

**SILICON PROCESSING
FOR
THE VLSI ERA**

**VOLUME 2:
PROCESS INTEGRATION**

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

LATTICE PRESS

Sunset Beach, California

滄海書局 總代理
台中市西屯區文華路88之1號
TEL: (04) 451-8787
FAX: (04) 451-5160

DISCLAIMER

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

Published by:

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

Cover design by Roy Montibon, Visionary Art Resources, Inc., Santa Ana, CA.

Copyright © 1990 by Lattice Press.

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

Library of Congress Cataloging in Publication Data
Wolf, Stanley

Silicon Processing for the VLSI Era
Volume 1 : Process Integration

Includes Index

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

86-081923

ISBN 0-961672-4-5

9 8 7 6 5

PRINTED IN THE UNITED STATES OF AMERICA

To my wife, Carrol Ann,
and my children, Jennifer Laura and Stanley Charles Ross

CONTENTS

PREFACE

CHAP. 1 - PROCESS INTEGRATION FOR VLSI AND ULSI

1

1.1 PROCESS INTEGRATION 5

- 1.1.1 Process Sequence Used to Fabricate an Integrated-Circuit MOS Capacitor, 5
- 1.1.2 Specifying a Process Sequence, 6
- 1.1.3 Levels of Process Integration Tasks, 7

1.2 PROCESS-DEVELOPMENT AND PROCESS-INTEGRATION ISSUES 8

REFERENCES 11

CHAP. 2 - ISOLATION TECHNOLOGIES FOR INTEGRATED CIRCUITS

12

2.1 BASIC ISOLATION PROCESSES FOR BIPOLAR ICs 13

- 2.1.1 Junction Isolation, 13
 - 2.1.1.1 Junction Isolation in the SBC Process
 - 2.1.1.2 Collector-Diffusion Isolation

2.2 BASIC ISOLATION PROCESS FOR MOS ICs (LOCOS ISOLATION) 17

- 2.2.1 Punchthrough Prevention between Adjacent Devices in MOS Circuits, 20
- 2.2.2 Details of the Semirecessed Oxide LOCOS Process, 20
 - 2.2.2.1 Pad-Oxide Layer.
 - 2.2.2.2 CVD of Silicon Nitride Layer.
 - 2.2.2.3 Mask and Etch Pad-Oxide/Nitride Layer to Define Active Regions.

vii

2.2.2.4	Channel-Stop Implant.	
2.2.2.5	Problems Arising from the Channel-Stop Implants.	
2.2.2.6	Grow Field Oxide.	
2.2.2.7	Strip the Masking Nitride/Pad-Oxide Layer.	
2.2.2.8	Regrow Sacrificial Pad Oxide and Strip (Kooi Effect).	
2.2.3	Limitations of Conventional Semi-Recessed Oxide LOCOS for Small-Geometry ICs,	27
2.3	FULLY RECESSED OXIDE LOCOS PROCESSES	28
2.3.1	Modeling the LOCOS Process,	31
2.4	ADVANCED SEMIRECESSED OXIDE LOCOS ISOLATION PROCESSES	31
2.4.1	Etched-Back LOCOS,	31
2.4.2	Polybuffered LOCOS,	32
2.4.3	SILO (Sealed-Interface Local Oxidation),	33
2.4.4	Laterally Sealed LOCOS Isolation,	35
2.4.5	Bird's Beak Suppression in LOCOS by Mask-Stack Engineering,	38
2.4.6	Planarized SILO with High-Energy Channel-Stop Implant,	38
2.5	ADVANCED FULLY RECESSED OXIDE LOCOS ISOLATION PROCESSES	39
2.5.1	SWAMI (Sidewall-Masked Isolation Technique),	39
2.5.2	SPOT (Self-Aligned Planar-Oxidation Technology),	41
2.5.3	FUOX (Fully Recessed Oxide),	41
2.5.4	OSELO II,	43
2.6	NON-LOCOS ISOLATION TECHNOLOGIES I: (TRENCH ETCH AND REFILL)	45
2.6.1	Shallow Trench and Refill Isolation,	45
2.6.1.1	BOX Isolation.	
2.6.1.2	Modifications to Improve BOX Isolation.	
2.6.2	Moderate-Depth Trench and Refill Isolation,	48
2.6.2.1	U-Groove Isolation.	
2.6.2.2	Toshiba Moderate-Depth Trench Isolation for CMOS.	
2.6.3	Deep, Narrow Trench and Refill,	51
2.6.3.1	Reactive Ion Etching of the Substrate.	
2.6.3.2	Refilling the Trench.	
2.6.3.3	Planarization after Refill.	

2.7	NON-LOCOS ISOLATION TECHNOLOGIES, II: SELECTIVE EPITAXIAL GROWTH (SEG)	58
2.7.1	Refill by SEG of Windows Cut into Surface Oxide,	59
2.7.2	Simultaneous Single-Crystal/Poly Deposition (SSPD),	60
2.7.3	Etching of Silicon Trenches and Refilling with SEG to Form Active Device Regions,	61
2.7.4	Selective-Epitaxial-Layer Field Oxidation (SELFOX),	61
2.7.5	SEG Refill of Trenches (as an Alternative to Poly Refill),	62
2.7.6	Epitaxial Lateral Overgrowth (ELO),	62
2.8	MISCELLANEOUS NON-LOCOS ISOLATION TECHNOLOGIES	63
2.8.1	Field-Shield Isolation,	63
2.8.2	Buried Insulator between Source/Drain Polysilicon (BIPS),	64
2.9	SUMMARY: CANDIDATE ISOLATION TECHNOLOGIES FOR SUBMICRON DEVICES	65
2.9.1	Basic Requirements of VLSI and ULSI Isolation Technologies,	65
2.9.2	The Need for Planarity,	65
2.9.3	How the Various Isolation Technologies Meet the Requirements,	66
2.10	SILICON-ON-INSULATOR (SOI) ISOLATION TECHNOLOGIES	66
2.10.1	Dielectric Isolation,	67
2.10.2	Wafer Bonding,	70
2.10.3	Silicon-on-Sapphire (SOS),	72
2.10.4	Separation by Implanted Oxygen (SIMOX),	72
2.10.5	Zone-Melting Recrystallization (ZMR),	75
2.10.6	Full Isolation by Porous Oxidized Silicon (FIPOS),	76
2.10.7	Novel SOI CMOS Processes with Selective Oxidation and Selective Epitaxial Growth,	77
REFERENCES		79

CHAP. 3 - CONTACT TECHNOLOGY AND LOCAL INTERCONNECTS FOR VLSI 84

3.1 THE ROLE OF CONTACT STRUCTURES IN DEVICE AND CIRCUIT BEHAVIOR 84

3.1.1 Contact Structures in Planar MOSFETs and Bipolar Transistors, 85

3.2 THEORY OF METAL-SEMICONDUCTOR CONTACTS 87

3.3 EXTRACTING VALUES OF SPECIFIC CONTACT RESISTIVITY FROM MEASUREMENTS 91

3.3.1 Extraction of the Specific Contact Resistivity from an Ideal Contact Structure, 92

3.3.2 Current Flow in Actual Contact Structures, 93

3.3.3 Contact Structures Used to Extract ρ_c , 94

3.3.4 Procedure for Accurately Extracting ρ_c from CBKR Test Structures, 97

3.3.5 Reported Values of ρ_c for Various Contact Structures, 100

3.3.6 Use of a Simple Contact-Chain Structure to Monitor Contact Resistance, 101

3.4 THE EVOLUTION OF CONVENTIONAL METAL-TO-SILICON CONTACTS 101

3.4.1 The Basic Process Sequence of Conventional Ohmic-Contact Structures to Silicon, 102

3.4.2 Additional Details Concerning the Processing Steps, 103

3.4.2.1 Formation of the Heavily Doped Regions in the Silicon.

3.4.2.2 Formation of Contact Openings (Etching).

3.4.2.3 Sidewall Contouring of the Contact Holes by Reflow.

3.4.2.4 Sidewall Contouring by Etching.

3.4.2.5 Deposition.

3.4.2.6 Metal Deposition and Patterning.

3.4.2.7 Sintering the Contacts.

3.4.3 Aluminum-Silicon Contact Characteristics, 111

3.4.3.1 The Kinetics of the Al-Si Interface During Sintering.

3.4.4 Use of Aluminum-Silicon Alloys to Reduce Junction Spiking, 116

3.4.5 Platinum Silicide-to-Silicon Contacts, 117

3.4.5.1 Process Sequence Used to Form PtSi-Si Contacts.

3.4.5.2 Limitations of the PtSi-Si Contact Structure.

3.5 DIFFUSION BARRIERS 121

3.5.1 Theory of Diffusion Barrier Layers, 121

3.5.2 Materials Used as Diffusion Barriers, 124

3.5.2.1 Sputter-Deposited Titanium-Tungsten (Stuffed Barrier).

3.5.2.2 Polysilicon (Sacrificial Barrier).

3.5.2.3 Titanium (Sacrificial Barrier).

3.5.2.4 Titanium Nitride (Passive Barrier).

3.5.2.5 CVD Tungsten.

3.5.2.6 Experimental Diffusion Barrier Materials.

3.6 MULTILAYERED OHMIC-CONTACT STRUCTURES TO SILICON 131

3.6.1 Al-Ti:W-PtSi-Si Contacts, 132

3.6.2 Al-TiN-Ti-Si Contacts, 132

3.6.3 Mo-Ti:W-Si and Mo-Ti-Si Contacts, 134

3.7 SCHOTTKY-BARRIER CONTACTS 134

3.8 THE IMPACT OF THE INTRINSIC SERIES RESISTANCE ON MOS TRANSISTOR PERFORMANCE 137

3.8.1 The Impact of R_s on MOSFET Performance, 137

3.8.2 Estimates of R_{sh} , R_{sp} , R_{ac} , and R_{co} , 138

3.8.3 Impact of R_s on Device Characteristics, 142

3.8.4 Summary of the Impact of Intrinsic Series-Resistance Effects on MOSFET Performance, 142

3.9 ALTERNATIVE (SELF-ALIGNED) CONTACT STRUCTURES FOR ULSI MOS DEVICES 143

3.9.1 Self-Aligned Silicide Contacts, 144

3.9.1.1 Self-Aligned Titanium Silicide Contacts.

3.9.1.2 Self-Aligned Cobalt Silicide Contacts.

3.9.1.3 Measuring r_c of Self-Aligned Silicide Contacts.

3.9.2 Buried-Oxide MOS Contact Structure (BOMOS), 153

3.10 FORMATION OF SHALLOW JUNCTIONS AND THEIR IMPACT ON CONTACT FABRICATION 154

3.10.1 Conventional Shallow-Junction Formation, 154

3.10.2 Alternative Approaches to Forming Shallow Junctions, 155

3.10.3 Impact of Shallow Junctions on Contact Formation, 160

3.11 BURIED CONTACTS AND LOCAL INTERCONNECTS 160

- 3.11.1 Butted Contacts and Buried Contacts, 160
- 3.11.2 Local Interconnects, 162
 - 3.11.2.1 *Selectively Formed TiSi₂.*
 - 3.11.2.2 *Ti:W over CoSi₂.*
 - 3.11.2.3 *TiN Formed over TiSi₂.*
 - 3.11.2.4 *Dual-Doped Polysilicon LI with Diffused Source/Drain Junctions.*
 - 3.11.2.5 *CVD W-Clad Polysilicon LI.*

