

Kirkland & Ellis Library

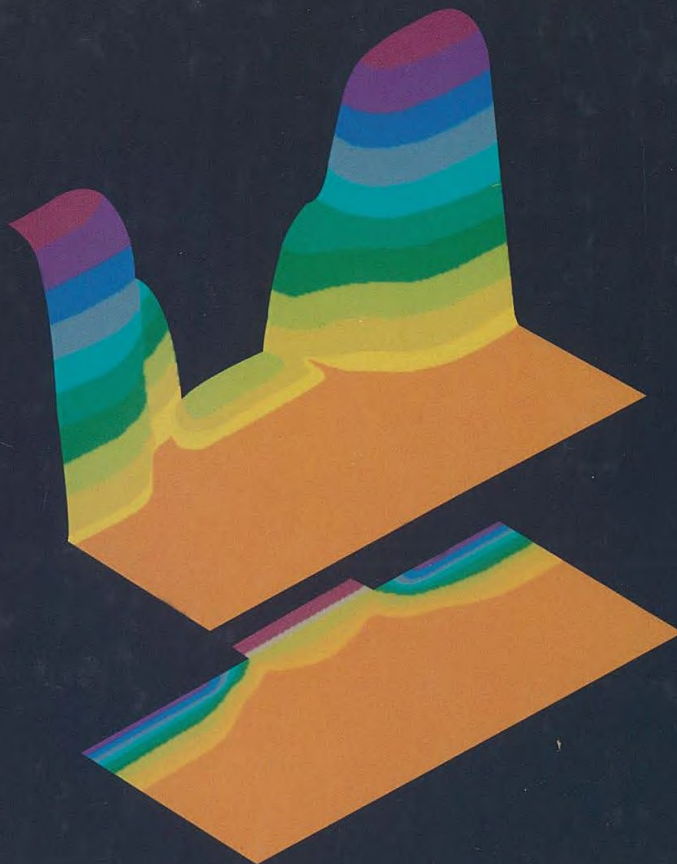


5 0628 01144179 4

# 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

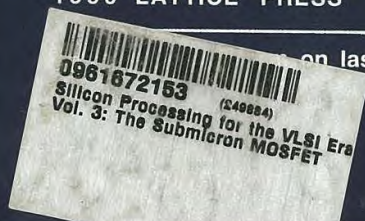
By **STANLEY WOLF**

---

**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

BOOK

ing

hnology

HARD N. TAUBER

Single-Crystal Growth and Wafering;  
essing, and Gettering; **Ch. 3** Vacuum  
Basics of Thin Films; **Ch. 5** Silicon  
apor 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  
**n. 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

egration

cess Integration for VLSI and ULSI;  
rated Circuits; **Ch. 3** Contact Technology  
**4**; Multilevel Interconnect Technology  
ices and Process Integration;  
**n. 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 Library



5 0628 01144

Sili

Vol

# CONTENTS

## PREFACE

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

#### **1.1 MODELS IN MICROELECTRONIC TECHNOLOGY 1**

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

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

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

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

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

|                                                                                                                                                 |           |
|-------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| 1.3.6 Analytical Solutions of the One-Dimensional Form of the Poisson Equation in the Depletion Approximation                                   | 23        |
| <b>1.4 CIRCUMSTANCES UNDER WHICH THE ANALYTIC SOLUTIONS OF THESE PDES BECOME INADEQUATE FOR SIMULATION PURPOSES</b>                             | <b>24</b> |
| 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        |
| <b>REFERENCES</b>                                                                                                                               | <b>28</b> |

|                                                                                                                                              |           |
|----------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| <b>Chap. 2 - NUMERICAL METHODS FOR SOLVING THE PARTIAL DIFFERENTIAL EQUATIONS WHICH MODEL THE PHYSICS OF SUBMICRON DEVICES AND PROCESSES</b> | <b>29</b> |
| <b>2.1 NUMERICAL SOLUTIONS OF DIFFERENTIAL EQUATIONS</b>                                                                                     | <b>31</b> |
| <b>2.2 NUMERICAL METHODS FOR SOLVING NONLINEAR PARTIAL DIFFERENTIAL EQUATIONS</b>                                                            | <b>32</b> |
| <b>2.3 GRID GENERATION AND REFINEMENT PROCEDURES</b>                                                                                         | <b>34</b> |
| <b>2.4 DISCRETIZATION OF DIFFERENTIAL EQUATIONS</b>                                                                                          | <b>36</b> |
| 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.                                                                                                        |           |
| <b>2.5 ORDERING THE SET OF DISCRETIZED PDEs INTO APPROPRIATE MATRIX FORM</b>                                                                 | <b>50</b> |

|                                                                                                                                                                                    |           |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| <b>2.6 LINEARIZING NONLINEAR ALGEBRAIC EQUATIONS</b>                                                                                                                               | <b>52</b> |
| <b>2.7 MATRIX METHODS FOR SOLVING SIMULTANEOUS LINEAR ALGEBRAIC EQUATIONS</b>                                                                                                      | <b>55</b> |
| <b>2.8 VISUALIZATION OF THE RESULTS</b>                                                                                                                                            | <b>57</b> |
| <b>2.9 DATA BASE OF THE SIMULATION PACKAGES</b>                                                                                                                                    | <b>58</b> |
| <b>2.10 EXAMPLES OF THE PROCEDURES USED TO NUMERICALLY SOLVE THE MODEL PARTIAL-DIFFERENTIAL EQUATIONS OF MICROELECTRONICS</b>                                                      | <b>59</b> |
| 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 <i>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.</i> |           |
| 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        |
| <b>2.11 SYSTEMS OF COUPLED EQUATIONS</b>                                                                                                                                           | <b>74</b> |
| 2.11.1 Gummel's Iterative Method                                                                                                                                                   | 75        |
| 2.11.2 Newton's Direct Method                                                                                                                                                      | 76        |
| <b>2.12 EVOLUTION OF MOSFET DEVICE SIMULATORS</b>                                                                                                                                  | <b>77</b> |
| 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        |
| <b>REFERENCES</b>                                                                                                                                                                  | <b>81</b> |
| <br><b>Chap. 3 - MOS TRANSISTOR DEVICE PHYSICS:</b>                                                                                                                                | <b>83</b> |
| <b>PART 1 - BASICS MOS PHYSICS and THE MOS CAPACITOR</b>                                                                                                                           |           |
| <b>3.1 BASIC MOS THEORY and PHYSICS of IDEAL MOS CAPACITORS (MOS-C)</b>                                                                                                            | <b>84</b> |
| 3.1.1 The Structure of the MOS-C and the MOSFET                                                                                                                                    | 85        |
| 3.1.1.1 <i>Comparison of MOS-Cs and Conventional Capacitor Structures.</i>                                                                                                         |           |
| 3.1.1.2 <i>Why is the MOS-C Studied First?</i>                                                                                                                                     |           |

**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 1033.3.4 More Accurate and General Model  
for MOS-C Behavior 1063.3.5 Charge-Sheet Model for Determining the Characteristics of the  
Ideal MOS-C 1123.3.5.1 *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<sub>2</sub>-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<sub>2</sub> 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.                                     |                |
| <b>4.6 THRESHOLD-VOLTAGE CONTROL IN MOSFETS</b>                                           | <b>182</b>     |
| 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            |
| <b>4.7 SUBTHRESHOLD CURRENTS IN LONG-CHANNEL MOSFETS</b>                                  | <b>191</b>     |
| ( $I_{DST}$ when $V_{GS} < IV_T$ )                                                        |                |
| 4.7.1 Subthreshold Swing $S_t$                                                            | 196            |
| 4.7.2 Gate-Induced Drain Leakage (GIDL)                                                   | 198            |
| <b>4.8 SUMMARY OF LONG-CHANNEL MOSFET BEHAVIOR</b>                                        | <b>200</b>     |
| <b>REFERENCES</b>                                                                         | <b>201</b>     |
| <br><b>Chap. 5 - MOS TRANSISTOR DEVICE PHYSICS:</b>                                       | <br><b>205</b> |
| <b>PART 3 - THE SUBMICRON MOSFET</b>                                                      |                |
| <b>5.1 DEVICE CHARACTERISTICS OF LONG-CHANNEL VERSUS SHORT-CHANNEL MOSFETS</b>            | <b>207</b>     |
| <b>5.2 EFFECT OF GATE DIMENSIONS ON THRESHOLD VOLTAGE</b>                                 | <b>208</b>     |
| 5.2.1 Short-Channel Threshold-Voltage Shift                                               | 208            |
| 5.2.1.1 Numerical Solution of the 2-D Poisson Equation for Calculating $\Delta V_{TSC}$ . |                |
| 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 $\Delta V_{TSC}$ .                  |                |
| 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 <i>Model which Proposes RSCE is Due to Lateral Dopant Nonuniformity at the Channel Si-SiO<sub>2</sub> Interface (Arising from Enhanced Diffusion of Channel Dopants Caused by Interstitial Injection During Poly Reox, Salicide Formation, or Implant Damage).</i> |            |
| 5.2.3.2 <i>Model Which Proposes RSCE Arises from Boron Segregation to Implant Damaged Regions at the Edge of the Channel.</i>                                                                                                                                              |            |
| 5.2.3.3 <i>Model Which Attributes RSCE to Transient Enhanced Diffusion of Channel Dopants to the Silicon Surface Arising from Implant Damage.</i>                                                                                                                          |            |
| 5.2.3.4 <i>Model which Attributes RSCE to Silicon Interstitial Capture in the Gate Oxide.</i>                                                                                                                                                                              |            |
| <b>5.3 SHORT-CHANNEL EFFECTS ON SUBTHRESHOLD SWING, S<sub>t</sub> (Subsurface Punchthrough)</b>                                                                                                                                                                            | <b>232</b> |
| 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        |
| <b>5.4 SHORT-CHANNEL EFFECTS ON THE I-V CHARACTERISTICS OF MOSFETS OPERATED IN THE STRONG THRESHOLD REGIME</b>                                                                                                                                                             | <b>247</b> |
| 5.4.1 Impact of the MOSFET Electric Fields on Mobility                                                                                                                                                                                                                     | 248        |
| 5.4.1.1 <i>Models of Mobility Degradation Due to Gate-Induced Electric Fields.</i>                                                                                                                                                                                         |            |
| 5.4.1.2 <i>Models of the Carrier-Velocity Dependence on Drain Electric Field (Velocity Saturation).</i>                                                                                                                                                                    |            |
| 5.4.2 A Simple, Analytic DC Circuit Model of the Short-Channel MOSFET in Strong Inversion                                                                                                                                                                                  | 254        |
| 5.4.2.1 <i>DC Circuit Model of the Short-Channel MOSFET in the Linear Regime.</i>                                                                                                                                                                                          |            |
| 5.4.2.2 <i>Effects of Scaling on g<sub>msat</sub>.</i>                                                                                                                                                                                                                     |            |
| 5.4.2.3 <i>Comparing the Values of I<sub>D</sub> Calculated with the Long- and Short-Channel MOSFET Models as a Function of Channel Length.</i>                                                                                                                            |            |
| 5.4.3 Modeling the Saturation Region of Operation in Short-Channel MOSFETs                                                                                                                                                                                                 | 260        |
| 5.4.3.1 <i>Modeling Short-Channel MOSFET Saturation Behavior Using the Abrupt Junction and Depletion Approximations.</i>                                                                                                                                                   |            |
| 5.4.3.2 <i>Constant Field-Gradient Model.</i>                                                                                                                                                                                                                              |            |
| 5.4.3.3 <i>Pseudo-Two-Dimensional Model.</i>                                                                                                                                                                                                                               |            |
| <b>5.5. SHORT-CHANNEL MOSFET MODELS USED IN THE SPICE CIRCUIT SIMULATOR</b>                                                                                                                                                                                                | <b>271</b> |

|                   |                                                                                                                                        |            |
|-------------------|----------------------------------------------------------------------------------------------------------------------------------------|------------|
| 5.5.1             | BSIM Model Equations                                                                                                                   | 272        |
| 5.5.1.1           | <i>BSIM Threshold Voltage Equation.</i>                                                                                                |            |
| 5.5.1.2           | <i>BSIM Equations for <math>I_D</math> when the MOSFET is Operated in Strong Inversion.</i>                                            |            |
| 5.5.1.3           | <i>BSIM Equations for <math>I_D</math> when the MOSFET is Operated in Weak Inversion (Subthreshold Regime).</i>                        |            |
| <b>5.6</b>        | <b>MOSFET SCALING</b>                                                                                                                  | <b>274</b> |
| 5.6.1             | Design of Submicron MOSFETs: Scaling Guidelines                                                                                        | 274        |
| 5.6.1.1           | <i>Constant Electric-Field, Constant Voltage &amp; Electrostatic Scaling.</i>                                                          |            |
| 5.6.1.2           | <i>Subthreshold Scaling.</i>                                                                                                           |            |
| 5.6.1.3           | <i>Off-Current Scaling Approach.</i>                                                                                                   |            |
| 5.6.2             | Manufacturing Considerations in Submicron MOSFET Device Design                                                                         | 285        |
| <b>5.7</b>        | <b>DESIGN OF SUBMICRON MOSFETS</b>                                                                                                     | <b>286</b> |
| 5.7.1             | Design Methodology Based on the Subthreshold Scaling Equation                                                                          | 286        |
| 5.7.2             | Design Methodology Based on Required Off-Current Value                                                                                 | 287        |
| <b>5.8</b>        | <b>p-CHANNEL DEVICES IN CMOS</b>                                                                                                       | <b>288</b> |
| 5.8.1             | PMOS Devices with $n^+$ -Polysilicon Gates                                                                                             | 289        |
| 5.8.1.1           | <i>Punchthrough Susceptibility.</i>                                                                                                    |            |
| 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           | <i>Threshold-Voltage Roll-Off in Buried-Channel PMOS.</i>                                                                              |            |
| 5.8.2.2           | <i>Subthreshold Swing (<math>S_L</math>) in Buried-Channel PMOS.</i>                                                                   |            |
| 5.8.2.3           | <i>Punchthrough Susceptibility of Buried-Channel PMOS.</i>                                                                             |            |
| 5.8.3             | Techniques Used to Suppress Short-Channel Effects in Buried-Channel PMOS                                                               | 305        |
| 5.8.3.1           | <i>Techniques Used to Keep the Implanted p-Layer as Thin as Possible.</i>                                                              |            |
| 5.8.3.2           | <i>Shallow <math>p^+</math> Source/Drain Junctions to Suppress Punchthrough in PMOSFETs.</i>                                           |            |
| 5.8.3.3           | <i>LDD Structures in PMOS to Limit Punchthrough.</i>                                                                                   |            |
| 5.8.3.4           | <i>New PMOS Device Structures Designed to Suppress Punchthrough.</i>                                                                   |            |
| 5.8.4             | PMOS Devices with $p^+$ -Polysilicon Gates                                                                                             | 311        |
| <b>REFERENCES</b> |                                                                                                                                        | <b>313</b> |

|                                                                                                                                       |            |
|---------------------------------------------------------------------------------------------------------------------------------------|------------|
| <b>Chap. 6 - ISOLATION STRUCTURES<br/>IN SUBMICRON CMOS</b>                                                                           | <b>321</b> |
| <b>6.1 CHARACTERIZING MOS ISOLATION</b>                                                                                               |            |
| 6.1.1 Test Structures for NMOS Isolation Structures                                                                                   | 322        |
| 6.1.2 Punchthrough Prevention between<br>Adjacent Devices in MOS Circuits                                                             | 325        |
| <b>6.2 THE BASIC LOCOS ISOLATION PROCESS</b>                                                                                          | <b>330</b> |
| 6.2.1 Details of the Basic Semi-recessed LOCOS Process                                                                                | 332        |
| 6.2.1.1 <i>Pad-Oxide Layer.</i>                                                                                                       |            |
| 6.2.1.2 <i>CVD of Silicon Nitride Layer.</i>                                                                                          |            |
| 6.2.1.3 <i>Mask and Etch Pad-Oxide/Nitride Layer to Define Active Regions.</i>                                                        |            |
| 6.2.1.4 <i>Channel-Stop Implant.</i>                                                                                                  |            |
| 6.2.1.5 <i>Problems Arising from the Channel-Stop Implants.</i>                                                                       |            |
| 6.2.1.6 <i>Grow Field Oxide.</i>                                                                                                      |            |
| 6.2.1.7 <i>Factors Which Impact Bird's Beak Length and Shape.</i>                                                                     |            |
| 6.2.1.8 <i>Use of High-Energy Ion Implantation to Introduce the<br/>                Channel-Stop Dopant After Field Oxide Growth.</i> |            |
| 6.2.1.9 <i>Strip the Masking Nitride/Pad-Oxide Layer.</i>                                                                             |            |
| 6.2.1.10 <i>Regrow Sacrificial Pad Oxide and Strip (Kooi Effect).</i>                                                                 |            |
| 6.2.2. Planarizing Semirecessed LOCOS Structures                                                                                      | 348        |
| <b>6.3 MODELING THE TOPOGRAPHY OF LOCOS STRUCTURES</b>                                                                                | <b>349</b> |
| 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<br>Using Physical Models                                                                    | 355        |
| <b>6.4 POLY-BUFFERED LOCOS</b>                                                                                                        | <b>357</b> |
| 6.4.1. Modified Poly-Buffered LOCOS Processes                                                                                         | 361        |
| 6.4.1.1 <i>Framed-Mask PBL (FMPBL)</i>                                                                                                |            |
| 6.4.1.2 <i>Polysilicon-Encapsulated LOCOS (PELOX)</i>                                                                                 |            |
| 6.4.1.3 <i>Nitride-Clad LOCOS (NCL)</i>                                                                                               |            |
| 6.4.1.4 <i>Mitigating the Field-Thinning Problem of LOCOS</i>                                                                         |            |
| <b>6.5 SHALLOW TRENCH AND REFILL ISOLATION</b>                                                                                        | <b>367</b> |
| 6.5.1 Early Shallow Trench and Refill Isolation                                                                                       | 368        |
| 6.5.2 Improved Shallow Trench<br>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  $\mu\text{m}$  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: 421**  
**GROWTH AND RELIABILITY**

**7.1 GATE OXIDE CHARACTERISTICS NEEDED 422**  
**FOR SUBMICRON MOSFETS**

**7.2 PHYSICAL PICTURE OF  $\text{SiO}_2$  423**  
**AND THE  $\text{Si/SiO}_2$  INTERFACE**

- 7.2.1 Physical and Chemical Properties of  $\text{SiO}_2$  423
- 7.2.2 The  $\text{Si/SiO}_2$  Interface 425

**7.3 CARRIER INJECTION IN THE  $\text{Si/SiO}_2$  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**

|                                                                                                            |            |
|------------------------------------------------------------------------------------------------------------|------------|
| <b>Chap. 8 - WELL FORMATION IN CMOS</b>                                                                    | <b>523</b> |
| <b>8.1 <i>p</i>-WELL, <i>n</i>-WELL, and TWIN-WELL TECHNOLOGIES</b>                                        | <b>524</b> |
| 8.1.1 <i>p</i> -Well CMOS                                                                                  | 525        |
| 8.1.2 <i>n</i> -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        |
| <b>8.2 CONVENTIONAL FORMATION OF WELLS AND CHANNEL STOPS</b>                                               | <b>532</b> |
| 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        |
| <b>8.3 RETROGRADE-WELL CMOS</b>                                                                            | <b>539</b> |
| 8.3.1 Retrograde Well Technologies to Improve Isolation                                                    | 546        |
| <b>8.4 FUTURE TRENDS IN WELL FORMATION</b>                                                                 | <b>552</b> |
| <b>8.5 AFTER-GATE IMPLANTATION (AGI)</b>                                                                   | <b>554</b> |
| <b>REFERENCES</b>                                                                                          | <b>555</b> |
| <b>Chap. 9 - HOT-CARRIER RESISTANT PROCESSING AND DEVICE STRUCTURES</b>                                    | <b>559</b> |
| <b>9.1 MODEL FOR THE MAXIMUM LATERAL ELECTRIC FIELD <math>E_{y\max}</math> IN NON-GRADED-DRAIN MOSFETS</b> | <b>561</b> |
| <b>9.2 SUBSTRATE CURRENTS DUE TO HOT CARRIERS</b>                                                          | <b>563</b> |
| <b>9.3 MODELS OF HOT-CARRIER DEGRADATION PHENOMENA</b>                                                     | <b>566</b> |
| 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        |
| <b>9.4 CHARACTERIZING MOSFET DEGRADATION DUE TO HOT-CARRIER EFFECTS</b>                                    | <b>572</b> |
| 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    | <i>Selecting a Criterion that Identifies Device Failure Due to Hot-Carrier Damage.</i>                                                            |            |
| 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    | <i>SCALE.</i>                                                                                                                                     |            |
| 9.4.6.2    | <i>Simulators that Relate the Device-Level Degradation Parameters to Circuit Performance (CAS and HOTRON).</i>                                    |            |
| <b>9.5</b> | <b>DECREASING HOT CARRIER DEGRADATION BY MODIFYING THE MOSFET STRUCTURE</b>                                                                       | <b>586</b> |
| 9.5.1      | Phosphorus-Drain Structure                                                                                                                        | 588        |
| 9.5.2      | Double Diffused Drain (DDD)                                                                                                                       | 588        |
| <b>9.6</b> | <b>LIGHTLY DOPED DRAIN (LDD) STRUCTURES</b>                                                                                                       | <b>591</b> |
| 9.6.1      | Modeling $E_{\text{ymax}}$ in LDD MOSFETs                                                                                                         | 592        |
| 9.6.1.1    | <i>Simple Model for Estimating <math>E_{\text{ymax}}</math> in LDD MOSFETs</i>                                                                    |            |
| 9.6.1.2    | <i>Analytical Model for Calculating <math>E_{\text{ymax}}</math> in LDD MOSFETs Using the Pseudo-2D Approach to Solving the Poisson Equation.</i> |            |
| 9.6.1.3    | <i>Model for the Series Resistance Introduced by the <math>n^-</math> Region of the LDD.</i>                                                      |            |
| 9.6.2      | Overview of the Process Sequence to Form LDDs                                                                                                     | 595        |
| <b>9.7</b> | <b>PERFORMANCE AND HOT-CARRIER LIMITATIONS OF THE CONVENTIONAL LDD</b>                                                                            | <b>598</b> |
| 9.7.1      | Performance Degradation in LDD MOSFETs                                                                                                            | 599        |
| 9.7.1.1    | <i><math>I_D</math> Decrease Due to Added Resistance of <math>n^-</math> Regions.</i>                                                             |            |
| 9.7.1.2    | <i>Doping Compensation Effects that can Increase <math>R_n</math> in Deep Submicron LDD MOSFETs.</i>                                              |            |
| 9.7.1.3    | <i>Increase of <math>R_{sh}</math> by Weak Overlap of the Gate and Drain.</i>                                                                     |            |
| 9.7.1.4    | <i>The Impact of Poly Re-Ox on Weak Overlap in MOSFETs.</i>                                                                                       |            |
| 9.7.1.5    | <i>Off-Axis Implant Effects in MOSFETs, Leading to Weak Overlap Problems and Asymmetrical MOSFET Device Characteristics.</i>                      |            |
| 9.7.1.6    | <i>Implanter Issues that Impact the LDD Process Sequence.</i>                                                                                     |            |
| 9.7.1.7    | <i>Process integration issues which arise when an LDD is integrated into a CMOS Technology.</i>                                                   |            |
| 9.7.2      | Hot Carrier Reliability Issues in Early LDDs                                                                                                      | 606        |

