^{-19} \text{ x } 2\text{x}10^{10} / 2.3\text{x}10^{-7} = -0.01 \text{ V}$$

$$V_{T(ud,lc)} = V_{T(ud,lc,\ ideal)} + V_{FB}$$

$$= 0.64\text{V} + (-1.03\text{V}) = -0.39 \text{ V}$$

Note that a negative value of $V_{T(ud,lc)}$ is yielded, implying that this MOSFET would be ON at $V_{GB} = 0$V (and would therefore behave as a depletion-mode device). The parameters used in this calculation of $V_{T(ud,lc)}$ are typical of the NMOS device parameters encountered in the early days of MOS ICs. Hence, at that time it was not possible to easily and/or reliably fabricate enhancement-mode NMOS transistors. In chap. 4 we shall show how this problem was overcome by the development of ion implantation. With this process it became possible to introduce dopants into the channel region near the silicon surface, such that the threshold voltage is shifted from the value it has in the case of a uniformly doped substrate, to the desired value. This enabled enhancement-mode NMOSFETs to be manufactured with appropriate threshold voltage values.

The typical value of V_T in NMOSFETs is 0.7V. This means that V_T of the MOS-C structure of Example 3-2 must be shifted by 1.11V. Equation 4-73 indicates that a shallow implant with a dose of ~$1.6\text{x}10^{12}$ boron ions/cm^2 will produce such a V_T shift in the MOSFET of Example 3-2.

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

Many aspects of the behavior of the MOS-C and the MOSFET are described in terms of the variation of the electrostatic potential in the silicon as a function of position. By plotting such relationships in a graphical format, a visual model is obtained which helps to portray many of the phenomena associated with MOS-C and MOSFET behavior.

When describing the MOS-C most graphical representations are one-dimensional, reflecting the one-dimensional nature of the behavior of such structures [i.e., usually $\varphi(x)$ versus x plots are presented]. However, when the potential in the MOSFET is being depicted, a two-dimensional representation is needed to appropriately depict the MOSFET behavior. Thus, we have decided to illustrate the potential versus position in the MOS-C using both the 2-D plot (as is needed for the MOSFET), as well as the 1-D potential plots more usually encountered when MOS-C behavior is being studied. This is done in Fig. 3-21, in which both 1-D and 2-D potential plots for a MOS-C cross-section are depicted. We note that $\varphi_{surf} = \varphi\ (x = 0)$ is constant versus y in the MOS-C, and thus the 1-D representation of the potential in the MOS-C would be

adequate. However, when the 2-D potential versus position diagrams of the MOSFET are presented (i.e., in chaps. 4 and 5), it will be easier to interpret their details by comparing them to the simpler 2-D cases introduced here.

In Fig. 3-21a we show the 2-D cross-section of a *p*-substrate MOS-C, with a bias V_{GB} connected between the gate and body. In Fig. 3-21b we show the case of the *ideal* MOS-C described in Example 3-1. We observe that when $V_{GB} = 0$ (Fig. 3-21b1), $\varphi(x,y) = 0$ everywhere, corresponding to the flat-band condition in the energy-band representation. In Fig. 3-21b2 we portray the case of $V_{GB} = V_T = 0.64V$ for the MOS-C of Example 3-1. Here we note that $\varphi_{surf} = 2\varphi_B = 0.58V$, with the variation of $\varphi(x)$ expressed by Eq. 3-23.

In Fig. 3-21c we note that the *real* MOSFET of Example 3-2 is depicted for the bias condition of $V_{GB} = 0V$. Since we noted that $V_T = -0.39V$, implying that a negative voltage of -0.39V must be applied to cause a condition of $\varphi_{surf} = 2\varphi_B = 0.58V$, which corresponds to the onset of strong inversion. Thus, when $V_{GB} = 0V$, $\varphi_{surf} = 0.97V$, which is larger than $2\varphi_B$, as shown in Fig. 3-21c1.

In order for this MOS-C to exhibit a value of $V_T = 0.7V$, we indicated in the previous section that a dose of 1.6×10^{12} boron ions/cm^2 must be implanted into the channel region. If this is done, when $V_{GB} = 0V$, φ_{surf} will then equal -0.05V, and the potential versus position diagram will as shown in Fig. 3-21d1. Note that when $V_{GB} = 0.7V$, $\varphi_{surf} = 0.58V$, but the potential versus x variation is different from that of the uniformly doped case of Fig. 3-21b. The effect of the implant dose on the depletion region width of the channel and other MOSFET phenomena are discussed in chap. 4.