REFERENCES**CHAP. 4 - MULTILEVEL INTERCONNECT TECHNOLOGY FOR VLSI AND ULSI 176****4.1 EARLY DEVELOPMENT OF INTERCONNECT TECHNOLOGY FOR INTEGRATED CIRCUITS 176**

- 4.1.1 Interconnects for Early Bipolar ICs, 176
- 4.1.2 Interconnects in Silicon-Gate NMOS ICs, 178
- 4.1.3 Evolution of Interconnects for Bipolar ICs, 179
- 4.1.4 Evolution of Interconnects for CMOS ICs, 180

4.2 THE NEED FOR MULTILEVEL INTERCONNECT TECHNOLOGIES 180

- 4.2.1 Interconnect Limitations of VLSI, 181
 - 4.2.1.1 *Functional Density.*
 - 4.2.1.2 *Propagation Delay.*
 - 4.2.1.3 *Ease of Design and Gate Utilization for ASICs and Wafer Scale Integration.*
 - 4.2.1.4 *Cost.*
- 4.2.2 Problems Associated with Multimetal Interconnect Processes, 187
- 4.2.3 Terminology of Multilevel Interconnect Structures, 188

4.3 MATERIALS FOR MULTILEVEL INTERCONNECT TECHNOLOGIES 189

- 4.3.1 Conductor Materials for Multilevel Interconnects, 189
 - 4.3.1.1 *Requirements of Conductor Materials Used for VLSI Interconnects.*
 - 4.3.1.2 *Local Interconnect Conductor Materials (Polysilicon, Metal-Silicides, and Polycides).*

- 4.3.1.3 *Aluminum Metallization.*
- 4.3.1.4 *Tungsten and Other Conductor Materials for VLSI Interconnects.*
- 4.3.2 Dielectric Materials for Multilevel Interconnects, 194
 - 4.3.2.1 *Requirements of Dielectric Layers in Multilevel Interconnects.*
 - 4.3.2.2 *Poly-Metal Interlevel Dielectric (PMD) Materials.*
 - 4.3.2.3 *CVD SiO₂ Films as Intermetal Dielectrics.*
 - 4.3.2.4 *Low-Temperature-TEOS SiO₂ Films as Intermetal Dielectrics.*
 - 4.3.2.5 *Other Materials and Deposition Processes Used to Form Intermetal Dielectrics.*

4.4 PLANARIZATION OF INTERLEVEL DIELECTRIC LAYERS 199

- 4.4.1 Terminology of Planarization in Multilevel Interconnects, 199
 - 4.4.1.1 *Degree of Planarization.*
 - 4.4.1.2 *The Need for Dielectric Planarization.*
 - 4.4.1.3 *The Price that Must be Paid as the Degree of Dielectric Planarization is Increased.*
 - 4.4.1.4 *Design Rules Related to Intermetal Dielectric-Formation and Planarization Processes.*
- 4.4.2 Step Height Reduction of Underlying Topography as a Technique to Alleviate the Need for Planarization, 208
 - 4.4.2.1 *Provide Substrate Topography that is Completely Planar.*
 - 4.4.2.2 *Provide a Planar Surface over Local Interconnect Levels.*
 - 4.4.2.3 *Minimize the Thickness of the Metal 1 Layer.*
 - 4.4.2.4 *Achieve Smoothing of Steps in DM1 by Sloping the Sidewalls of Metal-1 Lines.*
- 4.4.3 Deposition of Thick CVD SiO₂ Layers and Etching Back Without a Sacrificial Layer, 211
- 4.4.4 Oxide Spacers, 212
- 4.4.5 Polyimides as Intermetal Dielectrics, 214
- 4.4.6 Planarizing by Use of Bias-Sputtered SiO₂, 217
- 4.4.7 CVD SiO₂ and Bias-Sputter Etchback, 220
- 4.4.8 Planarization by Sacrificial Layer Etchback, 222
 - 4.4.8.1 *Degree of Planarization Achieved by Sacrificial Etchback.*
 - 4.4.8.2 *Advantages of the Sacrificial Etchback Process.*
 - 4.4.8.3 *Sacrificial Etchback Process Problems.*
 - 4.4.8.4 *Alternative Sacrificial Etchback Processes.*
- 4.4.9 Spin-On Glass (SOG), 449
 - 4.4.9.1 *SOG Process Integration.*
 - 4.4.9.2 *The Etchback SOG Process.*
 - 4.4.9.3 *The Non-Etchback SOG Process.*
- 4.4.10 Electron-Cyclotron-Resonance Plasma CVD, 237
- 4.4.11 Chemical-Mechanical Polishing, 238

4.5. METAL DEPOSITION AND VIA FILLING 240

- 4.5.1 Conventional Approach to Via Fabrication and Formation of Metal-to-Metal Contacts through the Vias, 240
 - 4.5.1.1 *Design Rules of Multilevel Metal Systems which are Impacted by Conventional Via Processing Limitations.*
- 4.5.2 Advanced Via Processing (Vertical Vias and Complete Filling of Vias by Metal), 244
 - 4.5.2.1 *Increases in Packing Density Resulting from Advanced Via Process Technology.*
- 4.5.3 Processing Techniques which Allow Vertical Vias to be Implemented, 245
 - 4.5.3.1 *Required Degree of Via Filling by Plugs.*
- 4.5.4 CVD W Techniques for Filling Vertical Vias and Contact Holes, 245
 - 4.5.4.1 *General Information on the CVD Tungsten Process.*
 - 4.5.4.2 *Blanket CVD W and Etchback.*
 - 4.5.4.3 *Selective CVD W.*
- 4.5.5 Other CVD Via Filling Processes, 253
 - 4.5.5.1 *Blanket CVD Polysilicon and Etchback for Contact Hole Filling.*
 - 4.5.5.2 *Selective Deposition of Poly.*
 - 4.5.5.3 *Selectively Formed Silicide Contact Plugs.*
 - 4.5.5.4 *CVD Aluminum.*
- 4.5.6 Alternatives to CVD for Filling of Vias, 254
 - 4.5.6.1 *Bias Sputtering of Al to Achieve Complete Filling of Via Holes.*
 - 4.5.6.2 *Laser Planarization of Al Films.*
 - 4.5.6.3 *Contact Hole and Via Filling by Selective Electroless Metal Deposition.*
- 4.5.7 Pillar Formation as an Alternative to Filling Contact Holes and Vias, 258

4.6 FILLED GROOVES IN A DIELECTRIC LAYER 259**4.7 MANUFACTURING YIELD AND RELIABILITY ISSUES OF VLSI INTERCONNECTS 260**

- 4.7.1 Factors Which Impact Manufacturing Yield, 261
- 4.7.2 Multilevel Interconnect-Related Yield Issues, 261
- 4.7.3 General Reliability Issues Associated with IC Interconnects, 264
 - 4.7.3.1 *Electromigration.*
 - 4.7.3.2 *Electromigration at the Contacts.*
 - 4.7.3.3 *Stress-Induced Metal Cracks and Voids.*
 - 4.7.3.4 *Corrosion.*
- 4.7.4 Reliability Issues Associated with Multilevel Interconnects, 268

- 4.7.4.1 *Hillock Formation and Prevention Measures.*
- 4.7.4.2 *Dielectric Void Reliability Problems.*

4.8 PASSIVATION LAYERS 273**4.9 SURVEY OF MULTILEVEL METAL SYSTEMS 276**

- 4.9.1 Bipolar Double-Level Metal Systems, 276
- 4.9.2 CMOS Double-Level-Metal Systems, 277
 - 4.9.2.1 *Non-Planarized DLM (2.0- μ m CMOS).*
 - 4.9.2.2 *Non-Planarized DLM: CVD-W Metal (2.0- μ m NMOS).*
 - 4.9.2.3 *Resist Etchback, Bias-Sputtered SiO₂, and SOG DLM for 1.5 μ m CMOS.*
 - 4.9.2.4 *Non-Sacrificial Layer Etchback DLM (1.0- μ m CMOS).*
 - 4.9.2.5 *Alternative CMOS DLM Process with Ti:W/Mo as Metal 1.*
 - 4.9.2.6 *DLM Processes for Submicron CMOS.*
- 4.9.3 Three-Level Metal Systems, 283
- 4.9.4 Four-Level Metal Systems, 285

4.10 SUMMARY OF MULTILEVEL INTERCONNECT TECHNOLOGY REQUIREMENTS FOR VLSI 286**REFERENCES 287****CHAP. 5 - MOS DEVICES AND NMOS PROCESS INTEGRATION 298****5.1 MOS DEVICE PHYSICS 298**

- 5.1.1 The Structure and Device Fundamentals of MOS Transistors, 298
- 5.1.2 The Threshold Voltage of the MOS Transistor, 301
- 5.1.3 Impact of Source-Body Bias on V_T (Body Effect), 304
- 5.1.4 Current-Voltage Characteristics of MOS Transistors, 305
- 5.1.5 The Capacitances of MOS Transistors, 307