|                                                                                                   |            |
|---------------------------------------------------------------------------------------------------|------------|
| <b>9.8 DISPOSABLE SPACERS</b>                                                                     | <b>608</b> |
| <b>9.9 SUMMARY OF THE LIMITATIONS OF EARLY LDD DEVICES</b>                                        | <b>611</b> |
| <b>9.10 SUMMARY OF THE METHODS TO BE USED IN DESIGNING EFFECTIVE LDD STRUCTURES</b>               | <b>612</b> |
| <b>9.11 IMPROVED CONVENTIONAL (OXIDE-SPACER-BASED) LDD STRUCTURES</b>                             | <b>613</b> |
| 9.11.1 Increasing the Doping in the $n^+$ Region of the LDD                                       | 614        |
| 9.11.1.1 <i>Moderate Lightly Doped Drain (M-LDD).</i>                                             |            |
| 9.11.1.2 <i>Metal-coated LDD (MLD).</i>                                                           |            |
| 9.11.2 Tailoring the Vertical Doping Profile of the $n^+$ Implant to Improve the Conventional LDD | 616        |
| 9.11.2.1 <i>Buried-LDD (B-LDD).</i>                                                               |            |
| 9.11.2.2 <i>Profiled-LDD (P-LDD).</i>                                                             |            |
| 9.11.3 Improving the Short-Channel Performance of the Conventional NMOS LDD                       | 621        |
| <b>9.12 FULLY OVERLAPPED LDD STRUCTURES</b>                                                       | <b>623</b> |
| 9.12.1 Inverse-T Gate LDD (ITLDD) and Gate-Drain Overlapped Device (GOLD) Structures              | 623        |
| 9.12.1.1 <i>Inverse-T Gate LDD Structure.</i>                                                     |            |
| 9.12.1.2 <i>Gate-Drain Overlapped LDD (GOLD).</i>                                                 |            |
| 9.12.1.3 <i>Modified IT-LDD Structures.</i>                                                       |            |
| 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        |
| <b>9.13 HOT-CARRIER EFFECTS IN PMOS TRANSISTORS</b>                                               | <b>637</b> |
| 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        |
| <b>9.14 STRENGTHENING OXIDES AGAINST HOT-CARRIER INJECTION</b>                                    | <b>641</b> |
| 9.14.1 Minimizing Hydrogen and $H_2O$ at the Si-SiO <sub>2</sub> Interface                        | 643        |
| 9.14.1.1 <i>The role of hydrogen in hot-carrier degradation.</i>                                  |            |

|          |                                                                                                         |     |
|----------|---------------------------------------------------------------------------------------------------------|-----|
| 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 $\text{SiO}_2$ .                                                                 |     |
| 9.14.3.2 | $\text{NH}_3$ Oxides as MOSFET Gate Dielectrics.                                                        |     |
| 9.14.3.3 | Low-Pressure and RTP $\text{NH}_3$ Oxides (Lightly Nitrided Reoxidized $\text{NH}_3$ -Nitrided Oxides). |     |
| 9.14.3.4 | Oxynitridation of Silicon in $\text{N}_2\text{O}$ Ambients.                                             |     |
| 9.14.3.5 | Oxynitridation of Oxides in $\text{N}_2\text{O}$ .                                                      |     |
| 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

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 shrunk 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

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.<sup>1</sup>

### 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 ( $I_{sub} \approx 0$ ).<sup>\*</sup> 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<sub>2</sub>

<sup>\*</sup> 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

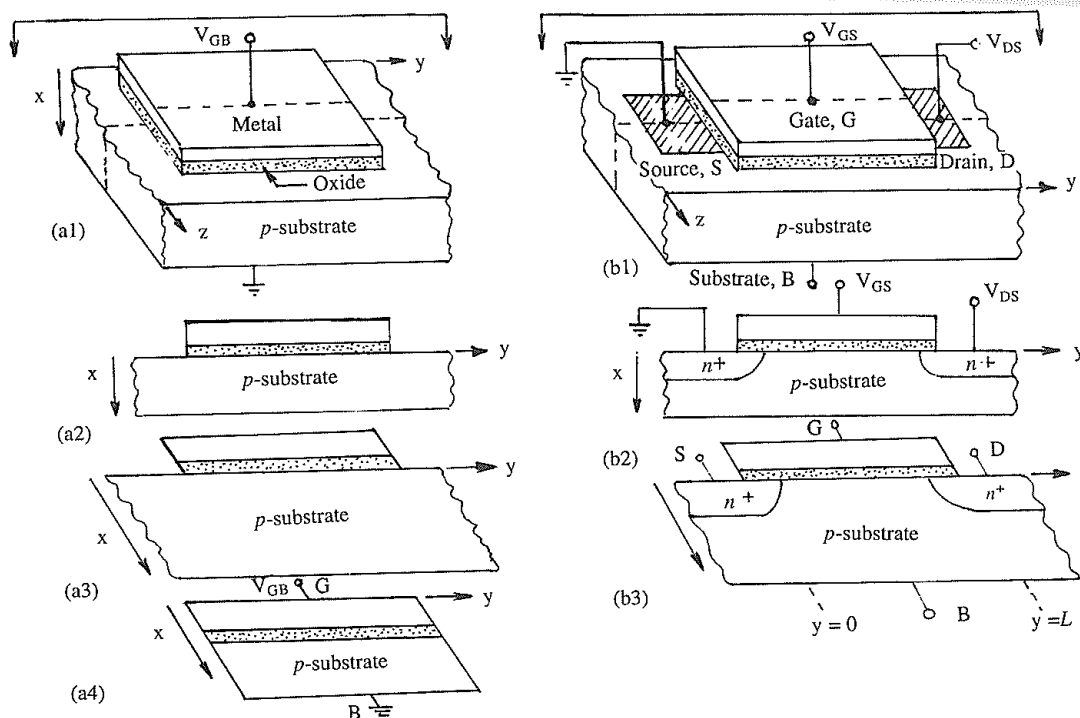


**Fig. 3-1** (a1) Perspective view of the MOS-C, (a2) Cross-section of the MOS-C cut along the middle of the structure, (a3) Cross-section shown lying flat, (a4) Cross-section with only the Si beneath the oxide shown. (b1) Perspective view of the MOSFET structure, (b2) Cross-section of the MOSFET cut 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 ( $n_0$  and  $p_0$ , 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:

$$n_0 = n_i \exp (E_F - E_i)/kT \quad (3 - 1a)$$

$$p_0 = n_i \exp (E_i - E_F)/kT \quad (3 - 1b)$$

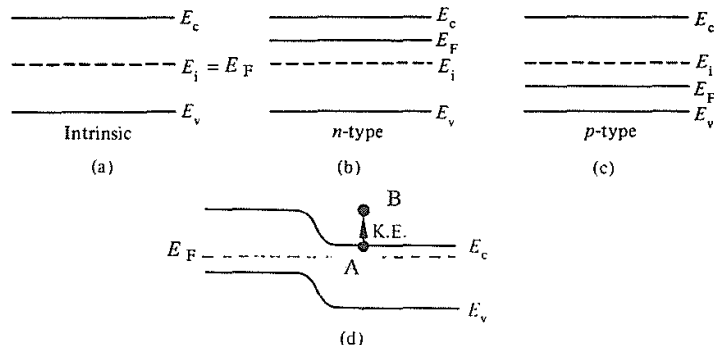


Fig. 3-2 Energy band representation of a semiconductor: (a) intrinsic; (b)  $n$ -type; (c)  $p$ -type; (d) Depicting the kinetic energy of an electron in the energy band diagram.

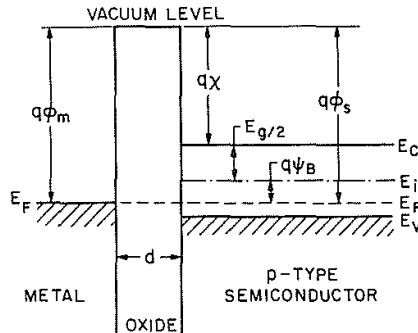


Fig. 3-3 Energy band diagram of an ideal MOS-C when  $V_{GB} = 0V$  (flat-band condition).

Note also that at equilibrium

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

### 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),  $q\phi_m$  and that of the semiconductor,  $q\phi_s$  is zero (i.e.,  $q\phi_{ms} = 0$ );
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  $q\phi_{ms} = 0$  (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, its 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

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

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  $C_{ox} = \kappa_{ox}\epsilon_0/t_{ox}$ , where  $\kappa_{ox}$  and  $t_{ox}$  are the dielectric constant and thickness of the oxide, and  $\epsilon_0$  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.

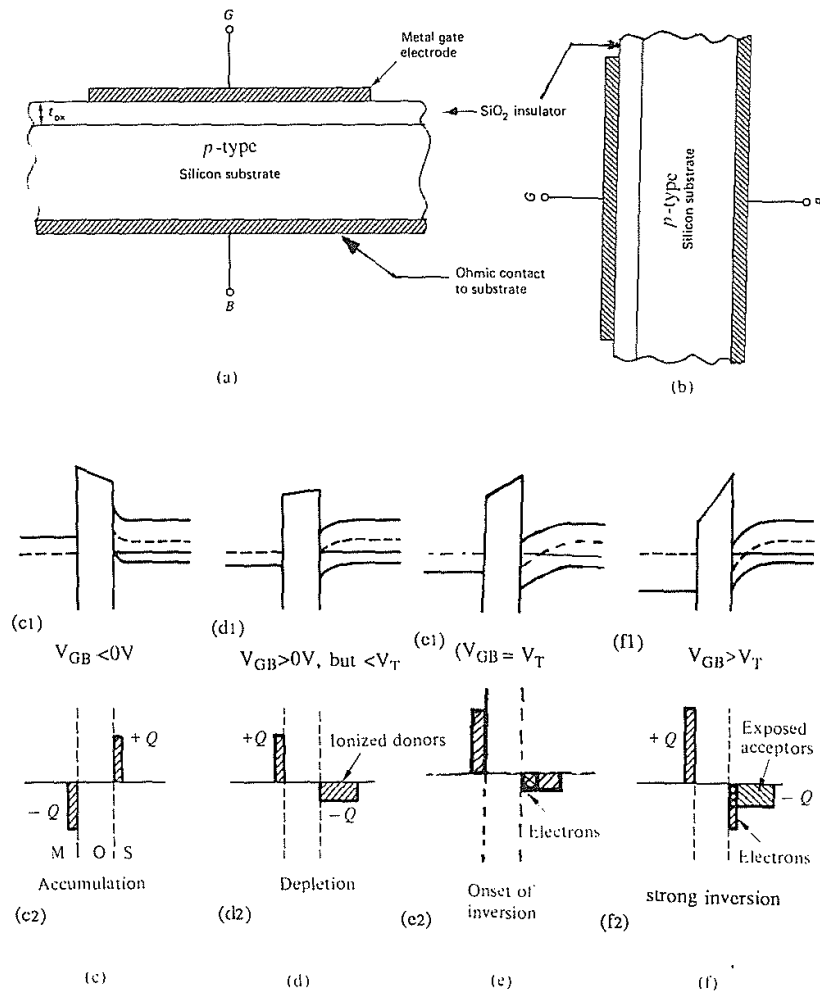


Fig. 3-4 (a) Cross-section of a MOS-C with p-substrate; (b) Cross-section of the same MOS-C, but rotated 90°; (c1) and (c2) Energy band and block charge diagrams when the MOS-C is biased to accumulation ( $V_{GB} < 0V$ ); (d1) and (d2) Same diagrams for depletion biasing ( $V_{GB} > 0V$ , but  $< V_T$ ); (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  $C_d = \kappa_{si}\epsilon_0/d$ . 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

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

Solving this equation for  $E_i(\text{surface})$ , and using the result in Eq. 3-1a, the concentration of electrons at the Si surface at the onset of strong inversion,  $n_{surf}(\text{inv})$  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,  $p_0 = N_A$ . 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  $Q_C$

$$Q_G \text{ (C/cm}^2\text{)} = -Q_C \text{ (C/cm}^2\text{)} \quad (3 - 5)$$

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 ( $Q_{B\max}$ ). Thus, in strong inversion

$$Q_G = -Q_C = -(Q_I + Q_{B\max}) \quad (3 - 6)$$

where  $Q_{B\max} = -qN_A d_{\max}$ . (Note that  $Q$ 's in this chapter are expressed in units of charge/unit area - specifically, C/cm<sup>2</sup>.) 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  $C_{d\max} = \kappa_{si}\epsilon_0/d_{\max}$ . 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<sup>-2</sup>.

### 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.,  $\epsilon_{si}$  and  $\epsilon_{ox}$ ):

1. The electrostatic potential as a function of position  $\phi(x)$ ;
2. The surface potential,  $\phi_{surf}$ , which is the potential at  $x = 0$ , or  $\phi(0)$
3. The electric field as a function of position  $E(x)$ ;
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-SiO<sub>2</sub> 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  $\phi$  (e.g., the voltage drop across the oxide  $\phi_{ox}$  and the voltage drop across the semiconductor from the Si surface to the bulk,  $\phi_{surf}$ ). 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*  $\phi$ . 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*  $\phi_{surf}$  (henceforth referred to as the *surface potential*), and;
- b. An expression for the electrostatic potential versus position  $\phi(x)$  [which uses  $\phi_{surf}$  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  $\Delta\phi$  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

$$\Delta\phi = \Delta E/(-q) \quad (3 - 7)$$

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).<sup>\*</sup> 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  $\phi$  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 [ $E_i(\text{bulk})$ ] to designate this level of zero on an energy-band diagram. Thus, elsewhere, the value of  $\phi(x)$  is measured with respect to  $E_i(\text{bulk})$  according to

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

\* 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.

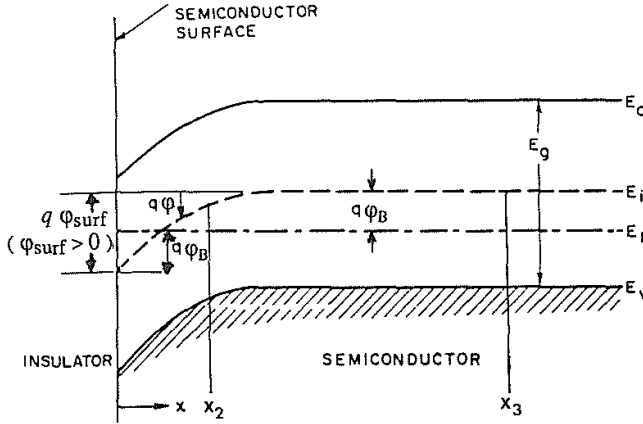


Fig. 3-5 The energy-band and potential diagram representations of the MOS-C (in this figure a MOS-C with a  $p$ -type substrate and biased to inversion is depicted).

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,  $\phi(x_2) \neq 0$ , but instead is  $\phi(x_2) = [E_i(\text{bulk}) - E_i(x_2)]/q$ .

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

$$V_{GB} = \phi_{ox} + \phi_{si} \quad (3-9)$$

If  $\phi(x = \infty) = 0$ , then the potential at the surface of the silicon ( $\phi_{surf}$ ) 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

$$V_{GB} = \phi_{ox} + \phi_{surf} \quad (3-10)$$

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

Let us also introduce another dimensionless parameter called the *doping parameter*  $\phi_B$  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

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

in the bulk region of the Si in the MOS capacitor, see Fig. 3-5). Note that  $\phi_B > 0$  for a  $p$ -type semiconductor and  $\phi_B < 0$  for an  $n$ -type semiconductor. For substrate doping concentrations usually encountered in MOSFETs ( $10^{14}$ - $10^{17} \text{ cm}^{-3}$ ), the magnitude of  $\phi_B$  is between  $9kT/q$  and  $16kT/q$  (where  $kT/q$  at 300K is  $\sim 0.025 \text{ V}$ ); that is, the magnitude of  $\phi_B$  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  $\phi_B$ . 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,  $\phi_{\text{surf}} = \phi_B$ . Since the onset of *strong inversion* is the condition when the surface electron concentration equals  $N_A$ , at that point  $E_i(\text{surf}) - E_i(\text{bulk}) = 2 [E_F - E_i(\text{bulk})]$ . By combining this fact with Eqs. 3-4 and 3-11, we see that the surface potential at the onset of strong inversion  $\phi_{\text{surf(inv)}} = 2\phi_B$ .

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

$$n_p(x) = n_{p0} \exp [q\phi(x)/kT] \quad (3 - 12a)$$

$$p_p(x) = p_{p0} \exp [-q\phi(x)/kT] \quad (3 - 12b)$$

where  $n_{p0}$  and  $p_{p0}$  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  $n_{\text{surf}}$  and  $p_{\text{surf}}$  are given by

$$n_{\text{surf}} = n_{p0} \exp (q\phi_{\text{surf}}/kT) \quad (3 - 13a)$$

$$p_{\text{surf}} = p_{p0} \exp (-q\phi_{\text{surf}}/kT) \quad (3 - 13b)$$

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

$$n_{\text{surf}} = N_A \exp [q(\phi_{\text{surf}} - 2\phi_B)/kT] \quad (3 - 13c)$$

This expression is plotted versus  $\phi_{\text{surf}}$  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  $\phi_{\text{surf}} = \phi_B$ ,  $n_{\text{surf}}$  becomes equal to the intrinsic concentration,  $n_i$ . Thus,  $\phi_{\text{surf}} = \phi_B$  is defined as the limit point between the regions of depletion and weak inversion. Second, as  $\phi_{\text{surf}}$  increases above  $\phi_B$ ,  $n_{\text{surf}}$  grows very rapidly. Eventually, when  $\phi_{\text{surf}} = 2\phi_B$ ,  $n_{\text{surf}} = N_A$ . This is commonly defined as the demarcation point between weak and strong inversion. Finally, beyond this point (i.e.,  $\phi_{\text{surf}} > 2\phi_B$ ),  $n_{\text{surf}}$  continues to increase rapidly, becoming much greater than  $N_A$  for small increases of  $\phi_{\text{surf}}$  beyond  $2\phi_B$ . 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  $N_A$ .

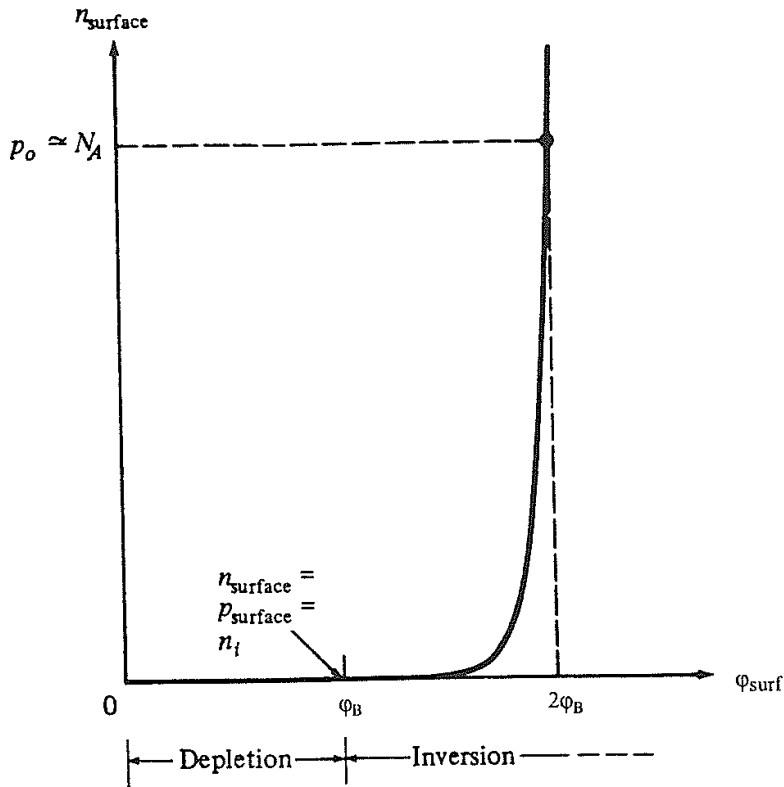


Fig. 3-6 Electron concentration at the surface of a MOS-C (with a  $p$ -type substrate) as a function of  $\phi_{\text{surf}}$ .<sup>6</sup> Linear axes are used. Reprinted with permission of McGraw-Hill Book Co.

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

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  $E_{\text{ox}} = -d\phi_{\text{ox}}/dx = \text{constant}$ . Thus, the voltage across the oxide  $\phi_{\text{ox}}$  is found from:

$$\phi_{\text{ox}} = - \int_{-t_{\text{ox}}}^0 E_{\text{ox}} dx = t_{\text{ox}} E_{\text{ox}} \quad (3-14)$$

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<sub>2</sub> interface of an ideal MOS-C is assumed to be zero, the electric displacement,  $D = \epsilon \mathcal{E}$  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

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

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

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

and this result can be substituted for  $\phi_{ox}$  in Eq. 3-10. In order to complete our goal of finding an expression for  $\phi_{surf}$  in terms of  $V_{GB}$  we see that an expression for  $\mathcal{E}_{si}(x = 0)$  in terms of  $\phi_{surf}$  is needed. Since  $\mathcal{E}(x) = -d\phi/dx$ , 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  $\phi(x = 0) = \phi_{surf}$ . We will carry out this integration for three cases in the upcoming sections.

2. We next solve the Poisson equation to find  $\phi(x)$ . 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  $\phi$  in a MOS-C to be one-dimensional (i.e., we assume  $\partial\phi/\partial y$  and  $\partial\phi/\partial z = 0$ ). Thus, it is only necessary to solve the Poisson equation in its one-dimensional form. We use the value of  $\phi_{surf}$  from step 1 as one boundary condition [ $\phi(x = 0) = \phi_{surf}$ ], and [ $\phi(x = \infty) = 0$ ] as the other, when we solve the Poisson equation to find  $\phi(x)$ .

3. To calculate the mobile charge concentrations  $n_p(x)$  and  $p_p(x)$ , we can use the values of  $\phi(x)$  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\*

$$Q_C = -\kappa_{si}\epsilon_0 \mathcal{E}_{si}(x = 0) \quad (3 - 17)$$

\* Note that only the  $\mathcal{E}$ -field at the Si-SiO<sub>2</sub> 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  $C_{ox} = \kappa_{ox}\epsilon_0/t_{ox}$  that

$$\varphi_{ox} = t_{ox} (\kappa_{si}/\kappa_{ox}) \mathcal{E}_{si} (x=0) = \{\kappa_{si}\epsilon_0 \mathcal{E}_{si} (x=0)\}/C_{ox} \quad (3-18)$$

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  $\varphi_{surf}$ , 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  $x = d$ ,  $p = N_A = p_{p0}$ , and  $n = n_{p0}$ ].