REFERENCES

1. E.H. Nicollian and J.R. Brews, *MOS Physics and Technology,* Wiley-Interscience, New York, 1982.
2. C.E. Young, "Extended Curves of the Space Charge, Electric Field, and Electrostatic Potential Inside a Semiconductor," *J. Appl. Phys.,* **32**, p. 329, (1961).
3. R.F. Pierret and J.A. Shields, "Simplified long-channel MOSFET theory," *Solid-State Electronics,* vol. 26, p. 143, 1983.
4. G. Baccarani et al., "Analytical i.g.f.e.t.model including drift and diffusion currents," *IEE Journal on Solid-State and Electron Devices,* vol. 2, p. 62, March 1978.
5. J.R. Brews, "A Charge-Sheet Model of the MOSFET," *Solid-State Electronics,* vol. 21, p. 345, 1978.
6. Y.P. Tsividis, *Operation and Modeling of the MOS Transistor,* McGraw-Hill, New York, 1987.
7. W.M. Werner, "The Work Function Difference of the MOS-System with Aluminum Field Plates and Polysilicon Field Plates," *Solid-State Electronics,* **17**, 769 (1974).
8. P. Balk Ed., *The Si-SiO$_2$ System,* Elsevier, New York, 1988.
9. B.E. Deal et al., "Characteristics of the Surface State Charge of Thermally Oxidized Silicon," *J. Electrochem. Soc.,* **114**, p. 266 (1967).
10. B.E. Deal et al., *J. Electrochem. Soc.,* **121**, p. 198C (1974).

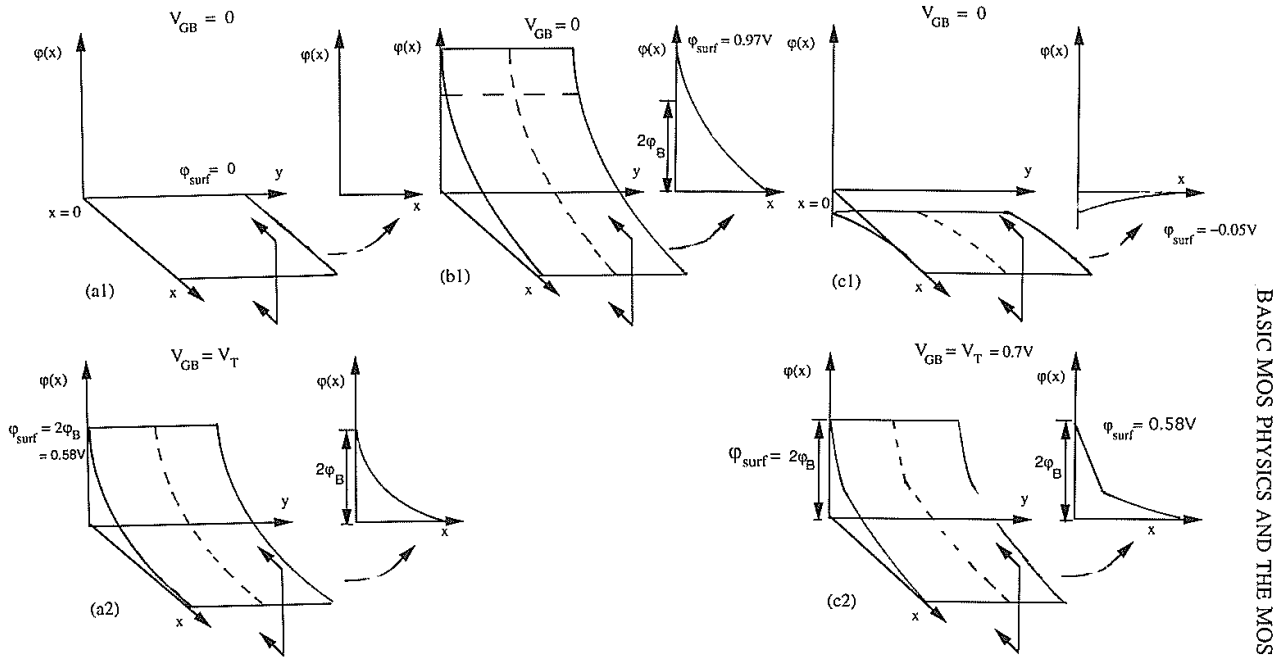


Fig. 3-21 Visualization of the electrostatic potential in *p*-type MOS-Cs with $N_A = 10^{15}$ cm^{-3}, and $t_{ox} = 150$Å. (a1) φ(x,y) in an *ideal* MOS-C ($\varphi_{ms} = 0$ and $Q_{tot} = 0$) when $V_{GB} = 0$V (flatband condition); and (a2) when $V_{GB} = V_T$. (b1) φ(x,y) in *real* MOSFET with uniformly doped substrate and n^+ polysilicon gate ($\varphi_{ms} = 1.02$V), and $Q_{tot} = 2\times10^{10}$ cm^{-2}, when $V_{GB} = 0$V. (c1) φ(x,y) in *real* MOSFET with n^+ polysilicon gate and V_T-adjust implant into substrate such that $V_T = 0.7$V, when $V_{GB} = 0$V. (c2) φ(x,y) when $V_{GB} = V_T$.