5.2 MAXIMIZING DEVICE PERFORMANCE THROUGH DEVICE DESIGN AND PROCESSING TECHNOLOGY 307

- 5.2.1 Output Current (I_D) and Transconductance (g_m), 308
- 5.2.2 Controlling the Threshold Voltage through Process and Circuit-Design Techniques, 309
- 5.2.3 Subthreshold Currents (I_{Dst} when $V_G < IV_T$), 311

5.2.4	Switching Speed,	313
5.2.5	Junction Breakdown Voltage (Drain-to-Substrate),	313
5.2.6	Gate-Oxide Breakdown Voltage,	314
5.2.7	High Field-Region Threshold-Voltage Value,	315
5.3	THE EVOLUTION OF MOS TECHNOLOGY (PMOS AND NMOS)	315
5.3.1	Aluminum-Gate PMOS,	316
5.3.2	Silicon-Gate MOS Technology,	318
5.3.3	Reduction of Oxide-Charge Densities,	319
5.3.4	Ion Implantation for Adjusting Threshold Voltage,	321
5.3.5	Isolation Technology for MOS,	323
5.3.6	Short-Channel Devices,	323
5.4	PROCESS SEQUENCE FOR FABRICATING NMOS INVERTERS WITH DEPLETION-MODE LOADS	324
5.4.1	Operation of an NMOS Inverter with a Depletion-Mode Load,	324
5.4.2	Process Sequence of a Basic E-D NMOS IC Technology,	327
5.4.2.1	<i>Starting Material.</i>	
5.4.2.2	<i>Active Region and Field Region Definitions.</i>	
5.4.2.3	<i>Gate-Oxide Growth and Threshold-Voltage Adjust Implant</i>	
5.4.2.4	<i>Polysilicon Deposition and Patterning.</i>	
5.4.2.5	<i>Formation of the Source and Drain Regions.</i>	
5.4.2.6	<i>Contact Formation.</i>	
5.4.2.7	<i>Metallization Deposition and Patterning.</i>	
5.4.2.8	<i>Passivation Layer and Pad Mask.</i>	
5.5	SHORT-CHANNEL EFFECTS AND HOW THEY IMPACT MOS PROCESSING	338
5.5.1	Effect of Gate Dimensions on Threshold Voltage,	338
5.5.1.1	<i>Short Channel Threshold Voltage Effect.</i>	
5.5.1.2	<i>Narrow Gate-Width Effect on Threshold Voltage.</i>	
5.5.2	Short-Channel Effects on Subthreshold Currents (Punchthrough and Drain-Induced Barrier Lowering),	341
5.5.3	Short-Channel Effects on I-V Characteristics,	343
5.5.4	Summary of Short-Channel Effects on the Fabrication of MOS ICs,	346
5.6	HOT-CARRIER EFFECTS IN MOSFETS	348
5.6.1	Substrate Currents Due to Hot Carriers,	349

5.6.2	Hot-Carrier Injection into the Gate Oxide,	350
5.6.3	Device-Performance Degradation Due to Hot-Carrier Effects,	352
5.6.4	Techniques for Reducing Hot-Carrier Degradation,	354
5.6.5	Lightly Doped Drains,	354
5.6.5.1	<i>Drain Engineering for Optimum LDD Structures.</i>	
5.6.5.2	<i>Asymmetrical Characteristics of LDD MOSFETs.</i>	
5.6.6	The Impact of IC Processing on Hot-Carrier Device Degradation,	361
5.6.7	Hot-Carrier Effects in PMOS Transistors,	362
5.6.8	Gate-Induced Drain-Leakage Current,	363

REFERENCES 363

CHAP. 6 - CMOS PROCESS INTEGRATION 368

6.1	INTRODUCTION TO CMOS TECHNOLOGY	368
6.1.1	The Power-Dissipation Crisis of VLSI and How CMOS Came to the Rescue,	368
6.1.2	Historical Evolution of CMOS,	370
6.1.3	Operation of CMOS Inverters,	373
6.1.4	Advantages (and Disadvantages) of Modern CMOS Technologies,	376
6.1.4.1	<i>Device/Chip Performance Advantages.</i>	
6.1.4.2	<i>Reliability Advantages of CMOS.</i>	
6.1.4.3	<i>Circuit Design Advantages of CMOS.</i>	
6.1.4.4	<i>Cost Analysis of CMOS.</i>	
6.1.5	Disadvantages of CMOS,	380
6.2	THE WELL CONTROVERSY IN CMOS	381
6.2.1	The Need for Wells in CMOS,	381
6.2.2	<i>p</i> -Well CMOS,	383
6.2.3	<i>n</i> -Well CMOS,	384
6.2.4	CMOS on Epitaxial Substrates,	385
6.2.5	Twin-Well CMOS,	387
6.2.6	Retrograde-Well CMOS,	389
6.2.7	Summary of CMOS Well-Technology Issues,	392
6.3	<i>p</i>-CHANNEL DEVICES IN CMOS	392
6.3.1	PMOS Devices with <i>n</i> ⁺ -Polysilicon Gates,	392
6.3.1.1	<i>Punchthrough Susceptibility.</i>	

- 6.3.2 PMOS Devices with p^+ -Polysilicon Gates, 397
- 6.3.3 Gate Materials having Symmetrical Work Functions (with Respect to both NMOS and PMOS Devices), 398

6.4 LATCHUP IN CMOS 400

- 6.4.1 Parasitic *pnpn* Structures in CMOS Circuits, 400
- 6.4.2 Circuit Behavior of *pnpn* Diodes, 402
- 6.4.3 Device Physics Behavior of *pnpn* Diodes, 403
- 6.4.4 Summary of Conditions That Must Exist in Order for Latchup to Occur, 406
- 6.4.5 Circuit Behavior of Actual *pnpn* Structures in CMOS Circuits, 406
 - 6.4.5.1 Value of β in CMOS Vertical Parasitic Bipolar Transistors.
 - 6.4.5.2 Value of β in CMOS Lateral Parasitic Bipolar Transistors.
- 6.4.6 Circuit and Device Effects that Induce Latchup, 408
 - 6.4.6.1 An external stimulus forward-biases the emitter-base of one transistor, and its collector current then turns-on the second transistor.
 - 6.4.6.2 An external stimulus causes current to flow through both bypass resistors, forward-biasing one or both bipolar transistors.
 - 6.4.6.3 Current is shunted through one of the parasitic transistors by some degradation mechanism, and the resulting collector current flows through the bypass resistor of the second transistor and turns it on.
- 6.4.7 Test Methods for Characterizing Latchup, 410
 - 6.4.7.1 Modelling Latchup in CMOS Technology.
- 6.4.8 Techniques for Reduction or Elimination of Latchup Susceptibility, 413
 - 6.4.8.1 Processing Techniques that Reduce the Current Gains of the Parasitic Bipolar Transistors.
 - 6.4.8.2 Processing Techniques that Reduce R_{sub} and R_w or Eliminate the *pnpn* Structure.
 - 6.4.8.3 Circuit Layout Techniques used to Decouple Parasitic Bipolar Transistors.

6.5 CMOS ISOLATION TECHNOLOGY 419

- 6.5.1 Trench Isolation for CMOS, 425
- 6.5.2 Isolation by Selective-Epitaxial Growth for CMOS, 426

6.6 CMOS PROCESS SEQUENCES 428

- 6.6.1 Basic *n*-Well CMOS Process Sequence, 428
- 6.6.2 Twin-Well CMOS Process Sequence, 431
 - 6.6.2.1 Starting Material.
 - 6.6.2.2 Forming the Wells and Channel Stops.

- 6.6.2.3 Active and Field Region Definition.
- 6.6.2.4 Gate Oxide Growth and Threshold Voltage Adjustment.
- 6.6.2.5 Polysilicon Deposition and Patterning.
- 6.6.2.6 Formation of the Source/Drain Regions.
- 6.6.2.7 CVD Oxide Deposition and Contact Formation.
- 6.6.2.8 Metal 1 Deposition and Patterning.
- 6.6.2.9 Intermetal Dielectric Deposition/Planarization and Via Patterning.
- 6.6.2.10 Metal 2 Deposition and Patterning.
- 6.6.2.11 Passivation Layer Deposition and Patterning.

6.7 MISCELLANEOUS CMOS TOPICS 441

- 6.7.1 Electrostatic Discharge Protection for CMOS, 441
 - 6.7.1.1 Diode Protection.
 - 6.7.1.2 Node-to-Node Punchthrough.
 - 6.7.1.3 Gate-Controlled Breakdown Structure.
 - 6.7.1.4 *pnpn*-Diode ESD Protection for Advanced CMOS Circuits.
- 6.7.2 Power Supply Voltage Levels for Future CMOS, 446
- 6.7.3 Low-Temperature CMOS, 446
- 6.7.4 Three-Dimensional CMOS, 447

REFERENCES 447

CHAPTER 7 - BIPOLAR AND BICMOS PROCESS INTEGRATION 453

7.1 BIPOLAR TRANSISTOR STRUCTURES FOR INTEGRATED CIRCUITS 453

- 7.1.1 The Transistor Action 454
 - 7.1.1.1 Basic Bipolar Transistor Physics.
 - 7.1.1.2 Bipolar Transistor Current Gain.
- 7.1.2 Integrated-Circuit Transistor Structures 458

7.2 DIGITAL CIRCUITS USING BIPOLAR TRANSISTORS 459

- 7.2.1 Basic Bipolar-Transistor Inverter Circuits 459
- 7.2.2 Bipolar Digital-Logic-Circuit Families 460