The simplifying assumptions dealing with the  $d$  (depletion region width),  $\phi_{surf}$  (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,  $\phi_{surf}$  also remains fixed (or is *clamped*) at the value it has at the onset of inversion, namely at  $\phi_{surf} (inv) = 2\phi_B$ .
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  $\phi_{surf}$  beyond strong inversion, we can compute  $\phi_{ox}$  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  $\phi(x)$ . That is, under depletion biasing conditions we have assumed  $n = p = 0$ , and so the charge density  $\rho(x)$  is merely a constant term ( $q N_A / \kappa_{Si} \epsilon_0$ ). In this case, the Poisson equation has the form:

$$\begin{aligned} d^2\phi/dx^2 &= q N_A / \kappa_{Si} \epsilon_0 & 0 < x < d \\ &= 0 & x > d \end{aligned} \quad (3 - 22)$$

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

$$\phi(x) = \phi_{surf} (1 - x/d)^2 \quad (3 - 23)$$

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

$$\mathcal{E}(x) = -d\phi/dx = 2\phi_{\text{surf}}(1 - x/d)(-1/d) \quad (3-24)$$

and

$$d^2\phi/dx^2 = 2\phi_{\text{surf}}/d^2 \quad [\text{from (3-23)}] \quad (3-25)$$

$$= qN_A/\kappa_{\text{Si}}\epsilon_0 \quad [\text{from (3-22)}] \quad (3-26)$$

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

$$d = [\kappa_{\text{Si}}\epsilon_0(2\phi_{\text{surf}}/qN_A)]^{1/2}. \quad (3-27)$$

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

$$d_{\text{max}} = (4\kappa_{\text{Si}}\epsilon_0\phi_B/qN_A)^{1/2} \quad (3-28)$$

Next, we solve Eq. 3-10 for  $\phi_{\text{surf}}$ . 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

$$Q_B = qN_A d \quad (3-29)$$

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

$$\phi_{\text{ox}} = -qN_A d/C_{\text{ox}} = \{[\kappa_{\text{Si}}\epsilon_0(2\phi_{\text{surf}}/qN_A)]^{1/2}\}/C_{\text{ox}} \quad (3-30)$$

Now if we let  $\eta = \{[2\kappa_{\text{Si}}\epsilon_0 qN_A]^{1/2}\}/C_{\text{ox}}$ , we can write  $\phi_{\text{ox}} = \eta\sqrt{\phi_{\text{surf}}}$ . Substituting this into Eq. 3-10 we get

$$V_{\text{GB}} = \phi_{\text{surf}} + \eta\sqrt{\phi_{\text{surf}}} \quad (3-31)$$

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

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

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

Now that  $\phi_{\text{surf}}$  has been obtained, we can find  $Q_B$  (from Eq. 3-29) and  $\phi(x)$  (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  $d_{\text{max}}$  from Eq. 3-28, and  $Q_{B\text{max}}$  from Eq. 3-29, (using  $d = d_{\text{max}}$ ). Then, since  $\phi_{\text{ox}} = V_{\text{GB}} - 2\phi_B$ , we can find  $Q_C$  from

$$Q_C = -(V_{\text{GB}} - 2\phi_B) C_{\text{ox}} \quad (3-33)$$

Now, since  $Q_I = Q_C - Q_{B\text{max}}$

$$Q_I = -[(V_{\text{GB}} - 2\phi_B) C_{\text{ox}}] - qN_A d_{\text{max}} \quad (3-34)$$

$$Q_I = -[(V_{GB} - 2\phi_B) C_{ox}] + \sqrt{4\kappa_{Si}\epsilon_0\phi_B q N_A} \quad (3 - 35)$$

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  $\phi$  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  $\phi$  versus  $x$  diagram curves upward, with the maximum value of  $\phi$  at the surface. However, for the *n*-type MOS-C a negative gate bias is needed. Thus, in such cases the  $\phi$  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  $Q_I \gg Q_B$  in strong inversion and the inversion layer is quite thin ( $\sim 30\text{-}300\text{\AA}$  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  $\phi(x)$ ,  $\phi_{surf}$ ,  $\mathcal{E}(x)$ , 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:

$$\frac{d^2\phi}{dx^2} = -\left(\frac{q}{\kappa_{Si}\epsilon_0}\right)[-N_A + N_D + p - n] \quad (3 - 36)$$

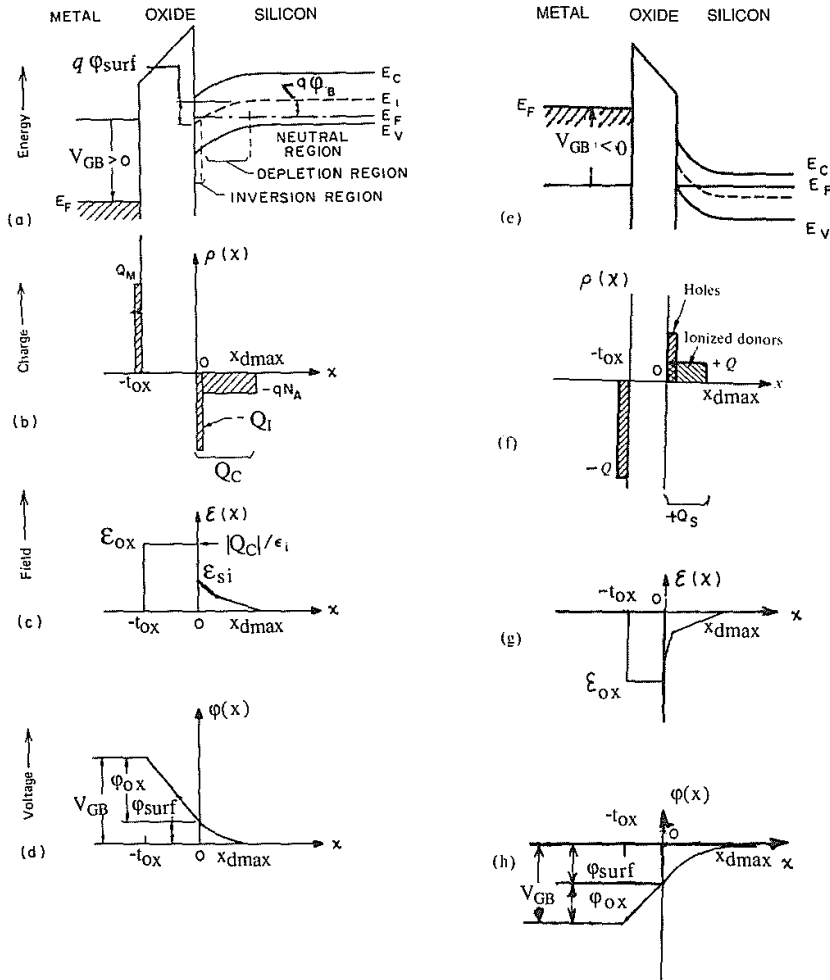


Fig. 3-7 Ideal *p*-type Si MOS-C when biased into inversion [ $V_{GB} > V_T$ ]: (a) Energy-band diagram; (b) Charge distribution; (c) Electric field; and (d) Potential distribution; (e) - (h) Comparable diagrams for an ideal *n*-type Si MOS-C.

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

$$\frac{d^2\phi}{dx^2} = - \left( \frac{q}{\kappa_{Si}\epsilon_0} \right) [-N_A + p_{po} e^{-q\phi/kT} + N_D - n_{po} e^{q\phi/kT}] \quad (3-37)$$

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

$$\frac{1}{2} \frac{d}{dx} \left( \frac{d\phi}{dx} \right)^2 = - \left( \frac{q}{\kappa_{Si}\epsilon_0} \right) [-N_A + p_{po} e^{-q\phi/kT} + N_D - n_{po} e^{q\phi/kT}] \left( \frac{d\phi}{dx} \right) \quad (3-38)$$

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

$$\left( \frac{d\phi}{dx} \right)_{\infty}^2 - \left( \frac{d\phi}{dx} \right)_x^2 = - \left( \frac{2q}{\kappa_{Si}\epsilon_0} \right) \int_0^{\phi} [-N_A + p_{po} e^{-q\phi/kT} + N_D - n_{po} e^{q\phi/kT}] d\phi \quad (3-39)$$

If we note that  $p_{po} \approx N_A$ , then  $n_{po} = n_i^2/N_A$ . In addition, since  $\phi_t = q/kT$ , and  $\mathcal{E} = -d\phi/dx$ , 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

$$- \frac{d\phi}{dx} = \mathcal{E}_{Si}(x) = \sqrt{\frac{2q N_A}{\kappa_{Si}\epsilon_0}} \sqrt{\phi_t e^{-\phi/\phi_t} + \phi - \phi_t + e^{-2\phi_B/\phi_t} (\phi_t e^{-\phi/\phi_t} - \phi - \phi_t)} \quad (3-40)$$

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

$$\mathcal{E}_{Si}(x = 0) = \sqrt{\frac{2q N_A}{\kappa_{Si}\epsilon_0}} \sqrt{\phi_t e^{-\phi_{surf}/\phi_t} + \phi_{surf} - \phi_t + e^{-2\phi_B/\phi_t} (\phi_t e^{-\phi_{surf}/\phi_t} - \phi_{surf} - \phi_t)} \quad (3-41)$$

In addition, by applying Gauss' law we can use Eq. 3-41 to find  $Q_C$  - i.e., from Eq. 3-17 [ $Q_C = -\kappa_{Si}\epsilon_0 \mathcal{E}_{Si}(x = 0)$ ] we get:

$$Q_C = \sqrt{2q N_A \kappa_{Si}\epsilon_0} \sqrt{\phi_t e^{-\phi_{surf}/\phi_t} + \phi_{surf} - \phi_t + e^{-2\phi_B/\phi_t} (\phi_t e^{-\phi_{surf}/\phi_t} - \phi_{surf} - \phi_t)} \quad (3-42)$$

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

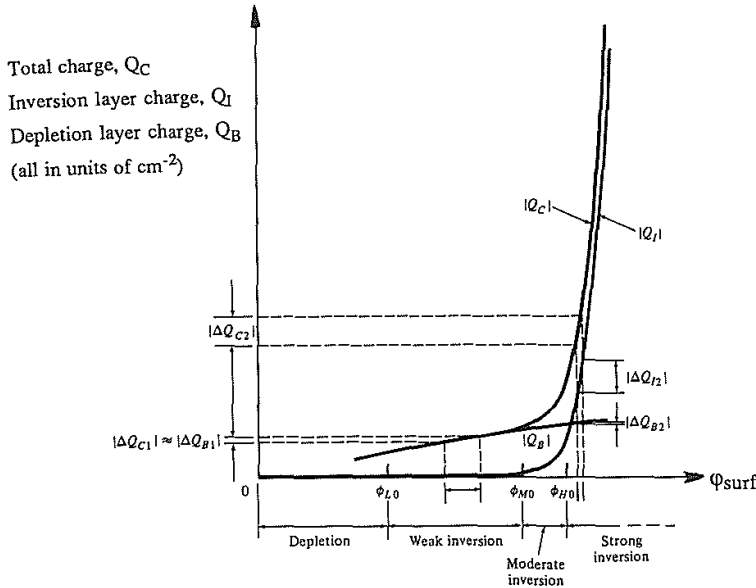


Fig. 3-8 Magnitude of inversion layer charge, depletion layer charge, and their sum (all per unit area) versus  $\phi_{\text{surf}}$ .<sup>6</sup> Reprinted with permission of McGraw-Hill Book Company.

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

To calculate  $\phi_{\text{surf}}$ , assuming we are given  $V_{GB}$ ,  $C_{ox}$  and  $N_A$ , we use

$$V_{GB} = \phi_{ox} + \phi_{\text{surf}} \quad (3-43)$$

and

$$\phi_{ox} = \{ \kappa_{si} \epsilon_0 \mathcal{E}_{si}(x=0) \} / C_{ox} \quad (3-44)$$

$$V_{GB} = \phi_{\text{surf}} + \frac{\sqrt{2q N_A \kappa_{si} \epsilon_0}}{C_{ox}} \sqrt{\phi_t e^{-\phi_{\text{surf}}/\phi_t} + \phi_{\text{surf}} - \phi_t + e^{-2\phi_B/\phi_t} (\phi_t e^{-\phi_{\text{surf}}/\phi_t} - \phi_{\text{surf}} - \phi_t)} \quad (3-45)$$

\* (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  $\phi_{\text{surf}}$  is varied is also included.

Note that to find  $\phi_{\text{surf}}$  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  $\phi$

$$Q_I = -q \int_0^{x_c} n(x) dx = -q \int_{\phi_{\text{surf}}}^{\phi_c} \frac{n(\phi) d\phi}{d\phi/dx} \quad (3-46)$$

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

$$Q_I = q \int_{\phi_{\text{surf}}}^{\phi_c} \frac{n_{po} e^{\phi/\phi_t}}{\sqrt{\frac{2q N_A}{\kappa_{Si} \epsilon_0}} \sqrt{\phi_t e^{-\phi/\phi_t} + \phi - \phi_t + e^{-2\phi_B/\phi_t} (\phi_t e^{-\phi/\phi_t} - \phi - \phi_t)}} d\phi \quad (3-47)$$

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

$$Q_B = q p_{po} \int_{\phi_{\text{surf}}}^{\phi_c} \frac{e^{-\phi/\phi_t} - 1}{\sqrt{\frac{2q N_A}{\kappa_{Si} \epsilon_0}} \sqrt{\phi_t e^{-\phi/\phi_t} + \phi - \phi_t + e^{-2\phi_B/\phi_t} (\phi_t e^{-\phi/\phi_t} - \phi - \phi_t)}} d\phi \quad (3-48)$$

To get  $\phi(x)$  we must integrate the Poisson equation a second time. That is, to find  $\phi$  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  $\phi$  to  $\phi_{\text{surf}}$  on the left:

$$- \int_{\phi}^{\phi_{\text{surf}}} \frac{d\phi}{\sqrt{\frac{2q N_A}{\kappa_{Si} \epsilon_0}} \sqrt{\phi_t e^{-\phi/\phi_t} + \phi - \phi_t + e^{-2\phi_B/\phi_t} (\phi_t e^{-\phi/\phi_t} - \phi - \phi_t)}} = x \quad (3-49)$$

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  $\phi(x)$  in this case. This was first done by Young in 1961.<sup>2</sup> 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  $\phi(x)$  versus  $x$  calculated using such numerical analysis techniques are shown for accumulation (Fig. 3-9a), depletion (Fig. 3-9b), onset of strong inversion,  $\phi_{\text{surf}} = 2\phi_B$  (Fig. 3-9c), and strong inversion,  $\phi_{\text{surf}} = 2\phi_B + 0.15V$  (Fig. 3-9d) for a MOS-C with  $\phi_B = 0.3V$ .<sup>3</sup>

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  $\sim 0.5 \mu m$ . The first observation supports the use of the following two approximations

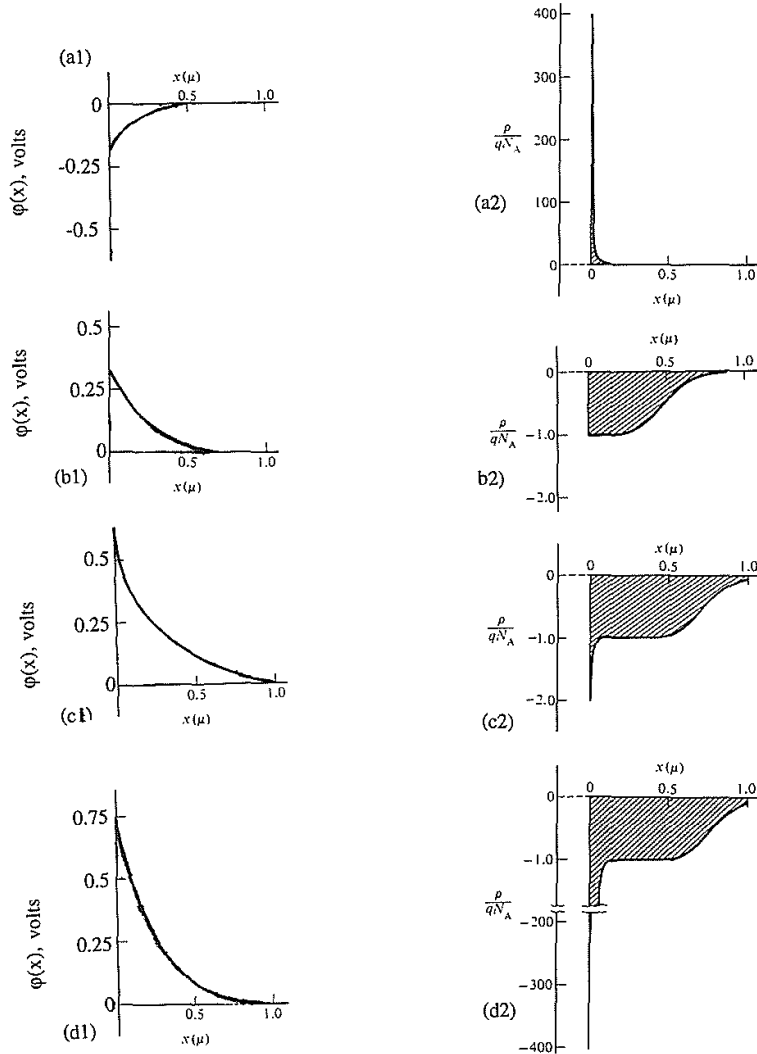


**Fig. 3-9** Exact solution for the potential  $\phi(x)$  in volts and normalized charge density  $(\rho/qN_A)$ , respectively, in the semiconductor region of a MOS-C with a  $p$ -type Si substrate, as a function of depth,  $x$ . Here we assume  $\phi_B = 0.3\text{V}$  and  $T = 300\text{K}$ : (a1 and a2) Accumulation ( $\phi_{\text{surf}} = -0.15\text{V}$ ); (b1 and b2) middle of depletion ( $\phi_{\text{surf}} = \phi_B = 0.3\text{V}$ ); (c1 and c2) Onset of inversion ( $\phi_{\text{surf}} = 2\phi_B = 0.6\text{V}$ ); and (d1 and d2) strong inversion ( $\phi_{\text{surf}} = 2\phi_B + 0.15\text{V} = 0.75\text{V}$ ). The charge concentration  $\rho/N_A$  diagrams are drawn on a linear scale.<sup>3</sup> (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  $\phi_{surf}$  using Eq. 3-45.
2. With this value of  $\phi_{surf}$  we can get  $Q_i$  from Eq. 3-47,  $\mathcal{E}(x)$  from Eq. 3-40, and  $\phi(x)$  from Eq. 3-49.
3. Once  $\phi(x)$  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 Bacarani et al.,<sup>4</sup> and Brews.<sup>5</sup> 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  $\phi_{surf} > \phi_B$  (unlike the previous model that assumed that no mobile charge exists in the channel until strong inversion, i.e., until  $\phi_{surf} > 2\phi_B$ );

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  $\phi_{surf}$  and  $\phi_{ox}$ ;
5. The value of  $\phi_{surf}$  beyond strong inversion is not clamped at exactly  $2\phi_B$ . 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  $E_{si}(x = 0)$ .  $Q_C$  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  $\phi_{surf} \geq \phi_B$  (where, as we noted in section 3.3.1,  $\phi_B$  is  $9\phi_t$  to  $16\phi_t$  for the range of doping concentrations used in MOSFETs), several terms in Eq. 3-42 become insignificant and can be dropped, yielding

$$Q_C \approx \sqrt{2q N_A \kappa_{Si} \epsilon_0} \sqrt{\phi_{surf} + \phi_t e(\phi_{surf} - 2\phi_B)/\phi_t} \quad (3 - 50)$$

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

$$d = [\kappa_{Si} \epsilon_0 (2\phi_{surf}) / q N_A]^{1/2} \quad (3 - 51)$$

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:

$$Q_B = q N_A d = (2\kappa_{Si} \epsilon_0 q N_A \phi_{surf})^{1/2} \quad (3 - 52)$$

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

$$Q_I = \sqrt{2q N_A \kappa_{Si} \epsilon_0} \left( \sqrt{\phi_{surf} + \phi_t e(\phi_{surf} - 2\phi_B)/\phi_t} - \sqrt{\phi_{surf}} \right) \quad (3 - 53)$$

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

$$V_{GB} = \phi_{surf} - Q_C/C_{ox} \quad (3-54)$$

or

$$V_{GB} = \phi_{surf} + \frac{\sqrt{2q N_A K_S \epsilon_0}}{C_{ox}} \sqrt{\phi_{surf} + \phi_t \exp(\phi_{surf} - 2\phi_B)/\phi_t} \quad (3-55)$$

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

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.,  $\phi_{surf} > \phi_B$ ). 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),<sup>6</sup> 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  $\phi_t \exp(\phi_{surf} - 2\phi_B)/\phi_t$  in Eq. 3-53 as small compared to

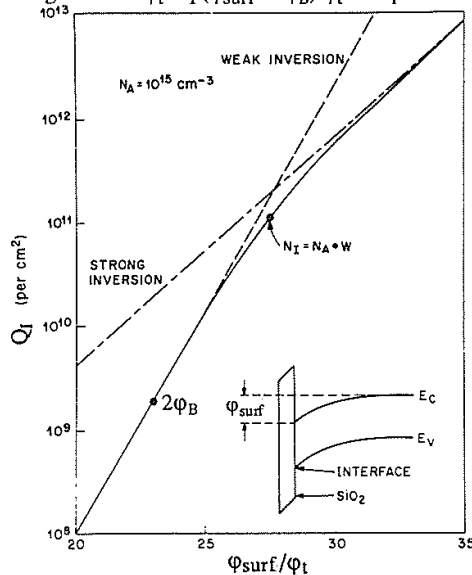


**Fig. 3-10** Inversion layer carrier density per unit area versus  $\phi_{surf}/\phi_t$  as computed from Eq. 3-53 and from approximations to this equation for the case of weak inversion (Eq. 3-58), and strong inversion (Eq. 3-62).<sup>5</sup> (© IEEE 1978).

the term  $\phi_{\text{surf}}$ . Then, if we let

$$\zeta = \phi_t \exp(\phi_{\text{surf}} - 2\phi_B)/\phi_t \quad (3-56)$$

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

$$\sqrt{\phi_{\text{surf}} + \zeta} \approx \sqrt{\phi_{\text{surf}}} + \frac{1}{2\sqrt{\phi_{\text{surf}}}} \zeta \quad (3-57)$$

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

$$Q_I (\text{weak inv}) \approx \frac{\sqrt{2q N_A \kappa_s \epsilon_0}}{2\sqrt{\phi_{\text{surf}}}} \phi_t e^{(\phi_{\text{surf}} - 2\phi_B)/\phi_t} \quad (3-58)$$

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

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

$$V_{GB} = \phi_{\text{surf}} - Q_B/C_{ox} \quad (3-59)$$

or using Eq. 3-52

$$V_{GB} = \phi_{\text{surf}} + [(\sqrt{2\kappa_{si}\epsilon_0 q N_A})/C_{ox}] \sqrt{\phi_{\text{surf}}} \quad (3-60)$$

$$= \phi_{\text{surf}} + \gamma \sqrt{\phi_{\text{surf}}} \quad (3-60a)$$

where  $\gamma = (\sqrt{2\kappa_{si}\epsilon_0 q N_A})/C_{ox}$ . Equation 3-60a can be solved for  $\phi_{\text{surf}}$  to yield

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

To relate  $Q_I$  to  $V_{GB}$ , we can substitute the value of  $\phi_{\text{surf}}$  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

$$Q_I (\text{strong inv}) \approx (\sqrt{2\kappa_{si}\epsilon_0 q} n_i \exp(\phi_{\text{surf}}/2\phi_t)) \quad (3-62)$$

The inversion-layer charge in strong inversion is thus seen to be an exponential function of the surface potential with a slope of  $1/(2\phi_t)$ . Therefore, a small increase in the surface potential induces a large change in  $Q_I$  (strong inv). This implies that  $\phi_{\text{surf}}$  will remain almost unchanged even in the face of a large increase in  $V_{GB}$  or  $Q_I$  beyond strong inversion. The approximation that  $\phi_{\text{surf}}$  (inv) is clamped at  $2\phi_B$  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  $\phi_{\text{surf}}$ (inv) is

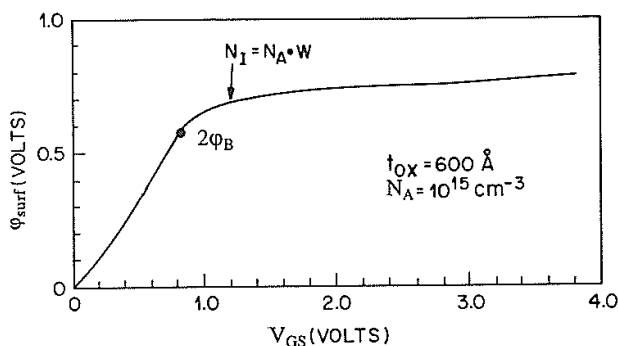


**Fig. 3-11**  $\phi_{surf}$  versus  $V_{GB}$  calculated from the charge sheet model for an oxide thickness of 60 nm, and a substrate doping level of  $N_A = 10^{15} \text{ cm}^{-3}$ . The dot indicates  $\phi_{surf} = 2\phi_B$ .<sup>5</sup> (© IEEE 1978).