7.3 MAXIMIZING BIPOLAR TRANSISTOR PERFORMANCE THROUGH DEVICE DESIGN & PROCESSING TECHNOLOGY 464

7.3.1	Current Gain	464
7.3.2	Early Voltage	466
7.3.3	High-Level Injection Effects (Kirk Effect)	467
7.3.4	Operating-Voltage Limits in Bipolar Transistors	468
7.3.4.1	Reachthrough Breakdown.	
7.3.4.2	Punchthrough Breakdown.	
7.3.4.3	Breakdown Voltage and High-Level Injection Limits in Advanced Bipolar Transistors.	
7.3.5	Parasitic Series Resistances in Bipolar Transistors	472
7.3.5.1	Collector Series Resistance, R_C .	
7.3.5.2	Base Series Resistance, R_B .	
7.3.5.3	Base-Spreading Resistance, R_{B2} (and Emitter Current Crowding).	
7.3.5.4	Emitter Series Resistance, R_E .	
7.3.6	Parasitic Junction Capacitances in Bipolar Transistors	475
7.3.6.1	Storage Capacitances in Bipolar Transistors.	
7.3.7	Bipolar Transistor Unity-Gain Frequency, f_T	477
7.3.8	First Order npn Device Design	477
7.3.9	Switching Speed Behavior in Bipolar ICs	478
7.3.9.1	Propagation-Delay Time Calculation in Bipolar Transistors.	
7.3.9.2	Propagation Delay in Digital MOS versus Digital Bipolar Circuits.	
7.3.9.3	General Switching Speed Behavior of Digital Bipolar Circuits.	
7.4	NON-OXIDE-ISOLATED BIPOLAR npn TRANSISTOR STRUCTURES	482
7.4.1	Triple-Diffused (3D) Process	483
7.5	STANDARD-BURIED-COLLECTOR PROCESS	483
7.5.1	Characteristics of npn Transistors Fabricated with the Standard-Buried-Collector (SBC) Process	483
7.5.1.1	Limitations of Junction-Isolated SBC Transistors for VLSI Circuits.	
7.5.2	Standard-Buried-Collector Process Flow	486
7.5.2.1	Starting Material.	
7.5.2.2	Buried Layer Formation.	
7.5.2.3	Epitaxial Growth.	
7.5.2.4	Formation of Isolation Regions.	
7.5.2.5	Deep-Collector Contact Formation (Optional).	
7.5.2.6	Base Region Formation.	
7.5.2.7	Emitter Region Formation.	
7.5.2.8	Contact and Interconnect Layer Formation.	
7.5.2.9	Washed Emitters.	
7.5.2.10	Schottky Contacts.	
7.6	OXIDE-ISOLATED BIPOLAR TRANSISTORS	498

7.7 ADVANCED BIPOLAR TRANSISTOR STRUCTURES FOR VLSI AND ULSI 500

7.8 ADVANCED EMITTER STRUCTURES 501

- 7.8.1 Polysilicon Emitters 501
 - 7.8.1.1 Models that Describe Polysilicon-Emitter Behavior.
 - 7.8.1.2 Process Technology for Polysilicon-Emitter Fabrication.
- 7.8.2 Heterojunction Bipolar Transistors (HBTs) 506

7.9 SELF-ALIGNED BIPOLAR STRUCTURES 510

- 7.9.1 Double-Polysilicon Self-Aligned Structures 510
 - 7.9.1.1 Limitations of Double-Polysilicon SA Structures.
 - 7.9.1.2 Current-Gain Degradation Due to Sidewall Injection in SA Bipolar Structures.
 - 7.9.1.3 Link-Up Region Formation.
- 7.9.2 Single-Polysilicon Self-Aligned Bipolar Structures 516
- 7.9.3 Sidewall-Base-Contact Structures (SICOS) 520

7.10 TRENCH-ISOLATED BIPOLAR TRANSISTORS 522

7.11 BICMOS TECHNOLOGY 523

- 7.11.1 Device and Circuit Advantages of BiCMOS 524
 - 7.11.1.1 Comparison of BiCMOS and CMOS Propagation Delay Times.
 - 7.11.1.2 Power Consumption of BiCMOS versus CMOS Gates.
 - 7.11.1.3 Capability of Providing Either TTL or ECL Outputs From a BiCMOS Chip.
 - 7.11.1.4 Process Complexity Increases Associated with BiCMOS.
 - 7.11.1.5 Extending Process Equipment Life by Fabricating BiCMOS.

7.12 CLASSIFICATION OF BICMOS TECHNOLOGIES 529

- 7.12.1 Digital BiCMOS Technology 531
 - 7.12.1.1 Low-Cost Digital BiCMOS Technology.
 - 7.12.1.2 High-Performance Digital BiCMOS.
 - 7.12.1.3 Device-Design Issues Related to Optimizing a High-Performance Digital Modified-Twin-Well BiCMOS Process.
 - 7.12.1.4 An Example Process Sequence for Fabricating High-Performance 5-V Digital BiCMOS ICs.
- 7.12.2 Process Integration of Analog/Digital BiCMOS 543
 - 7.12.2.1 Process-Integration Issues of Medium-Voltage Analog BiCMOS.
 - 7.12.2.2 An Example of an Analog/Digital BiCMOS Process.

7.12.3 BiCMOS Applications	551
7.12.3.1 Digital Logic Circuits and Gate Arrays.	
7.12.3.2 Interface Driver Circuits.	
7.12.3.3 BiCMOS SRAMs.	
7.12.3.4 Analog/Digital Applications.	
7.13 Trends in BiCMOS Technology	556
7.13 COMPLEMENTARY BIPOLAR (CB) TECHNOLOGY	557
REFERENCES	560
CHAP. 8 - SEMICONDUCTOR MEMORY PROCESS INTEGRATION	557
8.1 TERMINOLOGY OF SEMICONDUCTOR MEMORIES	557
8.1.1 Random-Access and Read-Only Memories (RAMs and ROMs)	568
8.1.2 Semiconductor-Memory Architecture	568
8.1.3 Semiconductor-Memory Types	570
8.1.4 Read Access Times and Cycle Times in Memories	571
8.1.5 Recently Introduced On-Chip Peripheral Circuits	571
8.1.6 Logic-Memory Circuits	571
8.2 STATIC RANDOM-ACCESS MEMORIES (SRAMS)	572
8.2.1 MOS SRAMs	575
8.2.1.1 Circuit Operation of MOS SRAM Cells.	
8.2.1.2 SRAM Processing and Cell Layout Issues.	
8.2.1.3. High-Valued Polysilicon Load-Resistors for MOS SRAMs	
8.2.2 Bipolar and BiCMOS SRAMS	584
8.2.2.1 BiCMOS SRAMs.	
8.3 DYNAMIC RANDOM ACCESS MEMORIES (DRAMs)	587
8.3.1 Evolution of DRAM Technology	587
8.3.1.1 One-Transistor DRAM Cell Design.	
8.3.1.2 Operation of the One-Transistor DRAM Cell.	
8.3.1.3 Writing, Reading, and Refreshing DRAM Cells.	
8.3.1.4 Quantity of Charge Stored on DRAM Cells and Their Capacitance.	
8.3.1.5 High-Capacity (Hi-C) DRAM Cells.	
8.3.1.6 CMOS DRAMs.	
8.3.2 Design and Economic Constraints on Advanced DRAM Cells	597

8.3.3 Trench Capacitor DRAM Cells	600
8.3.3.1 Trench Capacitor Processing for DRAMs.	
8.3.3.2 First Generation Trench Capacitor-based DRAM Cells.	
8.3.3.3 Trench Capacitor Structures with the Storage Electrode Inside the Trench (Inverted Trench Cell).	
8.3.3.4 Trench Capacitor Cells with the Access Transistor Stacked Above the Trench Capacitor.	
8.3.4 Stacked Capacitor DRAM Cells	609
8.3.5 Soft-Error Failures in DRAMs	615
8.3.5.1 Techniques Used to Reduce the Soft-Error Rates in DRAMs.	
8.3.6 The DRAM as a Technology Driver	618
8.4 MASKED READ-ONLY MEMORIES (ROMs)	619
8.4.1 Masked ROM Implementation	620
8.5 PROGRAMMABLE ROMS (PROMS)	621
8.6 ERASABLE PROGRAMMABLE READ-ONLY MEMORIES (EPROMs)	623
8.7 ELECTRICALLY-ERASABLE PROMS (EEPROMs)	628
8.7.1 MNOS-Based EEPROMs	628
8.7.2 FLOTOX EEPROMs	629
8.7.3 Textured-Polysilicon EEPROMs	631
8.8 FLASH EEPROMs	632
8.9 NONVOLATILE FERROELECTRIC MOS RAMS	635
REFERENCES	637
CHAP. 9 - PROCESS SIMULATION	643
9.1 OVERVIEW OF PROCESS SIMULATION	644
9.1.1 Hierarchy of Simulation Tools for IC Development	644
9.1.2 Benefits and Limitations of Process Simulation	645
9.1.3 Overview of Process Simulators	647
9.1.3.1 Simulator Availability.	
9.1.4 General Aspects of Process Simulation	650
9.1.4.1 Analytical and Numerical Methods of Solving the Equations that Describe Processes.	

9.1.4.2 Phenomenological versus Physical Models.	
9.1.4.3 Gridding.	
9.1.4.4 Interfacing One Simulator with Another.	
9.2 ONE-DIMENSIONAL PROCESS SIMULATORS	653
9.2.1 SUPREM III (Stanford University Process Engineering Model III)	655
9.2.1.1 The Basic Operation and Capabilities of SUPREM III.	
9.2.1.2 Additional Comments on the Use of SUPREM III.	
9.2.2 SUPREM III Models: Ion Implantation	658
9.2.3 SUPREM III Models: Diffusion in Silicon and SiO ₂ , and Segregation Effects at the Si/SiO ₂ Interface	663
9.2.3.1 Diffusion Models Used in SUPREM III.	
9.2.3.2 Modeling Low Impurity-Concentration (Intrinsic) Diffusion in Silicon.	
9.2.3.3 Modeling High-Impurity Concentration (Extrinsic) Diffusion in Silicon.	
9.2.3.4 Oxidation-Enhanced Diffusion Modeling in SUPREM III.	
9.2.3.5 Dopant Segregation Effects at the Si-SiO ₂ Interface and Diffusion in SiO ₂ .	
9.2.4 SUPREM III Models: Thermal Oxidation of Silicon in One-Dimension	669
9.2.4.1 High Dopant-Concentration Cases.	
9.2.4.2 Modeling Other Factors Which Impact the Oxide Growth Rate.	
9.2.4.3 Accuracy of Modeling Oxide Growth with SUPREM III.	
9.2.5 SUPREM III Models: Epitaxial Growth	674
9.2.6 SUPREM III Models: Deposition, Oxidation, and Material Properties of Polysilicon Films	675
9.2.7 Creating a SUPREM III Input File	677
9.2.8 PREDICT	679
9.3 INTRODUCTION TO 2-DIMENSIONAL PROCESS SIMULATORS	680
9.3.1 Classes of 2-Dimensional Process Simulators	683
9.4 TWO-DIMENSIONAL DOPING-PROFILE AND OXIDATION PROCESS SIMULATORS	684
9.4.1 SUPRA (Stanford University Process Analysis Program)	684
9.4.1.1 SUPRA Ion Implantation Models.	
9.4.1.2 SUPRA Diffusion Models.	
9.4.1.3 SUPRA Oxidation Models.	
9.4.1.4 SUPRA Epitaxial Model.	
9.4.1.5 SUPRA Input File.	
9.4.2 SUPREM IV	687

9.4.2.1 SUPREM IV Models of Diffusion.	
9.4.2.2 SUPREM IV Models of Oxidation.	
9.4.2.3 SUPREM IV Models of Ion Implantation, Epitaxy, Deposition, and Etching.	
9.4.2.4 SUPREM IV Input File Format.	
9.4.2.5 Comparison of SUPRA and SUPREM IV for 2-D Process Simulation.	
9.4.3 Two-Dimensional Simulation of Thermal Oxidation	690
9.4.3.1 Empirical Models of 2-D Thermal Oxidation.	
9.4.3.2 Physical-Based Models of 2-D Thermal Oxidation.	
9.5 TWO-DIMENSIONAL TOPOGRAPHY SIMULATORS	696
9.6 SAMPLE (SIMULATION AND MODELING OF PROFILES IN LITHOGRAPHY AND ETCHING)	697
9.6.1 Simulating Optical Lithography Processes with SAMPLE	697
9.6.1.1 Optical Imaging Subprogram.	
9.6.1.2 Resist Exposure Subprogram.	
9.6.1.3 Resist Development Subprogram.	
9.6.2 Simulating Etching and Deposition with SAMPLE	706
9.6.3 Creating Input Files for SAMPLE	708
9.7 OTHER 2-D TOPOGRAPHY SIMULATORS	710
9.7.1 PROLITH	710
9.7.2 DEPICT	710
9.7.3 PROFILE	711
9.7.4 SIMBAD	713
9.7.5 SIMPL (Simulated Programs from the Layout)	714
9.7.6 SIMPL-DIX	716
9.7.7 Manufacturing-Based Process Simulators	718
9.8 DEVICE SIMULATORS	718
9.8.1 Simulation of MOS Device Characteristics under Subthreshold and Linear Operation (GEMINI)	719
9.8.2 Simulation of MOS Device Under All dc Operating Conditions (MINIMOS, CADDET, CANDE)	719
9.8.3 Bipolar Device Simulators (SEDAN, BIPOLE)	720
9.8.4 Combined MOS and Bipolar Device Simulators (PICSES, SIFCOD, PADRE, and FIELDAY)	721

9.9 CIRCUIT SIMULATORS AND ELECTRICAL PARAMETER EXTRACTORS	723
9.10 FUTURE CHALLENGES IN PROCESS SIMULATION	723
REFERENCES	724
APPENDIX A IC RESISTOR FABRICATION	731
APPENDIX B PROPERTIES OF SILICON AT 300 °K	737
APPENDIX C PHYSICAL CONSTANTS	738
INDEX	739

LIST OF TECHNICAL REVIEWERS

Each of the chapters was reviewed for technical correctness. The following persons graciously undertook the review task for the chapters indicated:

Chapter 2	Dr. Joseph R. Monkowski Lam Research Corp. - CVD Division Fremont, CA	Dr. Haiping Dun Intel Corp. Santa Clara, CA
Chapter 3	Dr. Robert S. Blewer Sandia National Laboratories Albuquerque, NM	Dr. Stan Swirhun Honeywell SSPL Bloomington, MN
Chapter 4	Dr. Farhad K. Moghadam Intel Corp. Santa Clara, CA	Dr. Terry Herndon MIT - Lincoln Laboratory Lexington, MA
Chapter 5	Mr. Andrew R. Coulson TRW Electronic Systems Group Redondo Beach, CA	
Chapter 6	Dr. John Y. Chen Boeing Electronics Seattle, WA	Dr. Samuel T. Wang International CMOS Technology, Inc. San Jose, CA
Chapter 8	Professor Al F. Tasch, Jr. University of Texas Austin, TX	
Chapter 9	Dr. Michael Kump Technology Modeling Associates, Inc. Palo Alto, CA	

PREFACE

SILICON PROCESSING FOR THE VLSI ERA is a text designed to provide a comprehensive and up-to-date treatment of this important and rapidly changing field. The text will consist of three volumes, of which this book is the second, subtitled, *Process Integration*. Volume 1, subtitled *Process Technology*, was published in 1986. Volume 3, to be subtitled *Assembly, Packaging and Manufacturing Technology* is scheduled for publication in 1993. In Volume 1, the individual processes utilized in the fabrication of silicon VLSI circuits are covered in depth (e.g., epitaxial growth, chemical vapor and physical vapor deposition of amorphous and polycrystalline films, thermal oxidation of silicon, diffusion, ion implantation, microlithography, and etching processes).

In this volume, we undertake to explain how the individual processes described in Volume 1 are combined in various ways to produce silicon integrated circuits. This task is referred to as *process integration*. The first part of the book deals with *sub-process integration*; that is, the effort involved in forming circuit structures that can be implemented into a variety of circuit types. These structures include *isolation structures* (Chap. 2), *metal-silicon contacts* (Chap. 3), and *device-interconnect structures* (Chap. 4).

The second part of the book covers the process integration tasks of full-device-type technologies, including *NMOS* (Chap. 5), *CMOS* (Chap. 6), *bipolar* and *BiCMOS* (Chap. 7), and *semiconductor memories* (Chap. 8). Chapter 9 describes the process simulation tools that are available for aiding in the process integration and development efforts.

Volume 3 will cover process control, VLSI manufacturing issues and facilities, contamination and yield, automation, assembly, packaging, and parametric testing.

The purpose of writing this text was to provide professionals involved in the microelectronics industry with a single source that offers a complete overview of the technology associated with the manufacture of silicon integrated circuits. Other texts on the subject are available only in the form of specialized books (i.e., that treat just a small subset of all of the processes), or in the form of edited volumes (i.e., books in which a group of authors each contributes a small portion of the contents).

Such edited volumes typically suffer from a lack of unity in the presented material from chapter-to-chapter, as well as an unevenness in writing style and level of presentation. In addition, in multi-disciplinary fields, such as microelectronic fabrication, it is difficult for most readers to follow technical arguments in such books.

especially if the information is presented without defining each technical "buzzword" as it is first introduced.

In our books such drawbacks are avoided by treating the subject of VLSI fabrication from a unified and more pedagogical viewpoint, and by carefully defining technical terms when they are first introduced. The result is intended to be a user friendly book for workers who have come to the semiconductor industry after having been trained in but one of the many traditional technical disciplines.

An important technical breakthrough has occurred in publishing that the author also felt could be exploited in creating a unique book on silicon processing. That is, revolutionary electronic publishing techniques have recently become available, which can cut the time required to produce a published book from a finished manuscript. This task traditionally took 15-18 months, but can be now reduced to less than 3 months. If traditional techniques are used to produce books in such fast-breaking fields as VLSI fabrication, these books automatically possess a built-in obsolescence, even upon being first published.

We have taken advantage of these rapid production techniques, and have been able to successfully meet the reduced production-time schedule. As a result, information contained in technical journals and conferences which was available within three months of the book's publication date has been included.

Written for the professional, the book belongs on the bookshelf of workers in several microelectronic disciplines. Microelectronic fabrication engineers who seek to develop a more complete perspective of the subject, or who are new to the field, will find it invaluable. Integrated circuit designers, test engineers, and integrated circuit equipment designers, who must understand VLSI processing issues to effectively interface with the fabrication environment, will also find it a uniquely useful reference. The book should also be very suitable as a text for graduate-level courses on silicon processing techniques, offered to students of electrical engineering, applied physics, and materials science. It is assumed that such students already possess a basic familiarity with semiconductor device physics. Problems are included at the end of each chapter to assist readers in gauging how well they have assimilated the material in the text.

The book is an outgrowth of several intensive seminars conducted by the author through the Engineering Extension of the University of California, Irvine. Over one thousand engineers and managers from more than 75 companies and government agencies have enrolled in these short courses since they were first presented in 1984.

A book of this length and diversity would not have been possible without the indirect and direct assistance of many other workers. To begin, virtually all of the information presented in this text is based on the research efforts of a countless number 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. 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 authors express heartfelt thanks to those who gave of their time, energy, and intellect.

Each of the chapters was reviewed after the writing was completed. The engineers and scientists who participated in this review were numerous. The main reviewers are

listed on the page before the preface, and we would like to thank them once again at this point for their contributions. That is, the following professionals (each one an expert in the topic covered by the chapter they reviewed), read an entire chapter for technical correctness, and provided appropriate comments and corrections: Robert S. Blewer, John Y. Chen, Andrew R. Coulson, Haiping Dun, Terry Herndon, Michael Kump, Farhad K. Moghadam, Joseph R. Monkowski, Stan Swirhun, Al F. Tasch, Jr., and Samuel T. Wang. In addition, J. B. Price, of Spectrum CVD, Chris A. Mack, of the National Security Agency, and Sidney Marshall, Editor of Solid State Technology, provided other valuable technical and editorial input. Robert Shier, of VTC Corp., and Jim Cable of TRW also kindly provided the author with timely and valuable technical literature.

The copy editing of the book was undertaken by Mary Nadler, and the clarity and grammatical correctness of the prose owes a great deal to her professional efforts. The aesthetically pleasing graphics of the cover were designed by Roy Montibon of Visionary Art Resources, Inc., Santa Ana, CA.

Stanley Wolf

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

Lattice Press
P.O. Box 340-W
Sunset Beach, CA, 90742

An order form, for your convenience, is provided on the final leaf of the book.

CHAPTER 1

PROCESS INTEGRATION

FOR VLSI and ULSI

Since the creation of the first integrated circuit in 1960, the density of devices that can be fabricated on semiconductor substrates has steadily increased. In the late 1970s the number of devices manufactured on a chip exceeded the generally accepted definition of *very large scale integration*, or *VLSI* (i.e., more than 100,000 devices per chip) (Fig. 1-1a). By 1990 this number had grown to more than 32 million devices per chip (16-Mbit DRAMs), and it is generally acknowledged that the era of *ultra-large-scale integration* (*ULSI*) has begun. The increasing device count has been accompanied by a shrinking minimum feature size (Fig. 1-1b), which as of 1990 had decreased to $\sim 0.5 \mu\text{m}$ in the most advanced commercially available chips.