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

$$Q_I \approx -C_{OX} (V_{GB} - V_T) \quad (3 - 63)$$

We should mention that  $\phi_{surf(inv)}$  does change slightly as  $V_{GB}$  is increased beyond strong inversion, as shown in Fig. 3-11, where  $\phi_{surf}$  from Eq. 3-55 is plotted versus  $V_{GB}$ . We see that  $\phi_{surf}$  in strong inversion actually increases beyond  $2\phi_B$  by ~150 mV before leveling off. Thus, models of  $Q_I$  that use  $\phi_{surf(inv)} = 2\phi_B$  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).<sup>6</sup> 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:

$$V_{GB} = \phi_{ox} + \phi_{surf} \quad (3 - 64)$$

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:

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

where  $N_{SUB}$  is the substrate doping (i.e.,  $N_A$  in  $p$ -type and  $N_D$  in  $n$ -type substrates), respectively. Then,  $\phi_{surf}$  at the onset of inversion  $= 2\phi_B$ . 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

$$\phi_{surf}(inv) = 2\phi_B = qN_A (d_{max})^2 / 2\kappa_{si}\epsilon_0 \quad (3 - 66)$$

Solving this for  $d_{max}$ , and substituting into Eq. 3-65 we get  $Q_B$  in terms of  $N_A$  and  $\phi_B$ . Finally, using this expression for  $Q_B$  in Eqs. 3-19 and 3-20 ( $\phi_{ox} = -Q_B/C_{ox}$ ), 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  $\epsilon_{ox}$ ,  $T$ ,  $n_i$ , and  $N_A$ :

$$V_T \text{ (of an ideal MOS capacitor)} = 2\sqrt{\kappa_{si}\epsilon_0 q N_A \phi_B / C_{ox}} + 2\phi_B \quad (3 - 67)$$

### 3.5 PHYSICS OF NON-IDEAL MOS CAPACITORS ( $n^+$ -PolySi-SiO<sub>2</sub>-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<sub>2</sub>-Si structure. In the metal-SiO<sub>2</sub>-Si structure (and other "real" MOS capacitors), the work function difference  $\phi_{ms}$  is virtually never zero, and various charges are also always present inside the oxide or at the Si-SiO<sub>2</sub> 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 ( $q\phi_m'$  or  $q\phi_s'$ , respectively), as shown in Fig. 3-12. Note also that  $q\phi_s' = q[\chi' + (E_c - E_F)_\infty] = q(\chi' + E_G/2 + \phi_B)$  for a *p*-type semiconductor, where  $\chi'$  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<sub>2</sub> case,  $q\chi' = 3.23$  eV.

Figure 3-12a shows the energy-band diagram for an Al-SiO<sub>2</sub>-*p*-type-Si structure before contact. For Al-SiO<sub>2</sub>,  $q\phi_m' = 3.2$  eV, while for *p*-type Si-SiO<sub>2</sub>  $q\phi_s' = (3.23 + 0.55 + q\phi_B) \sim 4.1$  eV. Obviously,  $q\phi_m'$  and  $q\phi_s'$  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.

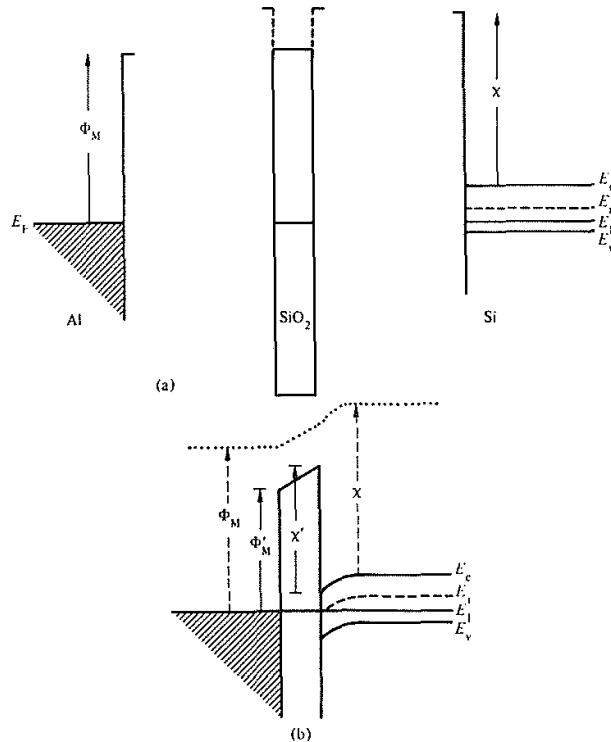


Fig. 3-12 The MOS capacitor structure and idealized energy-band diagram (a) before contact and (b) after contact.<sup>12</sup> Reprinted with permission of Addison-Wesley.

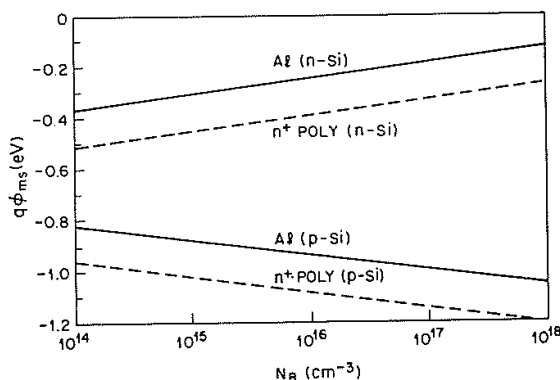


Fig. 3-13 Work-function difference  $\phi_{ms}$  (V) versus doping for  $n^+$  and  $p^+$  polysilicon, and Al gate electrodes.<sup>7</sup> Reprinted with permission of Solid-State Electronics.

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  $(\phi_m - \phi_s)$ ], would have to be applied to the gate (in this example  $\approx -0.9$  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  $\phi_{ms}(n\text{-ch}) = (\phi_m - \phi_s) = (-E_G/2q) - \phi_B$  (with its value between  $-0.9$  to  $-0.95$  V for  $10^{16} \text{ cm}^{-3} < N_A < 10^{17} \text{ cm}^{-3}$  - the appropriate doping range for submicron technologies if  $L > 0.5 \mu\text{m}$ ). Likewise, the  $p$ -channel  $\phi_{ms}$  for an  $n^+$  poly on an  $n$ -type substrate is  $\phi_{ms}(p\text{-ch}) = (\phi_m - \phi_s) = (-E_G/2q) + |\phi_B|$ , with its value between  $-0.15$  to  $-0.2$  V for the same doping range. Finally, for a  $p$ -channel device having a  $p^+$  poly on an  $n$ -substrate,  $\phi_{ms}(p\text{-ch}) = (\phi_m - \phi_s) = (E_G/2q) + |\phi_B|$ . Figure 3-13 shows the value of  $\phi_{ms}$  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<sub>2</sub> interface as shown schematically in Fig. 3-14. The effective net (i.e., total) charge per unit area at the Si-SiO<sub>2</sub> interface (in C/cm<sup>2</sup>) is denoted by the symbol  $Q_{\text{tot}}$ , whereas the net (total) number of charges per unit area ( $Q_{\text{tot}}/q$ ) is given by  $N_{\text{tot}}$ . 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<sup>2</sup>), and  $N_m$  (no. of mobile charges/cm<sup>2</sup>);

2. Fixed oxide charge per unit area  $Q_f$  (C/cm<sup>2</sup>), and  $N_f$  (no. of fixed oxide charges/cm<sup>2</sup>);
3. Interface trap charge per unit area  $Q_{it}$  (C/cm<sup>2</sup>), and  $N_{it}$  (no. of interface trapped charges/cm<sup>2</sup>);  $D_{it}$  - number of interface trapped charges per unit area *and* energy (no. of interface trapped charges/cm<sup>2</sup>-eV);
4. Oxide trapped charge  $Q_{ot}$  (C/cm<sup>2</sup>), and  $N_{ot}$  (no. of oxide trapped charges/cm<sup>2</sup>).

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<sup>-2</sup>. 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.

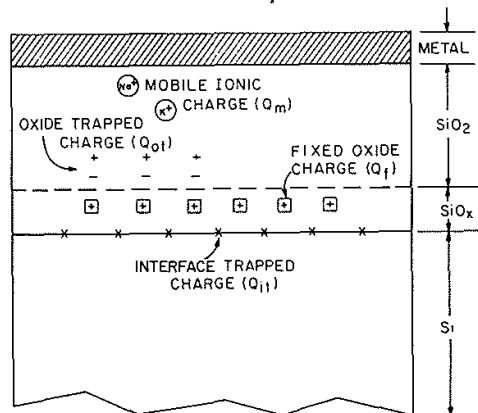


Fig. 3-14 Standardized terminology for oxide charges associated with thermally oxidized silicon.<sup>13</sup> (© IEEE 1980).

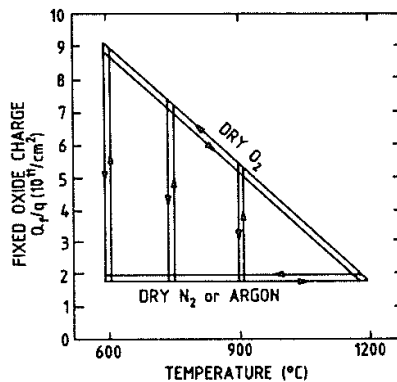


Fig. 3-15 The Deal triangle, which specifies the expected  $Q_f/q$  after dry  $O_2$  oxidation and after inert ambient annealing.<sup>9</sup> Reprinted by permission of the publisher, The Electrochemical Society, Inc.

**3.5.2.2 Fixed Oxide Charge.** The fixed oxide charge  $Q_f$  (usually positive) is located in the oxide within approximately  $30\text{\AA}$  of the Si-SiO<sub>2</sub> 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<sub>2</sub>. 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<sub>2</sub> 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  $\sim 2 \times 10^{11} \text{ cm}^{-2}$ . The time required to reach a minimum value of  $N_f$  is less than 5 min at  $1200^\circ\text{C}$  and about an hour at  $550^\circ\text{C}$ . 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  $1200^\circ\text{C}$ .

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  $\sim 1.1 \times 10^{10} \text{ cm}^{-2}$ ). As the lowest values of  $N_f$  are desired, (100) material is used to build MOS ICs.

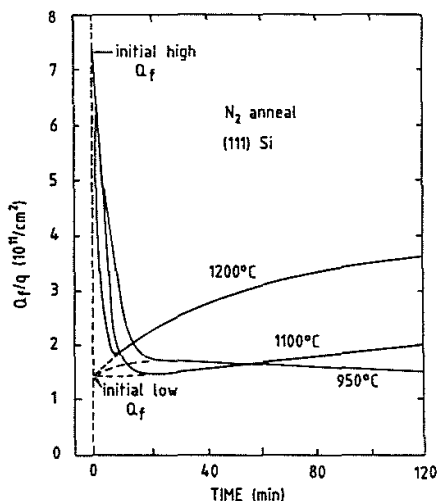


Fig. 3-16 Variation of fixed oxide charge as a function of time during nitrogen anneal at 950, 1100 or 1200°C.<sup>10</sup> Reprinted by permission of the publisher, The Electrochemical Society, Inc.

Since  $Q_f$  is located at the Si-SiO<sub>2</sub> interface, the change in  $V_T$  that it causes ( $\Delta V_T$ ) is simply calculated as the shift in  $V_T$  that would be produced by a sheet of positive charge placed at the interface, or

$$\Delta V_T = -Q_f/C_{OX} = q N_f/C_{OX} \quad (3-68)$$

**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<sub>2</sub> 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<sub>2</sub> 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).<sup>11</sup> 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

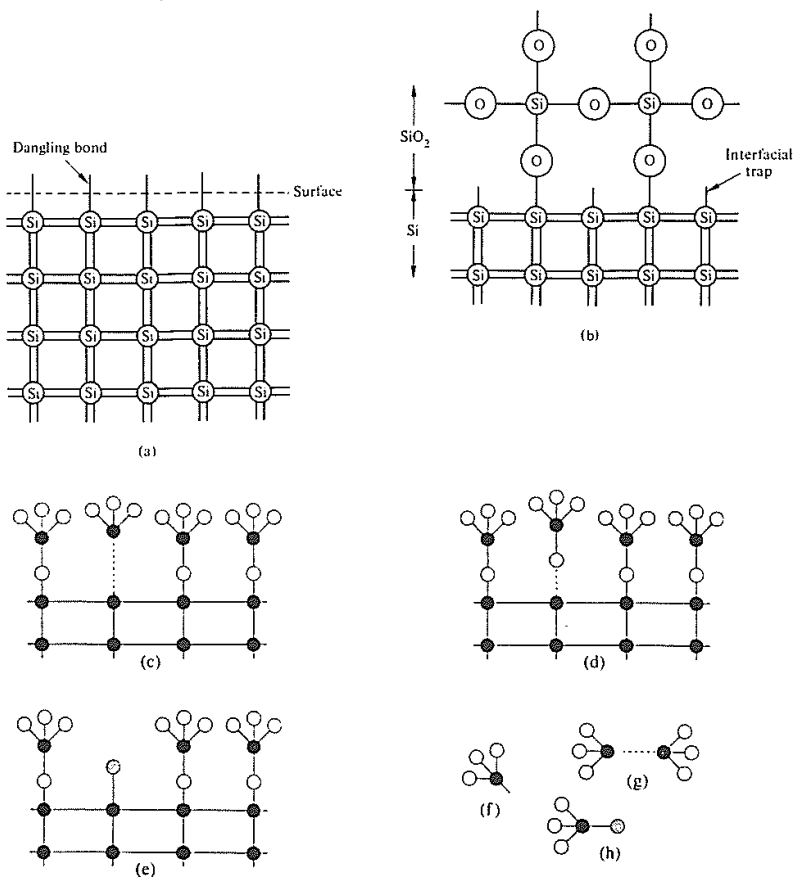


**Fig. 3-17** Physical model for the interfacial traps. (a) "Dangling bonds" which occur when the Si lattice is abruptly terminated along a given plane to form a surface. (b) Post oxidation dangling bonds (relative number greatly exaggerated) that become the interfacial traps. (c) Si-Si bond stretching. (d) Si-O bond stretching. (e) presence of a metallic impurity at the Si surface. (f) - (h) possible defects in the  $\text{SiO}_2$  bulk that may give rise to oxide trapped charge.

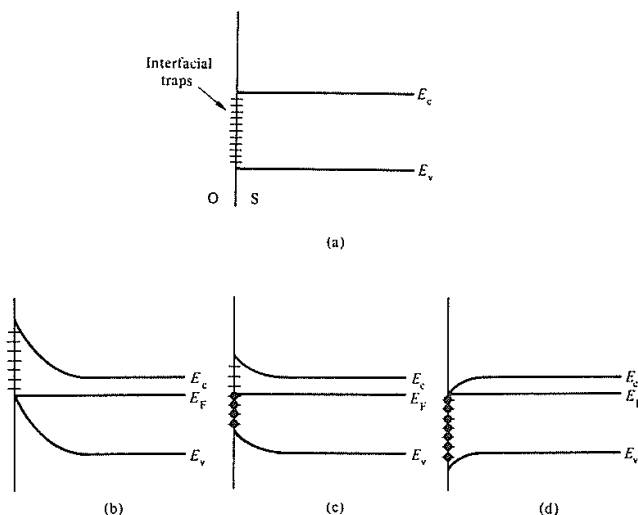


**Fig. 3-18** (a) Electrical modeling of interfacial traps as allowed electronic levels localized in space at the oxide-semiconductor interface. (b) - (d) Filling of the interface levels under (b) inversion, (c) depletion, and (d) accumulation biasing in an  $n$ -type device. From R.F. Pierret, *Modular 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<sub>2</sub> interface, the change in  $V_T$  due to  $Q_{it}$  can be simply calculated from

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

Since Eq. 3-69 shows that  $Q_{it}$  depends upon the bias condition (and is therefore a function of  $\phi_{surf}$ ), 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  $\sim 3:1$ . Second,

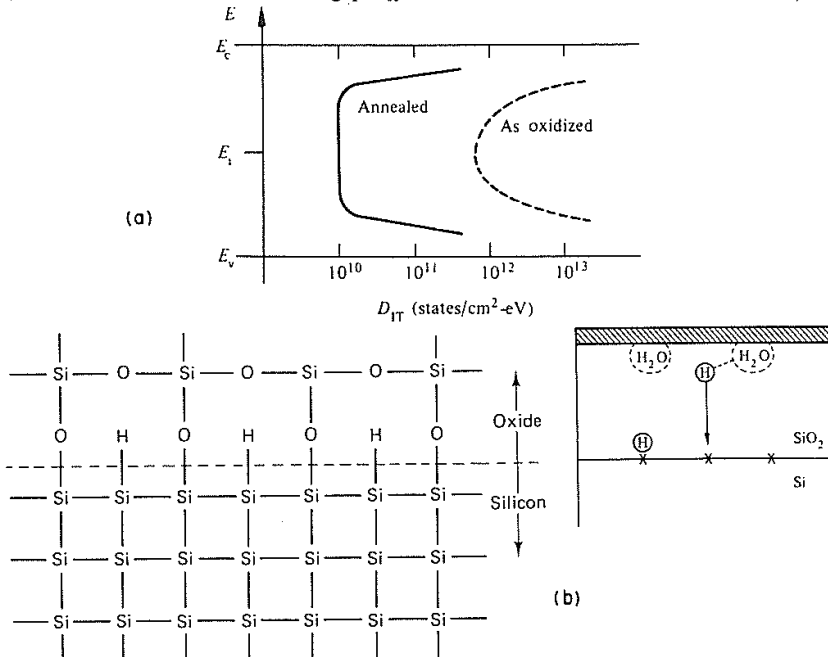


Fig. 3-19 (a) Energy distribution of interface states within the band gap. General form and magnitude of the interfacial trap density observed before and after interface state anneal. (b) Conceptual picture of how hydrogen is thought to passivate dangling bonds.

after oxidation in a dry  $O_2$  ambient, the midgap value of  $D_{it}$  is relatively high ( $\sim 10^{11}$ - $10^{12}$  states/cm<sup>2</sup>-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 ( $\leq 500^\circ\text{C}$ ) for 5-10 minutes minimizes  $D_{it}$ .  $D_{it}$  at midgap under an optimum set of processing conditions is  $\sim 10^{10}$ /cm<sup>2</sup>-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 ( $\cdot Si \equiv O_3$ ), and stretched Si-O or Si-Si bonds (in the latter a missing bridging oxygen atom leads to a defect,  $O_3 \equiv Si \cdot \cdot Si \equiv O_3$ ) 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 ( $450$ - $550^\circ\text{C}$ ) 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,

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

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

$$\begin{aligned}
\Delta V_T &= Q_{tot}/C_{ox} = qN_{tot} t_{ox}/3.9 \epsilon_0 \\
&= (1.6 \times 10^{-19} \text{ C})(5 \times 10^{10} \text{ cm}^{-2})(2 \times 10^{-6} \text{ cm}) / (3.5 \times 10^{-13} \text{ F/cm}) \\
&= 0.05 \text{ V}
\end{aligned}$$

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<sub>2</sub> 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  $\phi_{ms} \neq 0$  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

$$V_{FB} = \phi_{ms} - Q_{tot}/C_{ox} \quad (3-71)$$

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  $V_{SB} = 0V$ ) is given as

$$\begin{aligned}
V_T \text{ (uniform doping, long-channel)} &= V_{T(ud,lc)} = V_{FB} + 2\sqrt{\kappa_{Si}\epsilon_0 q N_{sub}\phi_B}/C_{ox} + 2\phi_B \\
&= \phi_{ms} - Q_{tot}/C_{ox} + 2\sqrt{\kappa_{Si}\epsilon_0 q N_{sub}\phi_B}/C_{ox} + 2\phi_B \quad (3-72)
\end{aligned}$$

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<sub>2</sub> interface. Their presence is accounted for by adding a term to the  $V_T$  equation similar to the term used to describe

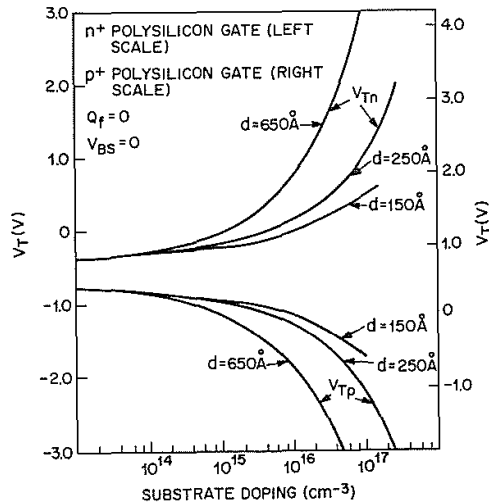


**Fig. 3-20** Calculated threshold voltages of  $n$ -channel ( $V_{Tn}$ ) and  $p$ -channel ( $V_{Tp}$ ) transistors as a function of their substrate's doping, assuming  $n^+$  polysilicon gate (left scale) and  $p^+$  polysilicon gate (right scale). Curves for gate oxide thicknesses  $d$  of 150 Å, 250 Å, and 650 Å 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 ( $\phi_{ms} = 0V$ ,  $Q_{tot} = 0$ ) with a uniform substrate doping  $N_A = 10^{15}/cm^2$ . The gate-oxide thickness is  $t_{ox} = 15$  nm.