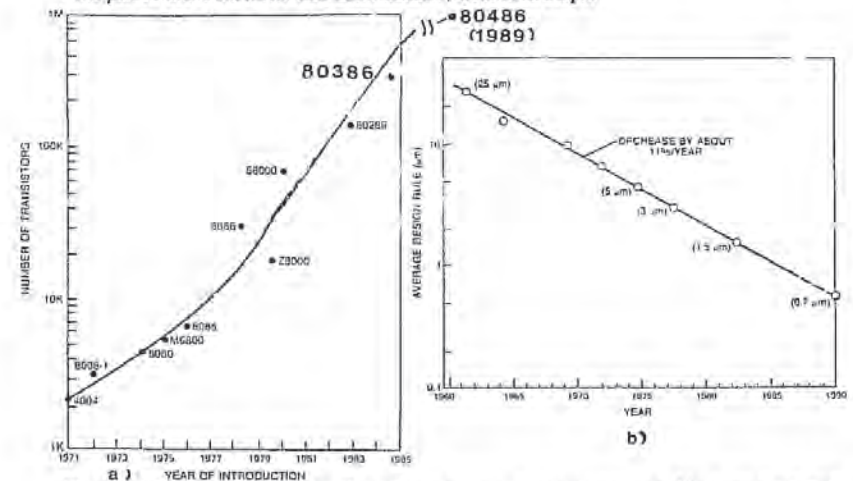


Fig. 1-1 (a) Increase in the number of transistors per microprocessor chip versus year of introduction, for a variety of 8-bit and 16-bit microprocessors, and (b) The decrease in minimum device feature size versus time on integrated circuits.

Progress seems likely to continue at a rapid pace, with even further reductions in the unit cost per function and in the power-delay product of integrated circuits projected. Silicon processing has been the dominant technology of IC fabrication and is likely to retain this position for the foreseeable future. The entire adventure of silicon device manufacturing represents a remarkable application of scientific knowledge to the requirements of technology. Our books* are intended to serve as a comprehensive and cohesive report on the state of the art of this technology, as practiced at the time of publication.

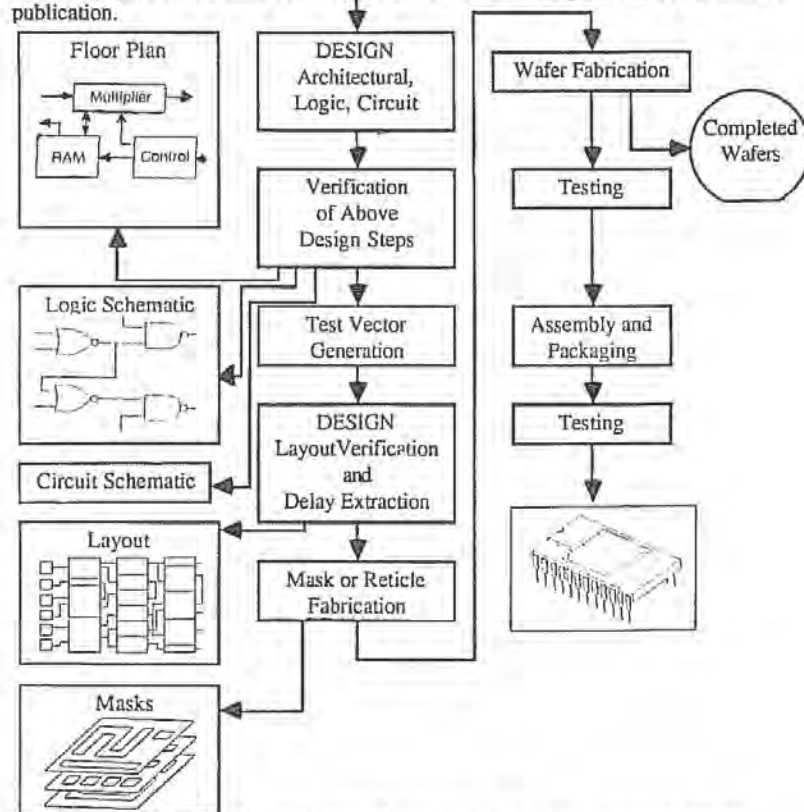


Fig. 1-2 Steps required for the manufacture of very large scale integrated circuits (VLSI).

* SILICON PROCESSING FOR THE VLSI ERA - Volume 1: Process Technology; Volume 2 - Process Integration; and Volume 3 - Assembly, Packaging, and Manufacturing Technology (the latter is scheduled to be published approximately in 1992).

Figure 1-2 illustrates the sequence of tasks followed in the manufacture of an integrated circuit. These tasks can be grouped into three phases: *design*, *fabrication*, and *testing*. Our books are concerned primarily with the fabrication phase. We describe the IC manufacturing steps that occur from the point at which the circuit design has been completed and the necessary design information has been rendered into the form of a circuit layout. In this form, the layout information is ready to be used to generate a set of masks (or reticles) that will serve as the tools for specifying the circuit patterns on silicon wafers. For VLSI and ULSI circuits, the layout information is stored in a computer.

The details concerning the individual fabrication processes (e.g., those associated with creating patterns, introducing dopants, and depositing films on silicon substrates to form integrated circuits) are the subjects of Volume 1. In this second volume, we

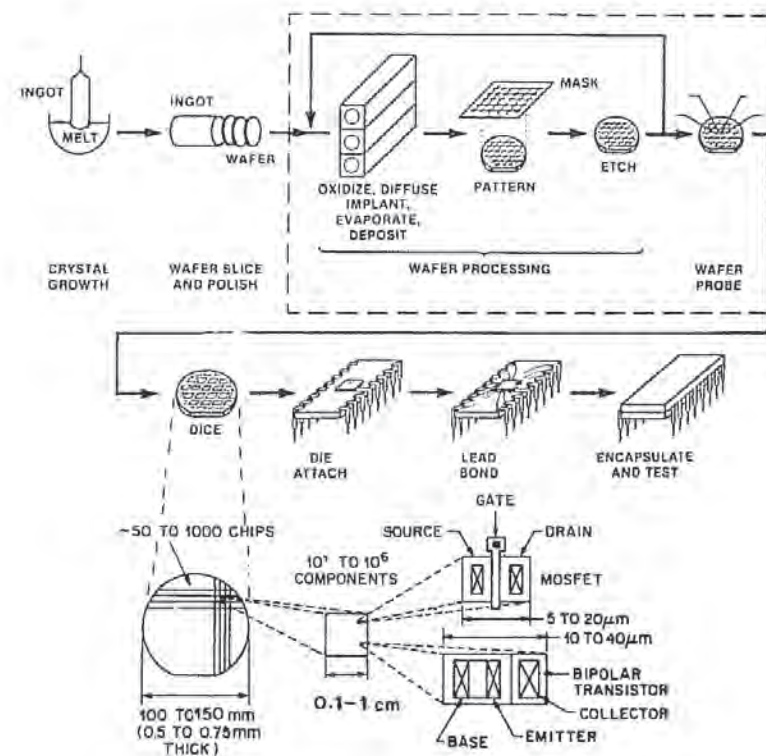


Fig. 1-3 The fabrication process sequence of integrated circuits.

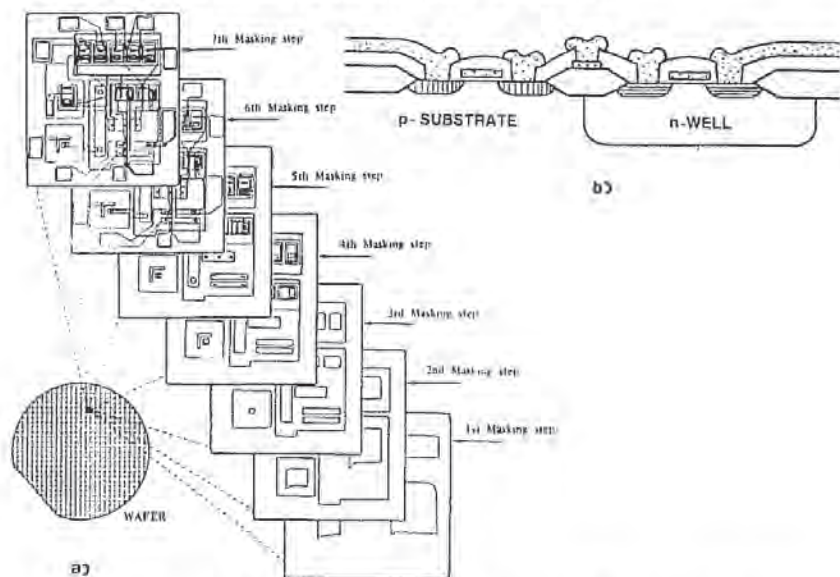


Fig. 1-4 (a) Example of the patterns transferred to a wafer during a seven-mask process sequence, and (b) Cross section of completed devices in a basic CMOS process.

undertake a discussion of the *sequences* of the steps performed to produce the device structures of the circuits. The text describes how the various individual processes are integrated together so that the end result is a completely realized integrated circuit.

One perspective of such a *process-integration* effort is illustrated in Fig. 1-4. It can be seen that a series of masking steps must be sequentially performed for the desired patterns to be created on the wafer surface. The other processing procedures performed between the masking steps serve to create the desired device structures. An example of the end result of a complete process sequence is shown in Fig. 1-4b as an IC cross-section.

This book is roughly divided into two parts. The first part deals with what we refer to as *subprocess integration* (Chapters 2, 3, and 4), while the second covers *complete process integration* (Chapters 5, 6, 7, and 8). Chapter 9 covers the subject of *process simulation*, an approach applicable to both the subprocess and complete process-integration development efforts.

The function of subprocess integration is to produce device structures that can be used in a variety of integrated-circuit technologies (i.e., isolation, contact, and interconnect structures). The purpose of the process-integration task is to design the process sequences used to manufacture complete IC technologies (i.e., NMOS, CMOS, bipolar, BICMOS, and IC memory devices).