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

$$\epsilon_{ox} = 3.9\epsilon_0 = 3.5 \times 10^{-13} \text{ F/cm}; \quad C_{ox} = \epsilon_{ox}/1.5 \times 10^{-6} = 2.3 \times 10^{-7} \text{ F/cm}^2$$

$$2\sqrt{\kappa_{si}\epsilon_0 q N_A \phi_B} = 2 [1.02 \times 10^{-12} \times 1.6 \times 10^{-19} \times 10^{15} \times 0.29]^{1/2} = 1.4 \times 10^{-8} \text{ C/cm}^2$$

$$2\sqrt{\kappa_{si}\epsilon_0 q N_A \phi_B} / C_{ox} = 1.4 \times 10^{-8} / 2.3 \times 10^{-7} = 0.06 \text{ V}$$

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

**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  $Q_{tot} = q(5 \times 10^{10}/cm^2)$ .

**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  $\phi_{ms} + Q_{tot}/C_{ox}$ . Thus, since

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

and

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

$$\begin{aligned} V_{T(ud,lc)} &= V_{T(ud,lc, ideal)} + V_{FB} \\ &= 0.64 \text{ V} + (-1.03 \text{ V}) = -0.39 \text{ V} \end{aligned}$$

Note that a negative value of  $V_{T(ud,lc)}$  is yielded, implying that this MOSFET would be ON at  $V_{GB} = 0 \text{ V}$  (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  $\sim 1.6 \times 10^{12}$  boron ions/cm<sup>2</sup> 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  $\phi(x)$  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  $\phi_{surf} = \phi(x=0)$  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),  $\phi(x,y) = 0$  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  $\phi_{surf} = 2\phi_B = 0.58V$ , with the variation of  $\phi(x)$  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  $\phi_{surf} = 2\phi_B = 0.58V$ , which corresponds to the onset of strong inversion. Thus, when  $V_{GB} = 0V$ ,  $\phi_{surf} = 0.97V$ , which is larger than  $2\phi_B$ , 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  $1.6 \times 10^{12}$  boron ions/cm<sup>2</sup> must be implanted into the channel region. If this is done, when  $V_{GB} = 0V$ ,  $\phi_{surf}$  will then equal  $-0.05V$ , and the potential versus position diagram will as shown in Fig. 3-21d1. Note that when  $V_{GB} = 0.7V$ ,  $\phi_{surf} = 0.58V$ , 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," *IEEE 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<sub>2</sub> 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).

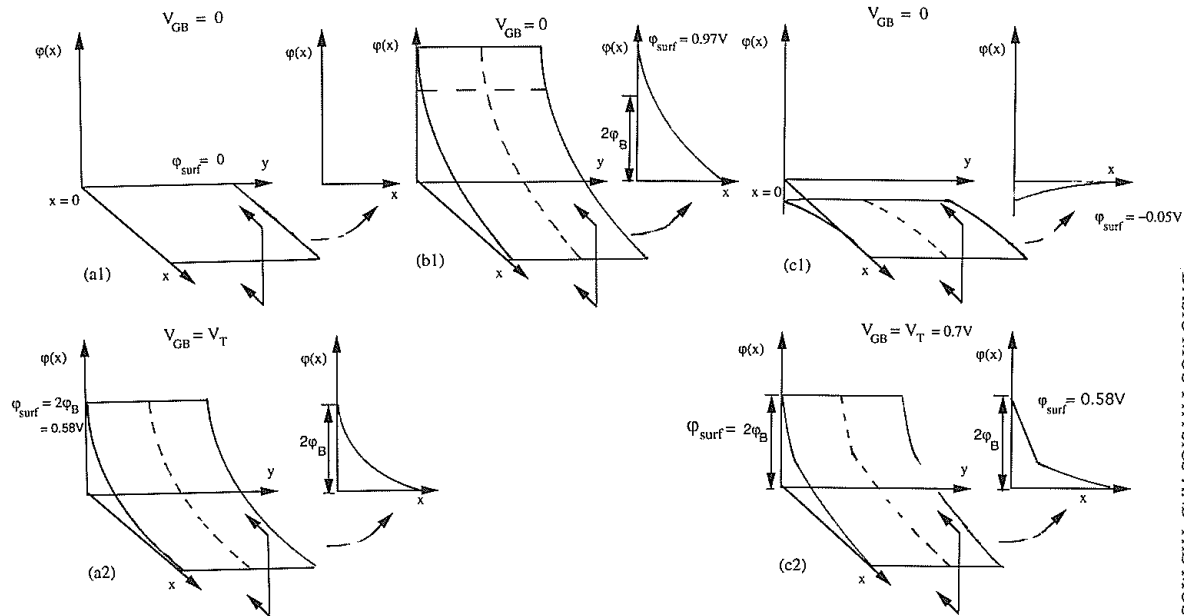


**Fig. 3-21** Visualization of the electrostatic potential in  $p$ -type MOS-Cs with  $N_A = 10^{15} \text{ cm}^{-3}$ , and  $t_{ox} = 150 \text{ \AA}$ . (a1)  $\phi(x,y)$  in an *ideal* MOS-C ( $\phi_{ms} = 0$  and  $Q_{tot} = 0$ ) when  $V_{GB} = 0V$  (flatband condition); and (a2) when  $V_{GB} =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

## 132 SILICON PROCESSING FOR THE VLSI ERA - VOLUME III

11. P. Roblin, S. Samman, and S. Bibyk, "Simulation of Hot-Electron Trapping and Aging in nMOSFETs," *IEEE Trans. Electron Dev.*, ED-35, December 1988, p. 2229.
12. R.F. Pierret, *Modular Series on Solid-State Devices; Vol. IV - Field Effect Devices.*, R.F. Pierret and G.W. Neudeck, Eds., Addison-Wesley Publ. Co., Reading MA, 1983.
13. B.E. Deal, "Standardized Terminology for Oxide Charges Associated with Thermally Oxidized Silicon," *IEEE Trans. Electron Dev.*, ED-27, p. 606, (1980).

## PROBLEMS

1. Explain why it is useful to study the behavior of the MOS-C prior to studying the behavior of the MOSFET.
2. Sketch the energy-band diagrams of a MOS-C with a heavily-doped  $n$ -polysilicon gate when (a)  $V_{GB} = 0$ , and (b) in the flat band condition when the Si substrate: (1)  $1\text{-}\Omega\text{-cm}$   $n$ -type, and (2)  $1\text{-}\Omega\text{-cm}$ ,  $p$ -type.
3. For an ideal MOS-C with  $t_{OX} = 30\text{ nm}$  and  $N_A = 5 \times 10^{15}\text{ cm}^{-3}$ , find the applied voltage and electric field at the interface needed to: (a) make the silicon surface intrinsic; and (b) to bring about strong inversion.
4. Using Eq. 3-13c prove that for  $n_{surf} = 10N_A$ ,  $\phi_{surf}$  is only 58 mV greater than  $2\phi_B$ .
5. Calculate the flatband voltage and the threshold voltage (a) in  $1\text{ }\Omega\text{-cm}$   $p$ -type silicon and (b) in  $1\text{ }\Omega\text{-cm}$   $p$ -type silicon. The MOS-Cs for each case have: (i) aluminum gate for which  $q\phi_m = 4.3\text{ eV}$ , (ii)  $t_{OX} = 100\text{ nm}$ ; and (iii) the oxide is free of charge except for a surface density of  $Q_f = 5 \times 10^{10}\text{ cm}^{-2}$ . The electron affinity of silicon is 4.05 eV.
6. Use the "threshold" statement in SUPREM III to extract and plot  $V_T$  vs.  $V_{SB}$  for increments of 0.25V between 0 and 5V for the device specified in Prob. 4. Compare the zero bias ( $V_{BS} = 0V$ ) value with the two corresponding values of  $V_T$  found in Prob. 4.
7. Consider Poisson's equation in the substrate of a MOS-C. Using the condition that there is charge neutrality in the substrate, show that the impurity doping terms  $N_D - N_A$  can be written as  $n_i (\exp(u_B) - \exp(-u_B)) = 2 n_i \sinh u_B$ , where  $u$  represents the potential normalized to  $(kT/q)$  and  $u_B$  is the substrate potential ( $u_B = -q\phi_B/kT$ ) in a  $p$ -type substrate.

8. Using the form derived in Problem 6, show that the Poisson equation can be written in terms of  $u$  as:

$$d^2u/dx^2 = (1/L_{Di})^2 [\sinh(u) - \sinh(u_B)]$$

where  $L_{Di}$  is the intrinsic Debye length

$$L_{Di} = (\epsilon_{Si} k T / 2q^2 n_i)^{1/2}$$

9. Calculate the maximum depletion layer depth in a MOS-C with a  $p$ -type substrate doped uniformly at  $2 \times 10^{15}\text{ cm}^{-3}$  if  $T = 77K$  and  $300K$ .

10. Consider the establishment of a  $p$ -type inversion layer in a MOS-C. Show that for a nondegenerate  $n$ -type semiconductor, the surface charge density  $Q_C$  needed to make  $p_{\text{surf}} = n_0$  (the equilibrium electron concentration in the bulk) is given by

$$Q_C = \sqrt{2\epsilon_{\text{Si}} q N_D (\phi_B)}.$$

# CHAPTER 4

## MOS TRANSISTOR DEVICE PHYSICS: PART 2

### LONG-CHANNEL MOSFETS

A description of the physics of the MOS capacitor (MOS-C, chap. 3) provides a solid foundation for analyzing the behavior of the MOSFET. Hence, in this chapter the device physics of long-channel MOSFETs is undertaken. In chap. 5 we examine the behavior of short-channel (i.e., submicron) MOSFETs.

Long-channel MOSFETs are discussed first for the following reasons:

1. The derivation of accurate long-channel-MOSFET drain-current models can be achieved by solving both the *Poisson equation* and the *drift-diffusion current density equations* in one-dimensional form, whereas accurate modeling of short-channel MOSFET behavior requires the solution of these equations in two (or even three) dimensions. Thus, it is logical to undertake the simpler analysis first.
2. The long-channel drain-current MOSFET models derived from the solution of the 1-D device equations can be expressed in analytical form. Hence, these models are widely used in circuit-design applications. In addition, quantitative insight about the device physics of MOSFET operation can also be obtained by examining these models. On the other hand, physically based analytical models of short-channel MOSFET behavior do not arise from the solution of the two-dimensional device equations, and thus it is harder to gain quantitative insight into MOSFET operation from them.
3. The analytical expressions for  $I_D$  from two of the four long-channel MOSFET models derived in this chapter are the  $I_D$  expressions used in SPICE for simulating MOSFET behavior in IC circuit design applications.
4. As MOSFETs are scaled down, they eventually become prone to 'short-channel' effects. In most cases, however, these effects are considered detrimental. That is, it is the task of the device designer to suppress them when scaling-down MOSFET structures. The goal is to produce 'short-channel' MOSFETs with 'long-channel' behavior. Thus, it is important to define long-channel behavior and to describe the physics that produce such characteristics in MOSFETs. This understanding provides insight into the measures that must be employed to suppress unwanted short-channel effects.

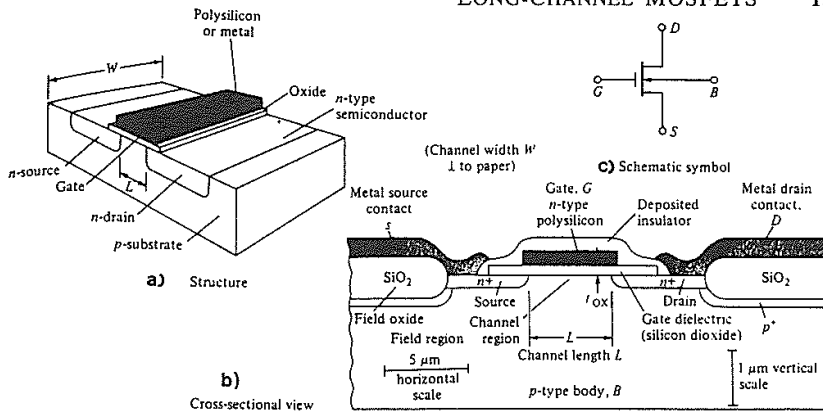


**Fig. 4-1** (a) Structure of an MOS device. (b) Cross sectional view. (c) Schematic symbol.<sup>1</sup> From D. A. Hodges and H. G. Jackson, Analysis and Design of Digital Integrated Circuits, Copyright, 1983 McGraw-Hill Book Co. Reprinted with permission.

In this chapter the following topics dealing with MOSFETs are covered:

- a. Introduction to the terminology of MOSFETs and their behavior
- b. Effect of Applying a Bias to the Inversion Layer of the MOS-C
- c. Definition of the Long-Channel MOSFET and the Gradual-Channel Approximation
- d. Qualitative Description of Long-Channel MOSFETs
- e. Quantitative Models of dc Long-Channel MOSFET Behavior
  - i. Pao-Sah Model
  - ii. Charge-Sheet Model
  - iii. Variable-Depletion-Charge Model
  - iv. Square-Law Model
- f. Threshold-Voltage Control in Long-Channel MOSFETs
  - i. Body Effect
  - ii. Threshold-voltage adjustment by ion-implantation
- g. Subthreshold behavior of Long-Channel MOSFETs
  - i. Physics and modeling of subthreshold current flow
  - ii. Subthreshold swing
  - iii. Gate-induced leakage current (GIDL)

#### 4.1 TERMINOLOGY OF MOSFET DEVICE STRUCTURES

Figure 4-1 shows the cross-section of the MOSFET, as introduced in chap. 3. The important difference between the MOS-C and the MOSFET is that in the latter a  $pn$

junction exists at each side of the MOS-C channel region. Each of these junctions are also close enough to make contact with the inversion layer. By making external connections to each of these edge junctions, the MOSFET becomes a four-terminal device (with the terminals designated respectively as the *gate* G, *body* B, *source* S, and *drain* D). The MOS-C, on the other hand, remains a two terminal device (*gate* and *body*). The source and drain terminals of the MOSFET allow a bias  $V_{DS}$  to be applied across the ends of the channel region, giving rise to a *drain current*  $I_D$ .

As described in chap. 3, the inversion layer can be established by applying an appropriate bias to the gate. For the device shown in Fig. 4-1a, the induced charge in the channel inversion layer is *n*-type, and the device is known as an *n*-channel MOSFET (or simply an NMOS transistor). A *p*-channel MOSFET (or PMOS) is obtained by interchanging the *n* and *p* regions. Just as in the ideal MOS-C it is assumed that current in the gate oxide (*gate current*,  $I_G$ ) is negligible. We also assume that the MOSFET is a symmetrical device, insofar as connecting the two junction terminals to a circuit (i.e., the drain current should remain the same if the two terminals are interchanged). Regardless of the way a MOSFET is connected, the terminal tied to the more positive voltage (in NMOS circuits) is defined to be the *drain*, and the other is the *source*.

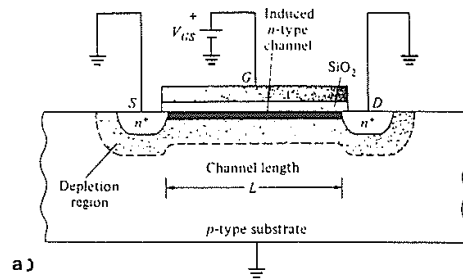
#### 4.1.1 Basics of MOSFET Operation

In simplest terms, the operation of a MOS transistor involves the application of an input voltage to the gate electrode. This establishes an electric field perpendicular (or transverse) to the Si-SiO<sub>2</sub> interface in the channel region of the device (Fig. 4-2a). The conductance of the channel region (in which the device output current  $I_D$  flows) can be modulated by varying this electric field. Since an electric field is responsible for controlling the output current flow, such devices are termed *field-effect transistors* (FETs).

If no gate bias is applied, however, the circuit path between source and drain consists of two back-to-back *pn* junctions in series. In this case, if  $V_{DS}$  is applied (such that the drain and source *pn* junctions are reverse biased),  $I_D$  will consist of only the reverse-bias diode leakage current, which is normally negligible.

When positive bias is applied to an NMOS transistor gate, however, electrons will be attracted to the channel region and holes will be repelled (as is the case in the MOS-C). Once the positive gate voltage becomes strong enough to form an inversion layer, an *n*-type channel is formed that connects the source and drain regions (Fig. 4-2a). A drain current  $I_D$  can then flow if a voltage  $V_{DS}$  is applied between the source and drain terminals. In the simplest analysis we assume that the voltage-induced *n*-type channel does not form unless the voltage applied to the gate exceeds the threshold voltage  $V_T$ .

MOS devices such as those just described (in which no conducting channel exists when  $V_{GS} = 0$ ), are referred to as *enhancement-mode* (or *normally OFF*) transistors (see Fig. 4-2b). With NMOS enhancement-mode transistors, a positive gate voltage,  $V_{GS}$ , greater than  $V_T$  must be applied to create the channel (or to turn them ON), while to turn ON PMOS enhancement-mode devices, a negative gate voltage (whose magnitude is  $>V_T$ ) must be applied. Note that in NMOS transistors a positive voltage must also



| TYPE                                 | CROSS SECTION | OUTPUT CHARACTERISTICS | TRANSFER CHARACTERISTICS |
|--------------------------------------|---------------|------------------------|--------------------------|
| n-CHANNEL ENHANCEMENT (NORMALLY OFF) |               |                        |                          |
| n-CHANNEL DEPLETION (NORMALLY ON)    |               |                        |                          |
| p-CHANNEL ENHANCEMENT (NORMALLY OFF) |               |                        |                          |
| p-CHANNEL DEPLETION (NORMALLY ON)    |               |                        |                          |

b)

Fig. 4-2 (a) Idealized NMOS cross section with positive  $V_{GS}$  applied showing depletion regions and induced channel. From D. A. Hodges and H. G. Jackson, Analysis and Design of Digital Integrated Circuits, Copyright, 1983 McGraw-Hill Book Co. Reprinted with permission. (b) Cross sections and output and transfer characteristics of four types of MOSFETs. From S.M.Sze, Semiconductor Devices - Physics and Technology, 1985, John Wiley & Sons. Reprinted with permission.

be applied to the drain to keep the drain-substrate reverse-biased, while in PMOS devices this voltage must be negative.

On the other hand, it is also possible to build MOS devices in which a conducting channel region exists when  $V_{GS} = 0V$  (see also Fig. 4-2b), and such MOS devices are described as being *normally ON*. Since a bias voltage to the gate electrode is needed to deplete the channel region of majority carriers, and thus turn them OFF, such devices are also commonly called *depletion-mode MOSFETs*. NMOS depletion-mode devices require a negative gate voltage to be turned OFF, while corresponding PMOS devices require positive gate voltages.

## 4.2 BIASING THE INVERSION LAYER (BODY EFFECT)

Before undertaking a detailed discussion of MOSFET behavior, we first study the consequences of *applying a dc bias to the inversion layer of a MOS-structure*. This topic is actually an extension of the subject of MOS-C behavior, but also serves as an important prelude to the discussion of MOSFET operation. Consequently, when we discuss the operation of the MOSFET we will refer back to the results developed in this section.

We examine the effect of biasing the inversion layer from both the qualitative and quantitative perspectives. In the latter we will show how the expressions for the mobile charge concentrations in the inversion layer of the MOS structure are altered when a such bias is applied, as well as how the application of a bias to the inversion layer modifies the threshold voltage equation derived for the MOS-C.

Biasing the inversion layer is accomplished by the application of a dc voltage between the substrate and a diffused region adjacent to, and in contact with, the inversion layer. This implies that a modified MOS-C structure must be used. Specifically, a diffused region must be added to the edge of the MOS-C channel region (see Fig. 4-3a), and a terminal must be connected to this region (designated by the letter C). The result, as shown, is the *three-terminal MOS device* depicted in Fig. 4-3a.

If we choose the substrate (body or B) terminal as the reference, the voltage applied between terminal C and the substrate (body) is denoted by  $V_{CB}$ , and the voltage between the gate (terminal G) and substrate by  $V_{GB}$  (see Fig. 4-3a).<sup>\*</sup> For simplicity we also assume an ideal *p*-type Si-substrate MOS structure (i.e.,  $\phi_{ms} = 0$  and  $Q_{tot} = 0$ ).

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

If a MOS-C with a *p*-type substrate is biased into inversion by a gate voltage  $V_{GB}$ , a *pn* junction exists between the surface and the bulk of the silicon (as a result of the region near the silicon surface having been inverted to *n*-type). If there is an adjacent *n*-doped region in contact with the *induced n*-type inversion layer, it is possible to apply a bias to the induced junction by applying a bias  $V_{CB}$  between terminal C and B (the substrate)

<sup>\*</sup> As elsewhere, we use the convention that the second letter of the voltage subscript designates the reference terminal.

as shown in Fig. 4-3a. (Note that here we have a MOS structure in which the induced junction is contacted only at one end of the channel.\*) We also assume that the polarity of  $V_{CB}$  is such that the induced  $pn$  junction is either unbiased or under reverse bias (i.e.,  $V_{CB}$  in Fig. 4-3a is  $\geq 0V$ ). We will refer to  $V_{CB}$  as the *channel voltage*.

Before presenting the equations relevant to this structure, let us first consider the qualitative effect of applying various  $V_{CB}$  bias values.

**4.2.1.1 Case When Both  $V_{GB}$  and  $V_{CB} = 0V$ .** In Fig. 4-3b we show the two-dimensional potential  $\phi$  versus position diagram for the three-terminal MOS structure when both  $V_{GB}$  and  $V_{CB} = 0V$ . We note that the channel region is in a flat-band condition and that a built-in voltage  $\phi_{bi}$  exists across the  $n^+p$  (doped) junction. We also recall that even if  $V_{GB} \neq 0V$ , as long as  $V_{CB}$  remains at  $0V$ , the structure is still in equilibrium, because  $I_G$  is assumed to be zero and there is no current in the  $n^+p$  junction if  $V_{CB} = 0V$ . Hence, this implies that when  $V_{CB} = 0V$  the channel region of this structure exhibits the same behavior as that of the MOS-C described in chap. 3.

**4.2.1.2 Case When  $V_{GB} \geq V_T$  and  $V_{CB} = 0V$ .** If we now consider the case where  $V_{CB} = 0V$ , but  $V_{GB} \geq V_T$ , we encounter a situation as shown in Fig. 4-3c. That is, the channel region is now inverted, and a *field-induced junction* now exists between the  $n$ -type inversion layer and the underlying  $p$ -type substrate. Since a bias is not applied across the field-induced junction, it is in equilibrium and is again characterized by the same Fermi level as in the diffused junction. The surface depletion region in the field-induced junction under these conditions is also described by the equilibrium theory presented in chap. 3.

Note, however, that the potential difference between the inversion layer and the  $n^+$  region is now much smaller than in the case when  $V_{GB} = 0V$ . Specifically, this potential difference is now  $\sim(\phi_{bi} - 2\phi_B)$ . This implies that the inversion layer is in "communication" with the  $n^+$  region adjacent to it. The term "communication" means that once the inversion layer is formed and a bias  $V_{CB}$  is applied between terminals C and B, this bias will appear across the entire length of the field-induced junction (i.e., no longer just across the doped  $n^+p$  junction at the edge of the channel). Thus, *if an inversion layer exists in the channel, then  $V_{CB}$  does have an influence on  $\phi_{surf}$ .*

For example, if  $V_{CB}$  is suddenly allowed to increase from  $0V$  to some non-zero value (while  $V_{GB}$  is held fixed), electrons will be observed to flow out of the inversion layer through the C terminal. Why this happens, and what the effects are on the behavior of the MOS structure are described next.

**4.2.1.3 Case When  $V_{GB} = 0V$ , But  $V_{CB} > 0V$ .** If we next apply a non-zero  $V_{CB}$  bias while keeping  $V_{GB} = 0V$ , equilibrium will cease to exist in the three-terminal structure. This is because  $V_{CB}$  causes a small reverse-bias leakage current to flow in the  $n^+p$  junction. (Although this reverse-bias leakage current is extremely small, its

\* However, in a MOSFET, contact to the inversion layer can be made at both ends of the induced junction since a diffused region exists at each end of the channel. In the MOSFET these two regions are called the source and drain, and the terminals are designated by the abbreviations S and D.

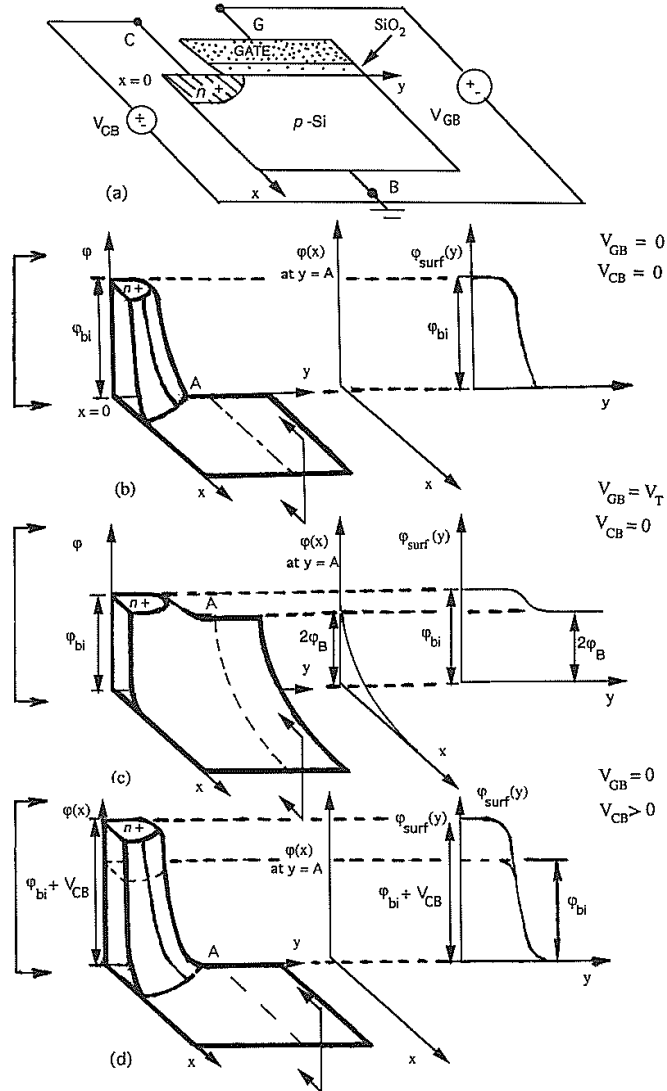


Fig. 4-3 (a) Applying a bias  $V_{CB}$  to an induced channel beneath the gate oxide by applying this bias  $V_{CB}$  between terminal C and the substrate of a three-terminal MOS structure. Two-dimensional potential  $\phi$  versus position diagrams in the three terminal MOS structure when: (b) both  $V_{GB}$  and  $V_{CB} = 0$ ; (c)  $V_{GB} > V_T$  and  $V_{CB} = 0$ ; (d)  $V_{GB} = 0$  and  $V_{CB} > 0$ .

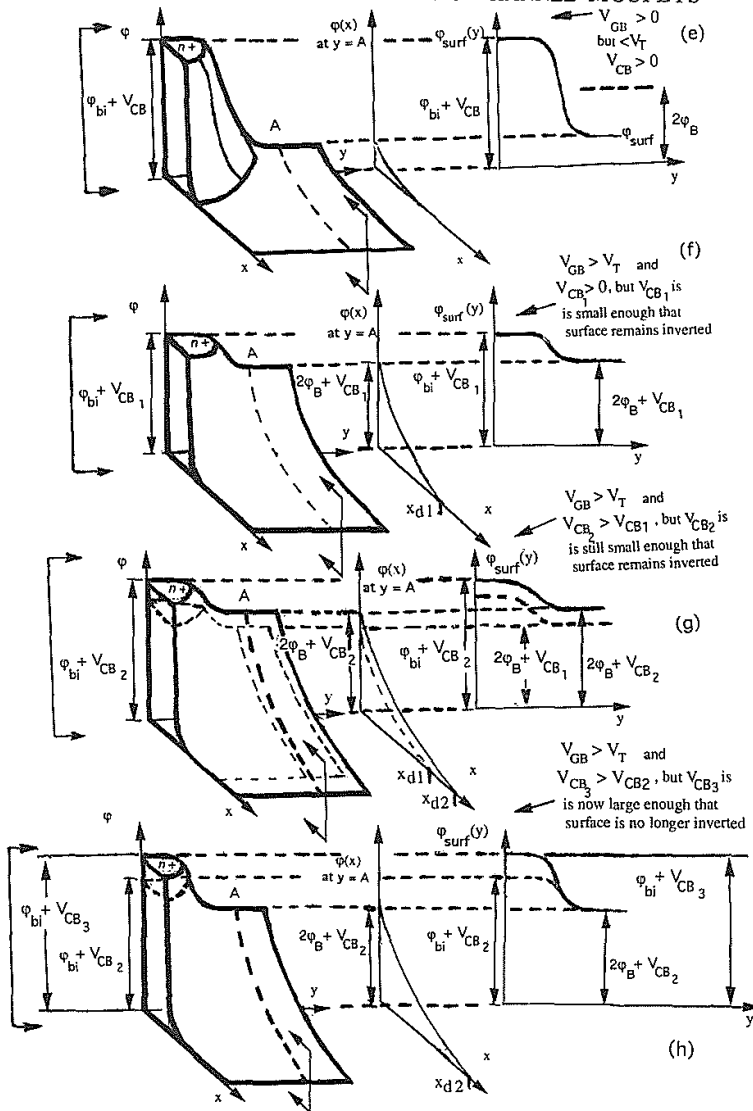


Fig. 4-3 (cont.) (e)  $V_{GB} > 0$ , but is  $< V_T$ , and  $V_{CB} > 0$ ; (f)  $V_{GB} > V_T$  and  $V_{CB} > 0$ , but  $V_{CB} = V_{CB1}$ , which we assume is small enough so that inversion still exists. (g)  $V_{GB}$  is the same as in part (f), but  $V_{CB} = V_{CB2}$ , where  $V_{CB2} > V_{CB1}$ , but  $V_{CB2}$  is still small enough to keep surface inverted. (h)  $V_{GB} = V_T$ ,  $V_{CB} = V_{CB3}$  (large enough to suppress the channel inversion layer).

existence implies that the MOS structure is no longer in equilibrium because some of the carrier energy now depends on an externally applied force, namely the extra electric field created in the depletion region by  $V_{CB}$ .) However, if  $V_{GB}$  remains fixed at 0V,  $V_{CB}$  is dropped entirely across the  $n^+p$  junction. Hence, the surface potential along the channel is not altered from when  $V_{CB} = 0V$ . Instead, only the potential difference between the  $n^+$  and  $p$  (substrate) regions is increased by  $V_{CB}$  (as shown in Fig. 4-3d), leaving the channel region to the right of the  $n^+p$  junction in a flat-band state. In addition, Fig. 4-3d indicates that *if no inversion layer is present, changing  $V_{CB}$  has negligible impact on the surface potential  $\phi_{surf}$  along the channel*. This last statement also applies to the case illustrated in Fig. 4-3e (i.e.,  $V_{GB} > 0V$ , but is less than  $V_T$ , and  $V_{CB}$  is increased from 0V to some value greater than 0V). Here we observe that there is still no impact on  $\phi_{surf}$  as long as no inversion layer is present in the channel.

**4.2.1.4 Case When  $V_{GB} \geq V_T$  and  $V_{CB} > 0V$ .** Let us now consider the case when  $V_{GB} > V_T$  and a reverse-bias voltage  $V_{CB}$  is applied to the induced  $pn$  junction (such that the surface remains inverted even after  $V_{CB}$  is applied). Since the surface is inverted, the potential in the inversion region is very close to the potential value in the diffused region, as shown in Fig. 4-3f (and as exists in the case shown in Fig. 4-3c). That is, a point at the surface is at about the same potential (with respect to the substrate deep in the bulk) as a point in the  $n^+$  region. Therefore, the surface is about as "attractive" for electrons as is the  $n^+$  region. Consequently, electrons have no reason to prefer only the  $n^+$  region as opposed to the "surface" at the right. As a result, the  $n^+$  region can "communicate" with the inversion layer (implying that a bias applied to the  $n^+$  region is also applied to the inversion layer). In addition, just as in the case of a MOS-C, the value of the surface potential  $\phi_{surf}$  remains clamped at the value it has at the onset of strong inversion (i.e., for the case of Fig. 4-3f, at a value of  $\phi_{surf} \sim 2\phi_B + V_{CB1}$ ). That is, as  $V_{GB}$  is increased beyond  $V_T$ ,  $\phi_{surf}$  remains unchanged, but the mobile charge concentration in the inversion layer continues to increase.

We also recall from  $pn$  junction theory that if a reverse-bias voltage is applied to a previously unbiased  $pn$  junction, the junction depletion-region width will increase. Thus, the depletion region on the  $p$ -side of the induced junction will also increase if a reverse-bias is applied between terminals C and B. However, if  $V_{GB}$  is kept fixed (such that  $V_{GS} \geq V_T$ ) as the reverse-bias voltage  $V_{CB}$  is increased (see Fig. 4-3g), the charge on the gate terminal will remain the same as when  $V_{CB} = 0V$ . Therefore, the *total* charge in the semiconductor must also remain unchanged to maintain balance with the unchanged value of charge on the gate. However, since the application of  $V_{CB}$  widens the depletion region (thereby uncovering more charge in this region of the induced junction), fewer electrons are needed in the inversion layer to balance the unchanged gate charge. Consequently, electrons must exit the inversion layer, resulting in a lower level of inversion. (Looking at it in another way, the increased positive potential at terminal C will attract electrons from the inversion layer, which will start flowing toward the  $n^+$  region, and from it into the top terminal of the voltage source.) However, if a sufficiently large value of  $V_{CB}$  is applied (while holding  $V_{GB}$  fixed), this effect can entirely extinguish the charge in the inversion layer. In other words, a gate voltage

capable of causing inversion in the three-terminal MOS structure when  $V_{CB} = 0V$  may no longer be able to sustain inversion if a non-zero value of  $V_{CB}$  is applied.

As also shown in Figs. 4-3f and g, as long as the surface remains in inversion as  $V_{CB}$  is increased, the depletion region under the field induced junction increases (from  $x_{d1}$  in Fig. 4-3f to  $x_{d2}$  in Fig. 4-3g). This means that the bands are bent more, and thus  $\phi_{surf}$  also increases. This supports the statement that as  $V_{CB}$  increases, the  $n^+$  region communicates this change in  $V_{CB}$  to the field-induced junction. However, once a sufficiently large number of electrons depart from the inversion layer so that it essentially disappears, further increases in  $V_{CB}$  no longer influence the channel region, and the depletion region no longer widens (as shown in Fig. 4-3h), and remains fixed at a value dependent on  $V_{CB}$  (e.g., at  $x_{d2}$ , see Figs. 4-3g and h) as expressed by Eq. 4-8.

If a non-zero  $V_{CB}$  is applied and it is desired to re-establish inversion,  $\phi_{surf}$  must be increased to a value larger than that needed when  $V_{CB} = 0$ . Specifically, at the onset of inversion, it is necessary to increase  $V_{GB}$  so that  $q\phi_{surf} \sim q(V_{CB} + 2\phi_B)$ . Note that the effect just described can also be exploited as a technique for raising the threshold voltage of a MOSFET, as will be discussed in section 4.6.2.

**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.** In summary, it is possible to establish "communication" with the channel region (i.e., an external bias voltage can be applied to the channel region of a MOS structure) if an inversion layer has been induced there. On the other hand, if the channel region is not inverted, the application of a bias voltage to a diffused region adjacent to the channel region will have no effect on the surface potential in the channel. Such "communication" (or lack of) has the following impact on the behavior of a MOSFET:

1. If the gate voltage applied to a MOSFET is less than the threshold voltage, the device is said to be operating in the subthreshold regime. From our discussion, this means that if a bias is applied to the drain ( $V_{DS}$ ) of a MOSFET being operated in subthreshold, this voltage will be dropped entirely across the diffused  $n^+p$  junction (i.e., the drain-substrate junction), and it will have no impact on  $\phi_{surf}$ . Hence,  $\phi_{surf}$  will be determined by the gate potential, which is constant everywhere on the gate. Consequently, the potential drop along the  $y$  direction at the surface of the channel will be essentially zero. We will use this information when we discuss subthreshold current flow in MOSFETs in section 4.7.
2. If the channel is inverted, and  $V_{CB}$  is increased (and the channel remains in inversion for the range of applied  $V_{CB}$  values), the depletion region of the field-induced channel increases. If  $V_{GB}$  is held fixed, the number of electrons in the inversion layer will decrease as  $V_{CB}$  is increased. At some value of  $V_{CB}$  the inversion layer charge will entirely disappear. These effects occur when a drain bias  $V_{DS}$  is applied to a MOSFET. That is, for a fixed gate voltage (greater than the threshold voltage value), as  $V_{DS}$  is increased, the depletion region of

will correspondingly decrease, making it less inverted. Thus, in a MOSFET operating in inversion and with  $V_{DS} > 0$ , the depletion region near the drain will be wider than at the source (i.e., assuming the source is the common terminal).

3. "Communication" with the channel also permits the diffused region to become an additional source (or sink) of mobile carriers to the channel (i.e., in addition to the carriers thermally generated in the depletion regions). As noted, two such diffused regions exist in a MOSFET, one which supplies mobile carriers (the source), and the other which removes them (the drain). Specifically, if a drain bias is applied to a MOSFET in inversion, the channel carriers will be attracted to and move into the drain. These are replenished by carriers entering the channel from the source. The resultant flow of charge in the channel constitutes a current in the drain terminal  $I_D$ . Since the channel is not a perfect conductor, this current also gives rise to a potential drop along the direction of the channel.

**4.2.1.6 Body Effect.** Until now we have chosen the substrate to be the reference terminal in the three-terminal MOS structure. However, when MOSFETs are used in circuits, the source terminal S (corresponding here to terminal C), is normally chosen to be the reference. Hence, *body effect* is the term used to describe the increase in the threshold voltage of a MOSFET (with reference to the source) upon the application of a reverse bias to the inversion layer. If, as shown in Fig. 4-4b, we choose terminal C to be the reference terminal (instead of the body terminal), the bias voltages are now applied between the gate and C ( $V_{GC}$ ), and between the body and C ( $V_{BC}$ ). In this case, for a given value of  $V_{BC}$ , a value of  $V_{GC}$  greater than that needed when  $V_{BC} = 0V$  must be applied to the MOS structure to establish strong inversion. In section 4.2.2.2 we will derive an expression that determines the increase in  $V_T$  as a function of  $V_{BC}$ .

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

For the three-terminal MOS structure introduced above, it is possible to develop the quantitative relationships in a general manner for all cases of inversion. However, these can also be simplified somewhat for the more limited (but still most widely used form) of the strong inversion case. Note that in deriving these expressions we also include  $V_{FB}$ , which accounts for the work-function difference and oxide charge that can exist in 'real' MOS structures.

**4.2.2.1 General Inversion Case.** In the general inversion case (which covers weak as well as strong inversion), five equations characterize the phenomena in the channel region (i.e., the Si region in the vicinity of the Si-SiO<sub>2</sub> interface to the right of the  $n^+$  junction in Fig. 4-3a) of three-terminal MOS structure (as they do for the MOS-C). If we invoke the charge-sheet and depletion approximations, four of these equations are identical to those for the expressions derived for the charge-sheet model of the MOS-C, and are repeated here for the readers convenience:

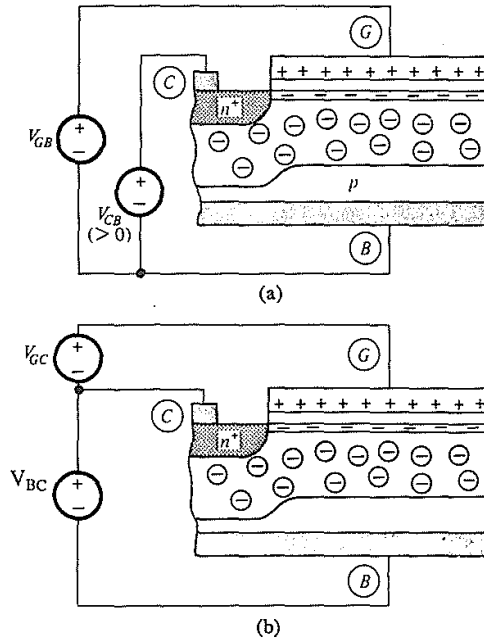


Fig. 4-4 Bias notation when: (a) terminal B is the reference terminal; (b) terminal C as the reference terminal instead of terminal B.<sup>5</sup> Reprinted with permission of McGraw-Hill Co.

$$V_{GB} = \phi_{\text{surf}} + V_{FB} + \phi_{\text{ox}} \quad (3-10) \quad (4-1a)$$

$$Q_C = Q_I + Q_B \quad (3-20) \quad (4-1b)$$

$$Q_C = -C_{\text{ox}} \phi_{\text{ox}} \quad (3-19) \quad (4-1c)$$

$$Q_B = -\gamma C_{\text{ox}} \sqrt{\phi_{\text{surf}}} \quad (3-29) \quad (4-1d1)$$

$$= [2qN_A \kappa_{\text{Si}} \epsilon_0 (\phi_{\text{surf}})]^{1/2} \quad (4-1d2)$$

The fifth equation is a modified form of Eq. 3-53, in which we replace  $2\phi_B$  by  $(2\phi_B + V_{CB})$  to get

$$Q_I = \sqrt{2qN_A \kappa_{\text{Si}} \epsilon_0} \left( \sqrt{\phi_{\text{surf}} + \phi_t e^{(\phi_{\text{surf}} - 2\phi_B - V_{CB})/\phi_t}} / \phi_t - \sqrt{\phi_{\text{surf}}} \right) \quad (4-2)$$

Then, combining Eqs. 4-1a, 4-1c, and 4-1b we get

$$V_{GB} = \phi_{\text{surf}} + V_{FB} - [Q_I(\phi_{\text{surf}}) + Q_B(\phi_{\text{surf}})]/C_{\text{ox}} \quad (4-3)$$

Next, using Eqs. 4-1d1 and 4-2 in Eq. 4-3,  $V_{GB}$  can be written as

$$V_{GB} = \phi_{surf} + V_{FB} + \gamma \sqrt{\phi_{surf} + \phi_t e(\phi_{surf} - 2\phi_B - V_{CB})/\phi_t} \quad (4-4)$$

Now, if  $V_{GB}$  and  $V_{CB}$  are given, Eq. 4-4 can be solved iteratively for  $\phi_{surf}$ . This value can then be inserted into Eq. 4-2 to calculate  $Q_I$ , or into Eq. 4-1d2 to calculate  $Q_B$ .

An expression for  $Q_I$  in terms of  $\phi_{surf}$  and  $V_{GB}$  can also be derived from the above equations. That is, by inserting Eqs. 4-1c and 4-1d1 into Eq. 4-1b above, we get

$$-C_{ox}\phi_{ox} = Q_I + \gamma C_{ox}\sqrt{\phi_{surf}} \quad (4-5)$$

Then, solving for  $\phi_{ox}$ , and substituting the result into Eq. 4-1a, we obtain

$$Q_I = -C_{ox}(V_{GB} - V_{FB} - \phi_{surf} - \gamma\sqrt{\phi_{surf}}) \quad (4-6)$$

**4.2.2.2 Strong Inversion Case.** For the three-terminal structure biased to strong inversion, the above general analysis can be simplified. That is, with  $V_{CB} = 0V$ ,  $\phi_{surf}$  of the three-terminal MOS structure at the onset of inversion is  $2\phi_B$ . With a reverse bias applied to the induced junction, the surface potential at the onset of strong inversion becomes

$$\phi_{surf}(inv) \approx 2\phi_B + V_{CB} \quad (4-7)$$

Since the applied  $V_{CB}$  extends the range of gate voltages under which the surface region remains in depletion,  $d_{max}$  is now larger. The expression for  $d_{max}$  for the case of an applied  $V_{CB}$  is (assuming the depletion approximation):

$$d_{max} = [2\kappa_{Si}\epsilon_0(2\phi_B + V_{CB})/qN_A]^{1/2} \quad (4-8)$$

and the depletion layer charge (per unit area),  $Q_{Bmax}$ , when  $d = d_{max}$  is expressed as

$$Q_{Bmax} = [2q\kappa_{Si}\epsilon_0N_A(2\phi_B + V_{CB})]^{1/2} \quad (4-9)$$

The expression for the inversion layer charge per unit area  $Q_I$  in strong inversion in terms of  $V_{GB}$  and  $V_{CB}$  is

$$Q_I = -C_{ox}[V_{GB} - V_{FB} - 2\phi_B - V_{CB} - \gamma(2\phi_B + V_{CB})^{1/2}] \quad (4-10)$$

Assuming again that  $Q_I$  is negligible at the onset of inversion, the total charge per unit area in the semiconductor is given by  $Q_B$ . Hence, the gate-substrate bias at the onset of strong inversion is

$$V_{GB} = V_{FB} + \phi_{surf} + \phi_{ox} = V_{FB} + (2\phi_B + V_{CB}) + \gamma(2\phi_B + V_{CB})^{1/2} \quad (4-11)$$

Noting that  $V_{GC} = V_{GB} - V_{CB}$ , the threshold voltage, referenced to terminal C, is given by the expression

$$V_T(\text{with ref. to term C}) = V_{T(\text{ud,lc})} = V_{FB} + 2\phi_B + \gamma(2\phi_B + V_{CB})^{1/2} \quad (4-12)$$

Note that the term  $\gamma$ , which was defined in Eq. 3-60a as  $[\sqrt{2qN_A\kappa_{Si}\epsilon_0}]/C_{OX} = [\sqrt{2qN_A\kappa_{Si}\epsilon_0}]t_{OX}/\epsilon_{OX}$  is a factor which determines how much  $V_T$  will increase when  $V_{CB}$  is increased. Hence,  $\gamma$  has been given the name *body effect coefficient*. By inspecting the expression for  $\gamma$  we note that it predicts the body effect is stronger for heavier substrate dopings and/or thicker gate oxides. Conversely, the body effect can be made smaller by using lower doping concentrations and/or thinner gate oxides. Figure 4-5 shows the threshold voltage shift versus substrate bias for various  $\gamma$  values.<sup>26</sup>

### 4.3 DEFINITION OF TERM: LONG-CHANNEL MOSFET

It is appropriate at this point to provide a detailed definition of the term *long-channel MOSFET*. From a qualitative perspective, long-channel MOSFETs are assumed to have channel regions sufficiently long and wide so that "edge" effects along the four sides of the channel can be neglected. Insofar as the channel length is concerned, we assume the channel length  $L$  is much longer than the sum of the source and drain

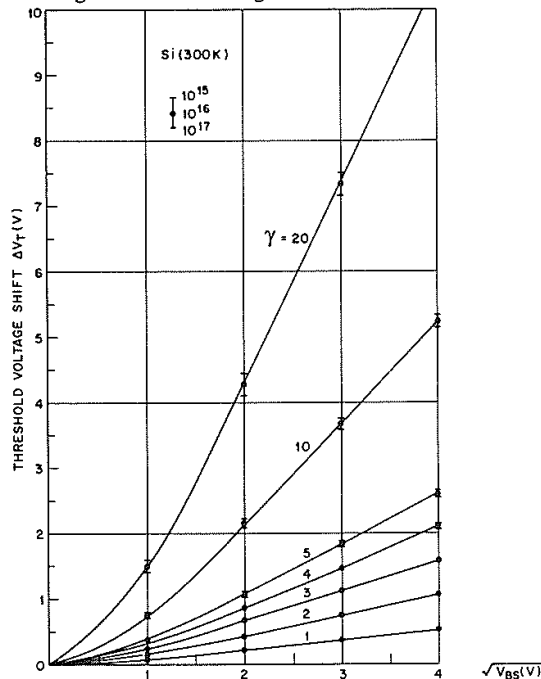


Fig. 4-5 Threshold voltage shift versus substrate reverse bias for various  $\gamma$  values.<sup>2</sup>  
Reprinted with permission of John Wiley & Sons.

depletion-region widths, or  $L \gg (d_{\text{source}} + d_{\text{drain}})$ . The channel width, on the other hand is assumed to be much greater than the gate depletion-region depth,  $d$  (or  $Z \gg d$ ).