1.1 PROCESS INTEGRATION

The approach used in building integrated circuits on monolithic pieces of silicon involves the fabrication of successive layers of insulating, conducting, and semiconducting materials. Each layer is patterned to form a structure that performs a specific function, usually linked with surrounding areas and subsequent layers. The fabrication steps used to manufacture an integrated circuit must therefore be executed in a specific sequence, which constitutes an IC *process flow* (or *process*).

1.1.1 Process Sequence Used to Fabricate an Integrated-Circuit MOS Capacitor Structure. A simple example of the sequence of steps that must be used to form an integrated circuit MOS capacitor is shown in Fig. 1-5.³ First, a relatively thick SiO_2 layer is grown on a bare *n*-type Si substrate (using oxide-growth techniques, as described in Vol. 1, chap. 7). This oxide layer (Fig. 1-5a) serves to isolate the metallization-interconnect film deposited at the end of the process from the Si substrate.

An opening in the SiO_2 layer is then formed using a photolithographic process (Vol. 1, chaps. 12 and 13) and an etch process (see Vol. 1, chaps. 15 and 16). Boron atoms are ion-implanted into the wafer surface (Fig. 1-5b). Wherever an opening exists in the SiO_2 layer, the boron atoms are implanted into the Si substrate. A sufficient

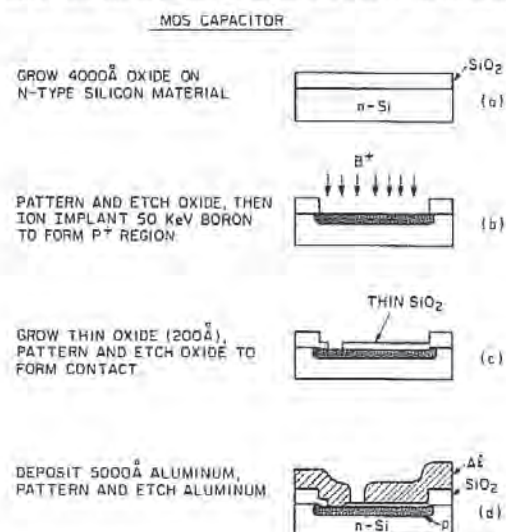


Fig. 1-5 Process sequence for fabrication an MOS capacitor. From S. M. Sze, Ed., *VLSI Technology*, 2nd Ed., Copyright 1988 McGraw-Hill. Reprinted with permission.

implant dose is used to cause these regions to become *p*-type. The energy of the implant and the SiO₂ thickness are also selected so that the boron atoms do not penetrate to the Si wherever the SiO₂ layer remains. In this manner, *p*-regions (which will serve as the lower plate of the MOS capacitor) are selectively formed in the Si.* In addition, if the voltage between the *p* and *n* regions of the Si substrate keeps the *pn* junctions reverse-biased, the lower plates of any capacitors built on the same Si substrate will be *junction-isolated* from one another (see chap. 2).

The remaining steps are shown in Figs. 1-5c and 1-5d. A thin SiO₂ film is grown over the exposed *p*-type Si region, and a *contact hole* is etched in the film to allow an Al film to make contact to the lower plate of the capacitor. The Al film, deposited next, also serves as the top plate of the capacitor. The formation of the contact holes and Al patterns involves alignment, another important process that must be successfully accomplished when integrating a process sequence. The capacitor structure will function properly only if the two plates of the capacitor are isolated from each other, except where an intentional contact must be made. Therefore, if the Al pattern that serves as the top plate of the capacitor is misaligned (resulting in an accidental contact between the Al and the *p*-type Si region), the two plates of the capacitor will be shorted. This alignment restriction is one of the limitations faced in device miniaturization, as discussed in Volume 1, chapter 13.

1.1.2 Specifying a Process Sequence. Once a process has been completely developed, a set of instructions is produced for fabrication of the specific technology. These instructions include the appropriate times, temperatures, gas flows, power settings, etc. needed for each process step in order for the desired structures to be produced. The instructions also list the specifications of the device structures that should be achieved at the conclusion of each of the process steps (e.g., line width, sheet resistance, film thickness, and step height), as well as the inspections and measurements that must be made to ensure that these specifications have been met. This set of instructions accompanies the group of wafers (or *lot*) as it moves through the process, and is therefore known as a *lot traveler*, or *run sheet*.

Since integrated circuits are normally processed in *batches* (or *lots*), each of the steps is usually performed on all of the wafers in a lot before that lot is moved to the next processing step. As each step is completed, the fabrication worker (*operator*) responsible for that step indicates that the process step has been accomplished and records on the *lot traveler* the time of completion, as well as any required data concerning the process conditions, resultant measured parametric data, and any pertinent comments. When a lot has been completed, the data gathered about all of the fabrication details of that *process run* is thus contained on the lot traveller. Lot travellers initially consisted of printed

* Note that this region is also *self-aligned* to the edges of the SiO₂ layer. This is an example of the self-alignment process, a fabrication approach that is critical for allowing progress to be made in device miniaturization techniques.

sheets of paper, but they are now stored on computers. Data concerning the manufacturing details of each lot-fabrication sequence is entered directly into the computer.

1.1.3 Levels of Process Integration Tasks. From a general perspective, process integration involves the combining of various individual processes (described in Vol. 1) to produce IC structures (or even an entire IC). Process integration tasks can thus involve varying degrees of complexity. Examples of the levels at which such tasks can be performed can be grouped according to increasing complexity, as follows:

1. *Development of a process sequence to modify an existing structure that is only one part of an integrated circuit.* Two examples of such process-integration tasks are: (a) the development of a modified local-oxidation isolation structure with reduced bird's beak or boron channel-stop encroachment (see chap. 2), and (b) the improvement of the characteristics of a barrier-layer material used in metal-Si contacts (e.g., by replacing Ti with Ti:W; see chap. 3).

2. *Development of a process sequence to produce a new device structure.* Three examples of this are: (a) development of a selective epitaxial isolation structure to replace LOCOS (see chap. 2); (b) implementation of a double-level-metal (DLM) structure to replace a single-level-metal interconnect approach (see chap. 4); and (c) creation of a TiN local-interconnect structure for CMOS (chap. 3).

3. *Enhancement of a complete process sequence by shrinking the minimum feature size used in the technology.* While such "technology-shrinks" are generally less complex than the tasks involved in the development of an entirely-new complete-process sequence, they usually entail more than just the implementation of higher-resolution lithography processes. That is, many of the other processes must also be modified, and entirely new ones may need to be developed. Two examples of such technology evolution that have been detailed in the literature are the shrinking of Intel's NMOS process from NMOS I ($L = 6 \mu\text{m}$) to HMOS III ($L = 1.5 \mu\text{m}$), and the shrinking of AT&T's Twin-Tub CMOS process from CMOS I ($L = 2 \mu\text{m}$) to CMOS VI ($L = 0.5 \mu\text{m}$).

4. *Implementation of a new technology through modification or enhancement of an older technology.* Two examples of this type of process-integration effort are:

a. Development of an *n*-well CMOS process based on the experience gained from prior NMOS technology development. In this case, NMOS devices are fabricated in the *p*-substrate, using the same process sequence as in an NMOS process developed earlier, while the fabrication of PMOS devices in the *n*-well (and the integration of these process steps into the NMOS sequence) represent new process integration tasks.

b. Development of a BiCMOS process through integration of the steps

needed to fabricate bipolar *npn* transistors into an existing CMOS process sequence (see chap. 7).

5. Development of a new technology from scratch. Examples of this effort are the development of new CMOS and BiCMOS processes without reliance on the enhancement of older process sequences. This kind of effort is normally the most difficult, and so it is rarely attempted. When a new process is to be developed from scratch (especially where new materials are to be used), the effort required to establish the interrelationships among the various process steps can be extremely expensive and time consuming. A more conservative approach is to modify or enhance a previously developed process. Consequently, when a company decides to launch a line of products based on a new process sequence, it often purchases the technology details from a company that has already developed such a process. The acquired process technology can either be used as is, or enhanced.

1.2 PROCESS-DEVELOPMENT AND PROCESS-INTEGRATION ISSUES

When a new process or device structure is proposed for development, the intent is to improve the circuit performance in some way or to build new products that could not be fabricated with existing technology. It is very important that the circuit designers, product engineers, device designers, and process engineers interact closely in such a development effort. The circuit designers and product engineers identify the performance requirements and package sizes needed for the applications to be served (e.g., SRAMs, ASICs, microprocessors, or logic circuits). From these specifications, the device designers can target candidate device structures, as well as the various structural options and process flows that can be potentially investigated for achieving these goals. The process engineers can then work on finding new materials and innovative processes to implement the advanced device structures. The odds of developing new device structures that offer improved performance (while still meeting reliability and manufacturability constraints) are greatly enhanced when each of the members of a technology development team understands the goals and limitations of the other disciplines involved.

Several key issues must be considered when a process-integration task is to be undertaken. One of the most important is that virtually all of the steps in a process sequence are significantly interrelated. For example, each of the thermal cycles in a process sequence contributes to the total vertical and lateral diffusion effects in a process. To end up with device structures that have the desired doping profiles at the end of a process sequence, it is necessary to take into account *all* of the thermal processing steps that a device undergoes. Furthermore, each new thermal step may impact the formation of defects in the silicon or may play a role in the chemical or

morphological characteristics of the structures created in the IC. Since the fabrication steps are so strongly interrelated, once a process has been well established, a cardinal rule – which must be rigidly followed – is that *no step in the process may be arbitrarily changed*.

Nevertheless, when a new process is developed, it is obvious that at least some changes must be made to an existing process. The ability to predict the effect of a process change on the circuit or device electrical parameters is an important part of the process-integration effort. In the early days of integrated-circuit manufacturing, hand calculations of the doping profiles in the silicon and SiO₂ based on simple one-dimensional analytical models were used to estimate the electrical characteristics of device structures fabricated using a particular process sequence. These predicted estimates were then compared to the electrical characteristics measured following the completion of process runs which were designed according to the analytically determined process steps (trial-and-error approach).