Even when the above criteria are well-satisfied, however, it is apparent that the electric fields and potential in the MOSFET still vary in at least two dimensions.\* It seems that to accurately model device characteristics, the semiconductor device equations in their two-, or even three-dimensional forms must be solved. But in modeling long-channel MOSFET behavior this complexity can in fact be circumvented. That is, in such structures the solutions of the semiconductor device equations in their one-dimensional form yield drain-current device characteristics that are quite close to those obtained from measurements of actual long-channel MOSFETs.

Why is it possible to treat long-channel MOSFET behavior as a one-dimensional problem and still obtain relationships that give accurate estimates of  $I_D$  as a function of the applied terminal voltages? First, consider the case of the MOSFET in which no drain bias is applied, and  $V_{GB}$  is the only external bias. We know that  $V_{GB}$  produces an electric field perpendicular to the device surface. Now, if  $L \gg (d_{\text{source}} + d_{\text{drain}})$  the junctions extend over such a small fraction of  $L$  that the  $\mathcal{E}$ -fields of these junctions have negligible impact on the  $\mathcal{E}$ -field due to the gate. Likewise, insofar as the channel width is concerned, if  $Z \gg d$  the extension of the gate depletion region beyond the sides of the channel along which there is no source or drain hardly alters the total ionic charge uncovered by the gate voltage. Thus, the fringing field effect in the  $z$  direction can also be ignored. If both of these conditions are met, we can approximate the  $\mathcal{E}$ -field in a long-channel MOSFET when  $V_{DS} = 0V$  as being perpendicular to the surface everywhere along the channel. In this case, the one-dimensional analysis we applied to the MOS-C can also be used to determine  $Q_i$ ,  $\mathcal{E}$ ,  $\phi$  and  $V_T$  in a long-channel MOSFET.

Next, if a drain bias  $V_{DS}$  is simultaneously applied, we assume that in long-channel MOSFETs the *gradual-channel approximation* is valid. This approximation assumes that the two aspects of drain current flow can be treated as strictly one-dimensional, independent problems. The two aspects are: (1) that the gate field acts approximately vertically to induce the channel (i.e., the mobile charge concentration that participates in  $I_D$ ), and; (2) that the drain-to-source voltage acts approximately horizontally to move these mobile charges, and thus to drive the current  $I_D$  in the channel. The assumption that these two one-dimensional effects are independent of one another permits the results of the MOS-C analysis (shown to be valid in the case of a MOSFET with no drain bias), to also be applied when deriving the  $I_D$ - $V_{DS}$  relationship of the MOSFET. Since the gradual-channel approximation is such a key simplification invoked when developing the models of long-channel MOSFET behavior, we need to examine it more rigorously.

\* This is the case whether or not a drain current flows in the MOSFET. That is, if a gate bias is applied, but  $I_D = 0$ , there will still be electric fields arising from the source and drain junctions at the edges of the channel. If a drain current flows, it is obvious that there must be a component of the electric field along the  $y$ -direction. In fact, each of these cases should be examined separately when considering our definition in detail.

### 4.3.1 The Gradual Channel Approximation (GCA)

If a drain bias  $V_{DS}$  is applied to a MOSFET in inversion, drain current  $I_D$  flows. Owing to the channel's finite resistance and the passage of current, a potential variation occurs along the channel length (i.e., if the source is grounded, the channel potential near the source is close to ground, whereas near the drain it is close to  $V_{DS}$ ). This causes the inversion-layer (mobile) charge density to decrease with position toward the drain because the gate is assumed to be an equipotential region (i.e., the voltage along the gate is invariant with position  $y$  along the channel), making the gate-to-channel voltage decrease with position toward the drain. Since  $I_D$  must remain constant at all points in the channel under dc biasing conditions, the electric field parallel to the flow of  $I_D$  ( $E_y$ ) must increase toward the drain to offset the decrease in mobile charge density. Thus, a gradient in the  $E_y$  value must exist in the channel (i.e.,  $\partial E_y/\partial y \neq 0$ ). The *gradual-channel approximation* (GCA), however, assumes that this gradient  $\partial E_y/\partial y$  is much smaller than the  $E$ -field gradient in the  $x$ -direction,  $\partial E_x/\partial x$  (i.e.,  $\partial E_x/\partial x \gg \partial E_y/\partial y$ ).

The two-dimensional form of the Poisson equation with  $\partial^2\phi/\partial x^2$  and  $\partial^2\phi/\partial y^2$  replaced respectively by  $-\partial E(x,y)/\partial x$  and  $-\partial E(x,y)/\partial y$  can be written as:

$$\frac{\partial E(x,y)}{\partial x} + \frac{\partial E(x,y)}{\partial y} = \frac{\rho(x,y)}{\kappa_{si} \epsilon_0} \quad (4-13)$$

If the GCA approximation is valid, the  $\partial E(x,y)/\partial y$  term in Eq. 4-13 can be neglected. Then, Eq. 4-13 can be approximated as a one dimensional differential equation with the form:

$$\frac{dE_x(x)}{dx} \approx \frac{\rho(x,y)}{\kappa_{si} \epsilon_0} \quad (4-13a)$$

The GCA is generally valid when the channel of the MOSFET is long enough (e.g.,  $L \sim 2 \mu\text{m}$  or longer), in which case the electric field strength along the  $y$ -direction  $E_y$  changes gradually from a small value near the source to values of the order of  $V_{DS}/L$  near the drain. Thus, neglecting the  $\partial E(x,y)/\partial y$  term in such devices is appropriate, and the one-dimensional form of the Poisson equation is sufficiently accurate for calculating  $Q_I$  used in the expressions for  $I_D$ . However, at large drain biases, the GCA always becomes invalid (even in long-channel MOSFETs), at least in a certain section of the channel near the drain where the gradient of the longitudinal field  $E_y$  becomes comparable to or larger than the gradient of the transverse field  $E_x$ . In such cases, as will be described later, the drain current then becomes nearly independent of drain bias, and the MOSFET is said to be in *saturation*. This implies that under conditions of bias that cause saturation, the  $I_D$  behavior of long-channel MOSFETs can no longer be adequately modeled by invoking the GCA along the entire length of the channel. Instead, modifications to the long-channel models must be made (as described in chap. 5).

The importance of being able to invoke the gradual channel approximation should be emphasized, as this is the basis of the derivation of the long-channel MOSFET  $I_D$ - $V_{DS}$  models. That is, by invoking the GCA, the problem of having to treat the semiconductor equations in their two-dimensional form to calculate  $I_D$  in a MOSFET is

avoided. Instead, the task is broken down into two one-dimensional analyses, each independent of the other. As a result,  $I_D$  is determined by the superposition of two one-dimensional solutions. Namely, by assuming that the *gradient* of the "horizontal" field in the channel is negligible, the one-dimensional *Poisson equation* can be solved to get the total charge in the silicon at each point  $y$  along the channel (i.e., the sum of the inversion layer charge  $Q_I$  [see Fig. 3-8] and the ionic charge  $Q_B$  in the depletion region). This means that the same formulas derived for the MOS-C are also applicable for calculating the local values of  $Q_I(y)$  and  $Q_B(y)$  in the long-channel MOSFET (provided, of course, that the effect of the channel voltage caused by the drain-source bias is also included in the MOS-C formulas). Then we note that within the conducting channel of a MOSFET the current flows almost exclusively in the  $y$ -direction. Thus, we can ignore the fact that there must be a slight  $x$ -component in the "horizontal"  $\mathcal{E}$ -field. This allows us to solve the one-dimensional *drift-diffusion current-density equation* in the  $y$  direction for  $I_D$  as a function of the terminal voltages [using the local  $Q(y)$  values obtained from the solution of the Poisson equation in this current-density equation].

Note, however, that if the GCA is not valid, but it is still invoked in order to enable one-dimensional analysis to be used for calculating  $Q_I$ , it is unlikely that accurate values of  $I_D$  will be yielded. This is the situation that arises in so-called short-channel devices (as described in chap. 5). In such cases, to obtain good agreement between calculated and measured values of  $I_D$ , the two-dimensional forms of the semiconductor equations must be solved (usually numerically).

**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.** The validity of the GCA for a particular MOSFET structure can be checked by making rough estimates of  $\partial\mathcal{E}_x/\partial x$  and  $\partial\mathcal{E}_y/\partial y$  values under given bias conditions. Since we are most interested in  $I_D$  when the MOSFET is in strong inversion, we will establish expressions that allow the GCA to be tested if the MOSFET is being operated in strong inversion.\*

We begin this derivation by recalling that under strong inversion the inversion layer charge density  $Q_I$  is much larger than the ionic charge density  $Q_B$  in the depletion region (see section 3.3.3), and that the inversion layer is very thin ( $\sim 30$ - $300\text{\AA}$ ). Thus, in strong inversion  $\partial\mathcal{E}_x/\partial x$  in the silicon is greatest in the inversion layer region (see Fig. 4-6). Now  $\partial\mathcal{E}_x/\partial x$  can be estimated by approximating the vertical electric field value at the bottom edge of the inversion layer as zero. Using this assumption implies that  $\partial\mathcal{E}_x/\partial x \sim [\mathcal{E}_{si}(x=0) - 0]/t_{inv} = [\mathcal{E}_{si}(x=0)]/t_{inv}$ , where  $t_{inv}$  is the effective inversion layer

\* Note that in weak inversion the surface potential  $\phi_{surf}(y)$  along the channel in long-channel MOSFETs is almost constant (see section 4.7). Thus,  $\partial\mathcal{E}_y/\partial y$  is very small, implying that  $\partial\mathcal{E}_x/\partial x \gg \partial\mathcal{E}_y/\partial y$ , because  $\mathcal{E}_x$  drops from its value at  $x = 0$  to zero at the bottom edge of the channel depletion region, which is a relatively small distance. Thus, in long-channel MOSFETs the GCA is also valid in weak inversion. However, in short-channel MOSFETs, this will no longer necessarily be the case, as we shall see in chap. 5.

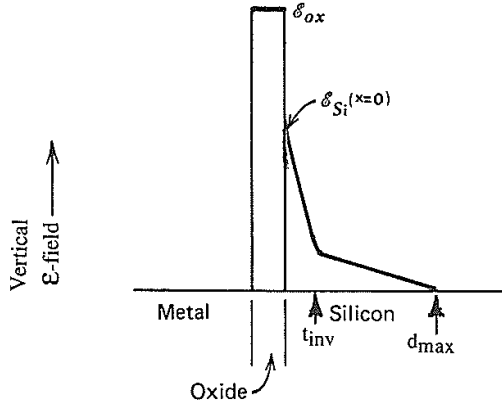


**Fig. 4-6** The vertical electric field in a MOS-C (or MOSFET channel) when a large enough gate bias to cause inversion is applied. Note that since  $Q_I > Q_B$  in inversion, that the gradient of the  $\mathcal{E}$ -field over the inversion layer thickness  $t_{inv}$  is larger than that over the depletion region width  $d_{max}$ , and we can therefore approximate the  $\mathcal{E}$ -field value at the bottom of the inversion layer thickness as being equal to zero for the purposes of estimating  $t_{inv}$ .

thickness. From Eq. 3-15,  $\mathcal{E}_{si}(x=0) = \epsilon_{ox}\mathcal{E}_{ox}/\kappa_{si}\epsilon_o$ , and from Eq. 3-14,  $\mathcal{E}_{ox}$  is  $\sim(V_{GS} - V_T)/t_{ox}$ . Thus,

$$\frac{\partial \mathcal{E}_x}{\partial x} \approx \left( \frac{\epsilon_{ox}}{\kappa_{si}\epsilon_o t_{inv}} \right) \frac{(V_{GS} - V_T)}{t_{ox}} \quad (4 - 14a)$$

Now, according to Eq. 3-12a, the concentration of electrons at some point in the semiconductor at equilibrium is exponentially dependent on the potential, with the exponential constant being  $1/\phi_t = kT/q$ . This implies that the majority of the charges is contained within a distance from the surface over which  $\phi$  drops by  $kT/q$  (i.e., since Eq. 3-12a predicts that if the electron concentration at the surface [ $x = 0$ ] is  $n_{surf}$ , then at the point where the potential drops below  $\phi_{surf}$  by an amount  $kT/q$ ,  $n_{surf}$  will have dropped to a value  $1/e$  of its value at  $x = 0$ , or to  $0.37n_{surf}$ ). Thus, we assume that  $t_{inv}$  is roughly equal to that distance below  $x = 0$ . If we also assume that the vertical electric field  $\mathcal{E}_x$  in the inversion layer is roughly the potential difference across the inversion layer  $\Delta\phi_{inv}$  divided by its thickness  $t_{inv}$  ( $\mathcal{E}_x \sim \Delta\phi_{inv}/t_{inv}$ ) we get  $t_{inv} \sim (kT/q)/\mathcal{E}_{si}(x=0)$ . Thus, substituting this expression for  $t_{inv}$  into Eq. 4-14a we finally obtain the following expression for roughly estimating  $\partial \mathcal{E}_x/\partial x$ :

$$\frac{\partial \mathcal{E}_x}{\partial x} \approx \left( \frac{\epsilon_{ox}}{\kappa_{si}\epsilon_o} \right)^2 \frac{q (V_{GS} - V_T)^2}{kT t_{ox}^2} \quad (4 - 14b)$$

If we also assume that the value of  $\mathcal{E}_y$  varies gradually from a small value near the source to values of the order of  $V_{DS}/L$  near the drain,  $\partial\mathcal{E}_y/\partial y$  can be roughly estimated as  $\partial\mathcal{E}_y/\partial y \sim V_{DS}/L^2$ . Thus, if we assume that the GCA is valid when the ratio  $(\partial\mathcal{E}_x/\partial x)/\partial\mathcal{E}_y/\partial y$  is large, then it is also valid if the following inequality is large

$$\frac{\partial\mathcal{E}_x/\partial x}{\partial\mathcal{E}_y/\partial y} = \left(\frac{\epsilon_{ox} L}{\kappa_{si}\epsilon_0}\right)^2 \frac{q (V_{GS} - V_T)^2}{kT t_{ox}^2 V_{DS}} \gg 1 \quad (4 - 15)$$

For a MOSFET at 300K with  $L = 1.0 \mu\text{m}$ ,  $t_{ox} = 300\text{\AA}$ ,  $V_{GS} - V_T = 0.5\text{V}$ , and  $V_{DS} = 0.5\text{V}$ , the left-hand side of the inequality (4-15) is  $\sim 2300$ , indicating that the GCA is a very good approximation for such a MOSFET. This also implies that the GCA can be valid even in submicron MOSFETs, provided that  $V_{GS} - V_T$  is not too small.

However, as we will see in chap. 5,  $\mathcal{E}_y$  is quite non-uniform along the length of the channel at high values of  $V_{DS}$ , becoming much stronger near the drain. Furthermore, from Eq. 4-10 we note that the inversion layer charge density can be approximated as  $Q_I \sim [V_{GS} - V_T - V(y)]$ , if we replace  $V_{GB}$  with  $V_{GS}$ , and  $V_{CB}$  with  $V(y)$  [where  $V(y)$  is the channel voltage induced by  $V_{DS}$ ]. However, this approximation is only valid when  $Q_I$  is positive, i.e., when  $V_{GS} - V_T > V(y)$ . Thus, if  $V_{DS}$  exceeds  $V_{GS} - V_T$  (i.e., the condition which occurs in the region of the channel near the drain when the MOSFET is operated in saturation), the GCA becomes invalid in this region of the channel. [Because, when  $V(y)$  is greater than  $V_{GS} - V_T$ , this implies that  $\mathcal{E}_y > \mathcal{E}_x$ , and that  $\partial\mathcal{E}_y/\partial y$  is therefore at least comparable to, if not greater than,  $\partial\mathcal{E}_x/\partial x$ ].

In summary, the GCA appears to be valid for MOSFETs operated in strong inversion even if the channel length is shorter than  $1 \mu\text{m}$  - provided that  $V_{DS}$  is small enough that the MOSFET is not being operated in saturation (i.e.,  $V_{DS} \leq V_{DSSat}$ , where  $V_{DSSat}$  is the voltage at the onset of saturation). On the other hand, if the MOSFET is operated in saturation, use of the GCA to calculate  $I_D$  will not yield accurate results, and alternative approximations (as described in chap. 5) must be employed to get good agreement with experimental  $I_D$ - $V_{DS}$  characteristics.

#### 4.3.2 Circuit Characteristics of Long-Channel MOSFETs

From the perspective of the MOSFET as a circuit component, long-channel behavior has been specified in a number of ways. A list of device parameters from a circuit perspective deemed to be characteristic of long-channel MOSFETs include the following:

1. The threshold voltage  $V_T$  is independent of channel length  $L$ .
2. The threshold voltage  $V_T$  is independent of drain bias voltage,  $V_{DS}$ .
3. The drain current in saturation  $I_{DSat}$  is independent of  $V_{DS}$ .
4. The drain current has a linear dependence on  $1/L$  (i.e., one quantitative criterion is that there should be no more than a 10% departure from linearity as the channel length is decreased).

5. The subthreshold current  $I_{Dst}$  is independent of drain bias.
6. The subthreshold swing  $S_t$  is independent of gate length.

As the behavior of the MOSFET is described later in this chapter, the terms associated with these device characteristics will be defined and discussed in more detail.

## 4.4 QUALITATIVE DESCRIPTION OF MOSFET OPERATION

The operation of the MOSFET is first described from a qualitative perspective. Specifically, our goal is to qualitatively explain the so-called *static characteristics* of the MOSFET; i.e., the drain current characteristics of the MOSFET as a function of the applied dc terminal voltages  $V_{GS}$  and  $V_{DS}$ . This behavior is normally depicted using an  $I_D$ - $V_{DS}$  plot, with  $V_{GS}$  as the parameter.

For simplicity we consider an ideal (i.e.,  $V_{FB} = 0$ ) long-channel NMOSFET ( $p$ -type substrate) as our example device structure, with the source being chosen as the reference terminal (Fig. 4-4b). We also assume that the source-substrate and drain-substrate junctions are always kept under reverse bias and that the resulting reverse-bias leakage currents are negligible (this is valid as long as the bias voltage is kept smaller than the junction breakdown voltage), and that  $V_{SB} = 0V$ . We also use energy-band versus position and potential versus position diagrams to help depict the qualitative MOSFET behavior. Note that in the case of the MOS-C one-dimensional diagrams were adequate. Here, however, two-dimensional energy-band and potential diagrams must be used. Axes and terminal voltage designations are the same as in Fig. 4-2a.

### 4.4.1 MOSFET Behavior in Response to Gate Bias Alone

We begin by describing how a MOSFET behaves in response to a gate bias  $V_{GS}$  alone. In this case, the region under the oxide responds in the same way as does an ideal MOS-C. However, because the source and drain regions are present in a MOSFET the overall device behavior differs from that of a MOS-C with a  $V_{GB}$  applied.

To analyze the MOSFET behavior, let us assume that all three bias voltages are initially set to zero (i.e.,  $V_{GS} = V_{DS} = V_{BS} = 0V$ ). In this case a flat band condition in exists in the gated region of the MOSFET (as is also observed when  $V_{GB} = 0V$  in an ideal MOS-C). The two-dimensional energy-band diagram corresponding to this condition is shown in Fig. 4-7b. We observe that  $E_c$  in the  $x$ -direction at the cut A-A is indeed flat. (The potential versus position diagram in the MOSFET for this bias condition is also shown in Fig. 4-8b). Assuming that  $\phi(x = \infty) = 0$ , in this case  $\phi_{surf} = 0$  as well.)

At the surface ( $x = 0$ ), however, a potential energy barrier of value  $qV_{bi}$  exists between the  $p$  region and each of the  $n^+$  regions (where  $V_{bi}$  is the built-in potential of the *doped*  $pn^+$  junction). This is shown in the  $E_c$  versus  $y$  direction plot for  $x = 0$  in Fig. 4-7b. For this case (and for cases of when a reverse bias exists across the doped junctions) only the very small leakage current of a reverse-biased  $pn$  junction flows.

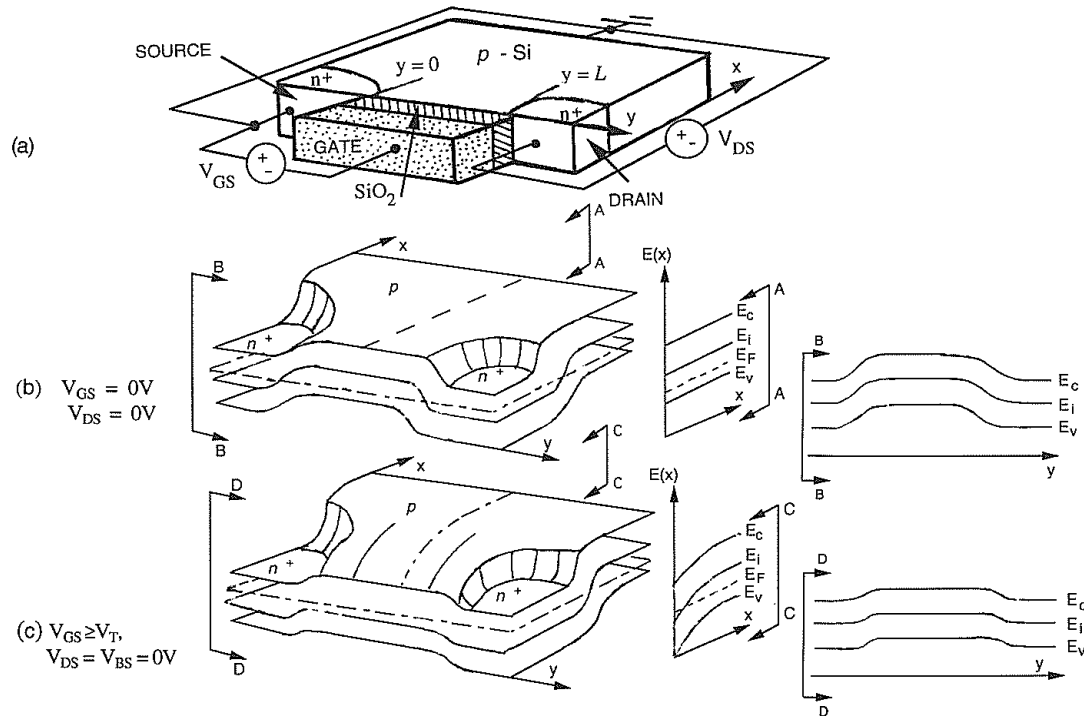


Fig. 4-7 (a) Cross section of an NMOS device. (b & c) 2-D perspective plots of the energy bands versus position of an ideal long-channel NMOS device when (b)  $V_{GS} = 0V$  (flat-band condition), and  $V_{DS} = 0V$ ; and (c)  $V_{GS} \geq V_T$ ,  $V_{DS} = V_{BS} = 0V$ .

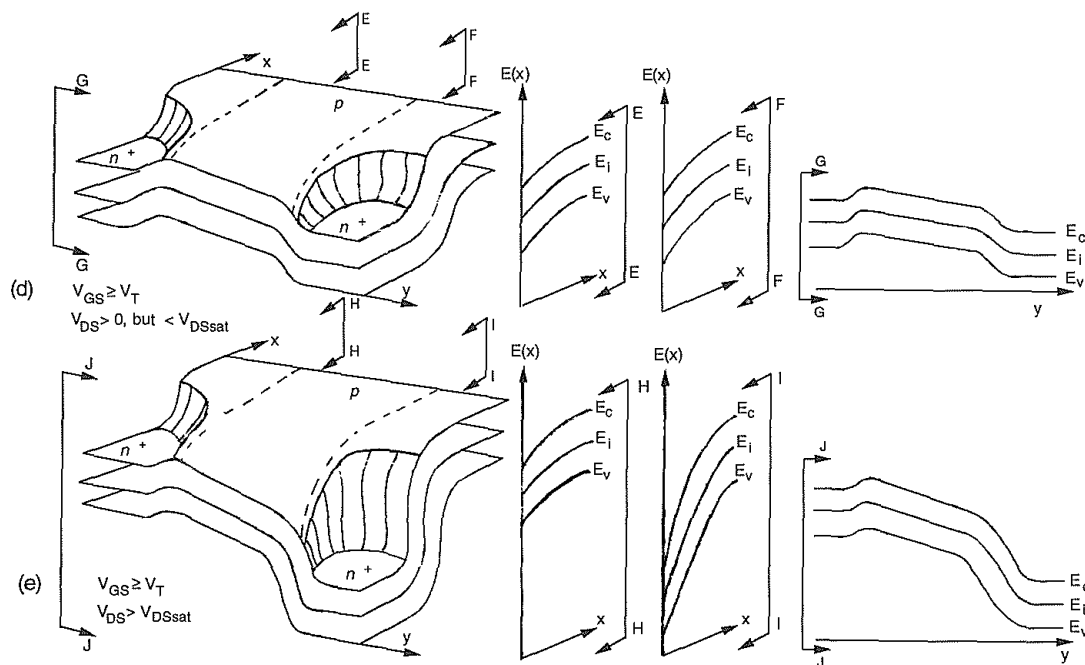


**Fig. 4-8** (cont.) (d & e) 2-D perspective plots of the energy bands as a function of position an ideal long-channel NMOS device when: (d)  $V_{GS} \geq V_T$  and  $V_{DS}$  is small (i.e.,  $< V_{DSsat}$ ); and (e)  $V_{GS} \geq V_T$  and  $V_{DS} > V_{DSsat}$ .  $V_{BS} = 0$  in both cases.

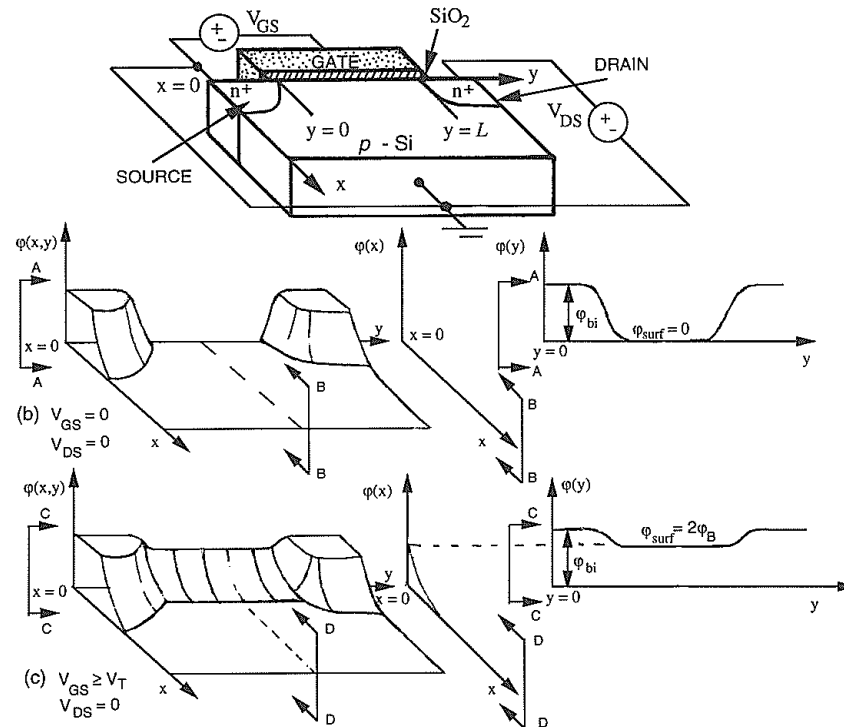


Fig. 4-8 (a) Cross section of an NMOS device. (b & c) 2-D perspective potential versus distance diagrams of an ideal NMOS device when: (b)  $V_{GS} = 0V$  (flat-band condition), and  $V_{DS} = 0V$ ; and (c)  $V_{GS} \geq V_T$ ,  $V_{DS} = V_{BS} = 0V$ .

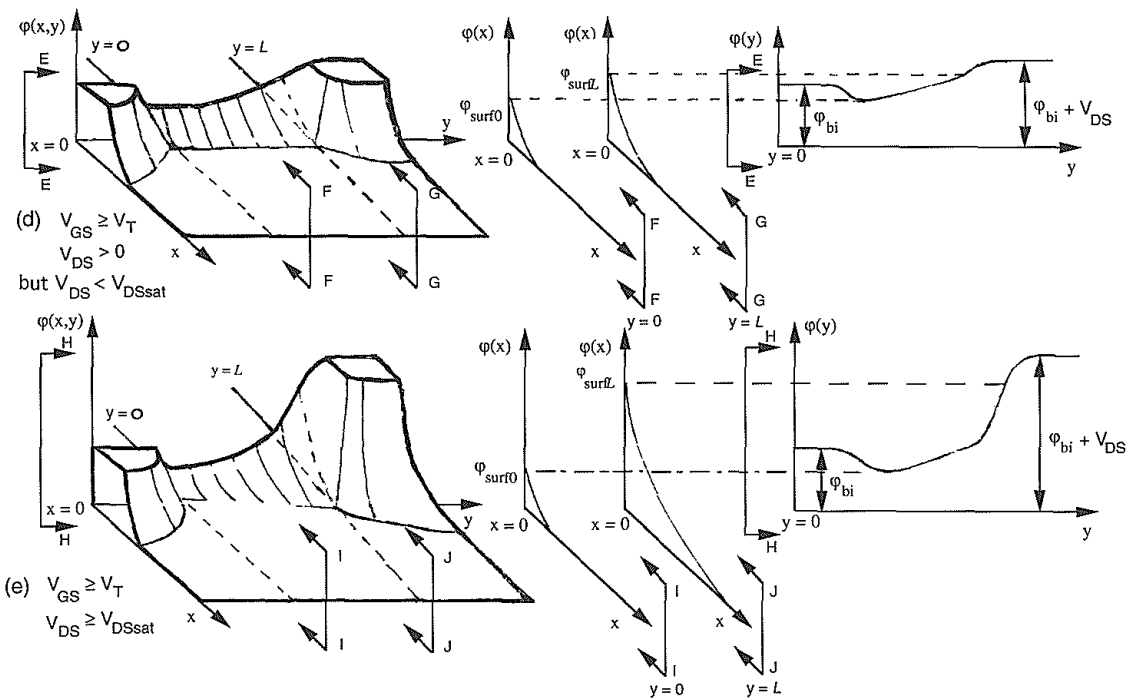


Fig. 4-8 (cont.) (d & e) 2-D perspective potential versus distance diagrams of an ideal long-channel NMOS device when (d)  $V_{GS} \geq V_T$  and  $V_{DS} > 0V$ , but  $< V_T$ ; and (e)  $V_{GS} > V_T$ , and  $V_{DS} > V_{DSsat} \sim V_{GS} - V_T$ .

Thus, the circuit path between the  $n^+$  source and drain regions effectively appears as an open circuit. This means that even if a voltage is applied between the source and drain ( $V_{DS} \neq 0V$ ), essentially no drain current will flow. Moreover, if we apply either a negative  $V_{GS}$  (or a positive  $V_{GS}$  such that  $V_{GS} < V_T$ ) the gated region contains either an excess of holes (*accumulation*, when  $V_{GS} < 0V$ ) or a deficit of holes and very few electrons (*depletion*, when  $V_{GS} < V_T$ ). In this case, the circuit path between source and drain still appears essentially as an open circuit. That is, no current flows in the drain terminal even if a drain voltage is applied. For all of these conditions, the MOSFET is said to be in an "OFF" state.

If a positive  $V_{GS}$  is applied such that  $V_{GS} > V_T$ , an inversion layer containing mobile electrons is formed adjacent to the Si surface throughout the entire gated region between the source and drain (here we still assume that  $V_{DS} = V_{BS} = 0V$ ). Thus, the circuit path between source and drain now contains an induced " $n$ -type" region (the inversion layer), as shown in Fig. 4-9. This means that a conducting path or *channel* exists between the source and drain. (However, if  $V_{DS} = 0V$ , no drain current will yet flow.) Now, if a drain voltage is applied (i.e.,  $V_{DS} > 0V$ ), drain current will flow. The device in this state is said to be "ON." Obviously, the larger the value of gate voltage ( $V_{GS}$ ), the more electrons pile up at the Si surface, and the greater the conductance of this channel. This demonstrates that the input voltage to the MOSFET can induce a source-to-drain channel and will determine the maximum conductance of the channel.\*

Figure 4-7c shows the energy band diagram of the MOSFET when  $V_{GS} > V_T$  and  $V_{DS} = V_{BS} = 0V$ , and Fig 4-8c also shows the potential diagram for this case. Note that the bands along the  $x$ -direction in the gated region are no longer flat (see section B-B) and that  $\phi_{surf} > 2\phi_B$  everywhere along the channel. The potential-energy barrier between the  $n^+$  and  $p$  regions is also reduced (see the  $E_c$  versus  $y$  plot in Fig. 4-7c). Hence, only a very low barrier to current injection exists between the source and the channel. The barrier height is smallest at the surface with its value controlled by the bias on the gate. At strong inversion and beyond, the barrier height is assumed to be remain constant at  $\sim q(V_{bi} - 2\phi_B)$ , because  $\phi_{surf}$  is assumed to be clamped at  $2\phi_B$  in strong inversion. However, as  $V_{GS}$  is increased, the inversion charge density continues to increase, increasing the conductance of the channel

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

We now consider the variation of  $I_D$  with  $V_{DS}$ . At first we assume that  $V_{GS}$  is kept fixed at a single value while  $V_{DS}$  is varied, and  $V_{BS} = 0V$ . Later we show how varying the value of  $V_{GS}$  impacts the  $I_D$ - $V_{DS}$  characteristic. In this section we also assume that

\* Note that when  $V_{GS}$  is positive but smaller than  $V_T$ , the gated region can be in state of weak inversion. In weak inversion only small, so-called *subthreshold currents* are able to flow between the drain and source. These subthreshold currents are not large enough to be useful as driving currents for most applications, but they are sufficiently large to cause troublesome leakage - which can complicate the design of some circuits. We will describe subthreshold currents in more detail in section 4.7 and in chap. 5, as part of discussion on the quantitative behavior of the long-channel MOSFET and the submicron MOSFET.

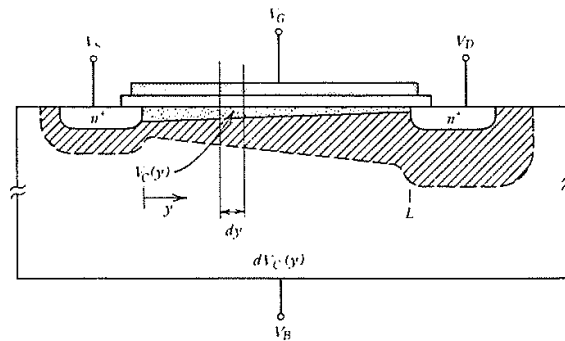


Fig. 4-9 Depletion region width variation from source to drain end of a MOSFET when  $V_{DS} > 0V$ .

sufficient gate bias is applied to ensure that the MOSFET is being operated in inversion (i.e.,  $V_{GS} > V_T$ ). We then describe what happens as a dc  $V_{DS}$  is applied, with its value being increased in small steps, starting from  $V_{DS} = 0V$ .

As noted earlier, when  $V_{DS} = 0V$ ,  $I_D = 0$ , even if  $V_{GS} > V_T$  (see Fig. 4-10b). However, when  $V_{DS}$  is increased slightly,  $I_D$  starts to flow. When  $V_{DS}$  is small (i.e., a few tenths of a volt or less), the surface channel behaves like a simple resistor (Fig. 4-10c). That is, an  $I_D$  proportional to  $V_{DS}$  flows into the drain. If a plot of  $I_D$  versus  $V_{DS}$  is used to demonstrate the  $I_D$ - $V_{DS}$  behavior, the response to small values of  $V_{DS}$  is shown as the line from the point 0 to the point A in Fig. 4-10a.

When  $V_{DS}$  is increased beyond a few tenths of a volt, the MOSFET exhibits a new phase of behavior. That is, current flowing in the channel gives rise to a significant voltage drop in the channel (unlike the case in the three-terminal structure introduced in section 4.2.1, where the channel voltage is constant along the entire channel length because essentially no current flows along it). Specifically, the *channel voltage* (introduced in section 4.2.1) is  $V_{DB}$ , at the drain end and  $V_{SB}$  at the source end. If  $V_{SB} = 0V$ , the channel voltage at the source end will equal zero. But even with  $V_{SB} = 0V$ , the channel voltage at the drain end will not be zero, since we assume  $V_{DS} > 0V$ . As shown in Fig. 4-9 such a non-zero  $V_{DB}$  causes the depletion region under the gate to widen toward the drain end of the channel. As described in section 4.2.1, this implies that the mobile charge concentration in the inversion layer simultaneously decreases. The smaller number of carriers causes the channel conductance to decrease, which in turn is manifested as a smaller slope in the  $I_D$ - $V_{DS}$  characteristic as  $V_{DS}$  is increased. As the drain voltage is further increased, the depletion region continues to widen and the slope of the  $I$ - $V$  characteristic continues to decrease (see Figs. 4-10a and 4-10c).

The energy-band and potential diagrams of the MOSFET when both a small  $V_{DS}$  and a  $V_{GS} > V_T$  are applied are shown in Figs. 4-7d and 4-8d, respectively. Even if the drain voltage is small enough so that the entire surface region remains in inversion, we see that near the drain the band bending is larger than at the source (see the  $E_C$  vs.  $y$  diagram in Fig. 4-7d). This implies that the depletion region under the gate is wider there. Since the gate-to-body voltage is constant across the entire gated region, less voltage is

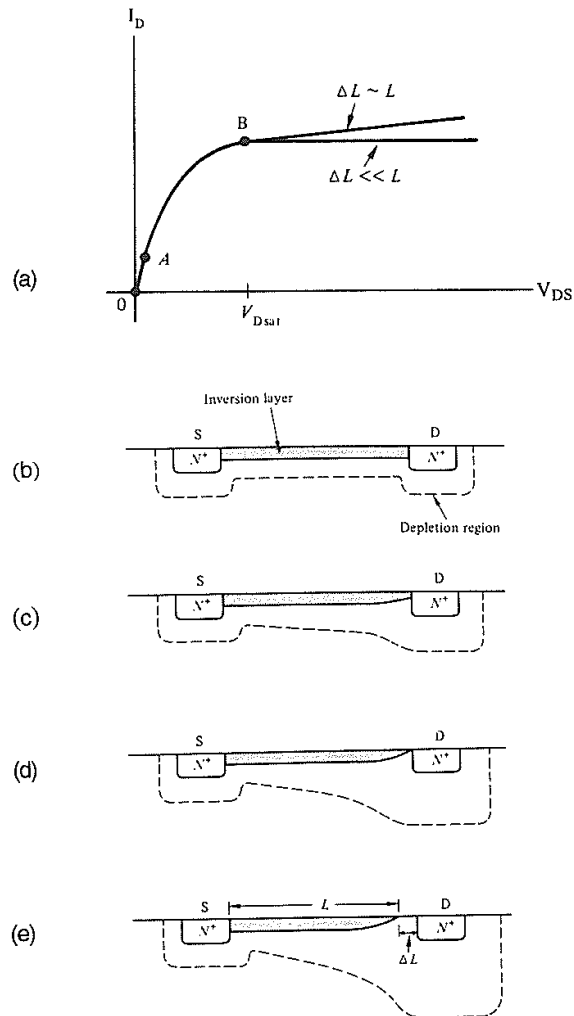


**Fig. 4-10** (a)  $I_D$ - $V_{DS}$  characteristic of a MOSFET being operated at a fixed  $V_{GS}$ , with  $V_{GS} > V_T$ . (b-e) Visualization of the various phases of  $V_{GS} > V_T$  MOSFET operation: (b)  $V_{DS} = 0V$ ; (c) channel (inversion layer) narrowing under moderate  $V_{DS}$  biasing; (d) pinchoff; and (e) post pinchoff ( $V_{DS} > V_{DSsat}$ ).<sup>30</sup> Reprinted with permission of Addison-Wesley Pub. Co.

available near the drain to establish the inversion layer charge. Hence, inversion is less strong in the channel near the drain.

Eventually, if  $V_{DS}$  is sufficiently increased such that  $V_{DS} > (V_{GS} - V_T)$ , the inversion layer ceases to exist near the vicinity of the drain. The disappearance of the inversion layer due to an increase in  $V_{DS}$  is termed *pinch-off* (see Fig. 4-10d). The pinch-off point on the  $I_D$ - $V_{DS}$  characteristic is shown as point B in Fig. 4-10a. We see that the slope of the  $I_D$  curve here has become approximately zero. Note, however, that pinch-off does *not* mean that  $I_D$  vanishes as  $V_{DS}$  exceeds the pinch-off voltage. In fact, it is the *gradual channel approximation* that is no longer valid in the region of the device in which pinch-off exists. That is, mobile charges still move through the depleted region near the drain, and drain current continues to flow. However, the equations for the charge distribution derived from the one-dimensional solution of the Poisson equation (which are valid for the case of the MOS-C) predict that no mobile charge exists in the pinch-off region. But in the depleted pinched-off region the one-dimensional Poisson equation is no longer valid. Instead, a two-dimensional numerical analysis (or an analytical expression based on a modified 1-D form of the Poisson equation - see section 5.4.3.3) must be used to calculate the potential and carrier concentrations in the pinched-off device region. A detailed numerical analysis of the saturation range has been carried out using the two-dimensional device simulator MINIMOS.<sup>3</sup> This work showed that for voltages greater than  $V_{DSSat}$  the current stream of the carriers moving into the pinched off region is no longer confined to the inversion layer region near the surface, but begins to move away from the surface and into the bulk. Note that this physical picture is plausible because once  $V_{DS}$  becomes larger than  $(V_{GS} - V_T)$  the polarity of the vertical electric field in the pinched-off region is reversed. This tends to push any mobile carriers flowing in the channel *away* from the surface. Thus, in the region of the channel where  $V_{DS} > (V_{GS} - V_T)$  the mobile carriers flow toward the drain *beneath* the surface.

Nevertheless, as the drain voltage increases beyond the pinch-off voltage  $V_{DSSat}$  the pinched-off section of the channel widens from just a point into a segment  $\Delta L$  in length (see Fig. 4-10e). Most of the voltage drop in excess of  $V_{DSSat}$  is dropped across this pinched-off section  $\Delta L$ .  $I_D$  remains approximately constant (*saturated*) for drain voltages greater than  $V_{DSSat}$  so long as  $\Delta L \ll L$ . If  $\Delta L$  becomes comparable to  $L$  (which is a condition that may occur in short-channel MOSFETs), the same voltage drop appears across a shorter channel ( $L - \Delta L$ ), and the post-pinch-off current  $I_{Dsat}$  in such devices will increase somewhat as  $V_{DS}$  is increased further beyond  $V_{DSSat}$ . Figures 4-7e and 4-8e show the energy-band diagrams and potential diagrams, respectively, for a long-channel MOSFET biased into pinch-off.

The response of the long-channel MOSFET to separate manipulation of gate and drain dc biases has been qualitatively described up to this point. This has allowed us to visualize the general variation of  $I_D$  with  $V_{DS}$  for a fixed value of  $V_{GS} > V_T$ . To allow the complete set of  $I_D$  versus  $V_{DS}$  characteristics to be obtained, we need to combine the results from the individual cases. These are as follows:

For  $V_{GS} < V_T$ , no channel exists and  $I_D \approx 0$  for all drain biases below junction breakdown. For each value of  $V_{GS} > V_T$ , a characteristic of the form shown in