As device sizes grew smaller, however, such an approach proved to be inadequate. That is, two-dimensional effects within the device structures made the measured electrical characteristics deviate significantly from the characteristics predicted with one-dimensional models. To produce small-geometry devices with the desired electrical characteristics, device and process engineers had two options: (1) to perform more experiments, or (2) to use computer-aided simulation of the fabrication process to model the device structures that are obtained from a specific process sequence (Fig. 1-6).¹

The first option would mean a drastic increase in the costs and cycle time of new process integration. Furthermore, even if the experiments did yield adequately performing devices, if complicated processes and structures were involved, the result would still be poor physical insight and inadequate quantitative analysis of the factors governing device operation.

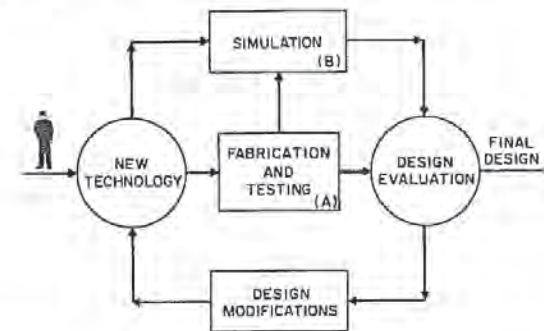


Fig. 1-6 Simulation and experiment are approaches to IC process development. Path A depicts the experimental approach. Path B is the approach that exploits computer-aided simulation.¹ (Copyright IEEE, 1984).

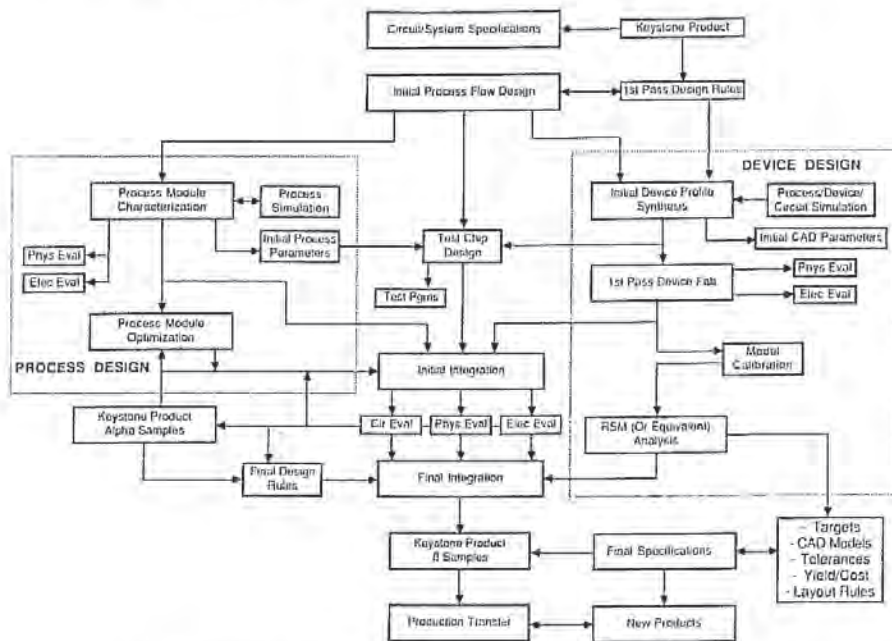


Fig. 1-7 Technology development flowchart, integrating process, device, and design considerations.² (Copyright IEEE, 1987).

The second approach has therefore become more widely adopted. The information about the device structures calculated from the process simulation (in the form of impurity profiles, electrical characteristics, and layer and pattern dimensions) can be used with circuit models of the devices to calculate the device's terminal characteristics. The circuit parameters (e.g., SPICE parameters) can then be extracted from the terminal characteristics and the device layouts. These parameters, together with the information calculated on the connectivity characteristics of the circuit (extracted from circuit-layout and process-simulation data on the interconnect structures of the circuit), can yield the frequency-response or switching characteristics of the resultant circuits. The simulated design can be verified by building circuits based on the results once a satisfactory set of circuit-performance parameters has been obtained. In addition to being far less costly and time consuming than laboratory experiments, this design approach produces detailed information on device operation in a well-controlled environment. The simulation tools available for process integration as described in chapter 9.

A comprehensive approach for developing new IC technologies has been outlined by Alvarez et al.² and is summarized in the flow-chart shown in Fig. 1.7. It can be seen that the process integration task is but one part of this larger technology development effort. Other important tasks that must also be successfully performed include device design, module development (i.e., development of each of the individual steps in the process sequence), specification of the circuit/system characteristics that the technology will be serving, and test-chip design.

Reference 2 also addresses the impact of the conflicting requirements that arise in the attempt to optimize all of the device characteristics (here, for the specific case of a BiCMOS technology). Statistical design of experiments (see Vol. 1, chap. 18) and response surface methodology (RSM) are used to identify the effect of tradeoffs between the various device characteristics on device structures. Process and device simulators (described in chap. 9) are used to conduct the analyses that identify the optimum device structures for the applications to be served. Process conditions can then be selected to allow such optimum device structures to be fabricated.

REFERENCES

1. W. Fichtner et al., *Proceedings IEEE*, 72, 96, (1984).
2. A. R. Alvarez, "BiCMOS Technology," 1987 IEDM Short Course on BiCMOS Technology, December 1987.
3. S. J. Hillenius, "VLSI Process Integration," Chap. 11, in *VLSI Technology*, 2nd. Ed., S.M. Sze, Editor, McGraw-Hill Book Company, New York, 1988.

PROBLEMS

1.1 Using the MOS device structure illustrated in Fig. 3-38, and assuming that a worst-case misalignment will be one-half the distance of the minimum feature size (e.g., the length of the contact window), draw a composite view of the situation that will result from a worst-case misalignment of the masks. Assume that the metal mask aligns to the gate oxide mask, and the gate oxide and contact masks align to the diffusion mask.

1.2 Design a mask set for a 3-pF MOS capacitor. Assume the oxide is 15 nm thick. Assume that the minimum window size is $1 \times 5 \mu\text{m}$ and the maximum registration errors are $1 \mu\text{m}$.

1.3 Consider the process steps required to fabricate a simple diffused resistor consisting of a long, rectangular p -type region in an n -type substrate with contacts formed on either end (see Appendix A).

- (a) What are the minimum masking steps needed to form the resistor?
- (b) What would the dimensions (length, L , and width, W) of the resistor have to be if the sheet resistance of the p -region is $50 \Omega/\text{sq}$ and the minimum feature size is $1 \mu\text{m}$?
- (c) What are the minimum masking steps needed to form a pinched diffused resistor?

CHAPTER 2

ISOLATION TECHNOLOGIES FOR INTEGRATED CIRCUITS

Implementing electric circuits involves connecting *isolated* devices through specific electrical paths. When fabricating silicon integrated circuits it must therefore be possible to isolate devices built into the silicon from one another. These devices can subsequently be interconnected to create the specific circuit configurations desired. From this perspective, we can see that isolation technology is one of the critical aspects of fabricating integrated circuits.

A variety of techniques have been developed to isolate devices in integrated circuits. One reason is that different IC types (e.g., NMOS, CMOS, and bipolar) have somewhat different isolation requirements. Furthermore, the various isolation technologies exhibit differing attributes, with respect to minimum isolation spacing, surface planarity, process complexity, and density of defects generated during fabrication of the isolation structure. Tradeoffs can be made among these characteristics when selecting an isolation technology for a particular circuit application. This chapter surveys the various isolation technologies, including their evolution up to those being evaluated for submicron devices.

Before the invention of integrated circuits, only discrete diodes, bipolar transistors and field-effect transistors (FETs) could be fabricated. In the early 1950s, these discrete devices exhibited relatively high reverse-bias junction leakages and low breakdown voltages (caused by the large density of traps at the surface of single-crystal silicon).

In 1958, a group of workers at Bell Telephone Laboratories, led by Atalla, found that when a thin layer of SiO_2 was grown on the surface of silicon where a *pn* junction intercepts the surface, the leakage current of the junction was reduced by a factor from 10 to 100. It was later understood that the oxide reduces and stabilizes many of the interface and oxide traps. Not only did such oxide-passivation of the silicon surfaces allow diodes and transistors to be fabricated with significantly improved device characteristics, but the leakage path along the surface of the silicon was also effectively shut off. Thus, one of the fundamental isolation capabilities needed for planar devices and integrated circuits had also been developed. In his report on the evolution of the MOS transistor, C. T. Sah remarks that the successful effort by the Bell Labs group to stabilize Si surfaces was the most important technological advance in microelectronics

during the 1950s, and that it blazed the trail that led to the development of the silicon integrated circuit.⁹⁰

With the advent of integrated circuits, it became necessary to provide electrical isolation between devices fabricated on the same piece of silicon. Since bipolar ICs were the first to be developed, a technology for isolating the collector regions of the bipolar devices was also the first to be invented (termed *junction isolation*).

PMOS and NMOS ICs did not need junction isolation, but it was still necessary to provide an isolation structure that would prevent the establishment of parasitic channels between adjacent devices. The most important technique developed was termed *LOCOS isolation* (for LOCal Oxidation of Silicon), which involved the formation of a semirecessed oxide in the nonactive (or *field*) areas of the substrate.

Eventually, bipolar ICs adopted a similar LOCOS-isolation technology (with the oxide performing a somewhat different function than in MOS circuits). The nature of CMOS also required that isolation exist between devices in adjacent tubs as well as between devices within each tub.

As device geometries reached submicron size, conventional LOCOS isolation technologies reached the limits of their effectiveness, and alternative isolation processes for CMOS and bipolar technologies were needed. Modified LOCOS processes (which overcome some drawbacks of conventional LOCOS for small-geometry devices), trench isolation, and selective-epitaxial isolation, were among the newer approaches adopted.

Devices that must function under high voltages and in harsh radiation environments require even more stringent isolation techniques. These techniques are generally referred to as *silicon-on-insulator* (SOI) isolation methods. They include such older approaches as *dielectric isolation* (DI) and *silicon-on-sapphire* (SOS), as well as more recently developed technologies, such as *separation by implanted oxygen* (SIMOX), *zone-melting-recrystallization* (ZMR), *full isolation by porous-oxidized silicon* (FIPOS), and *wafer bonding*. These technologies will be described in the final sections of the chapter.

2.1 BASIC ISOLATION PROCESSES FOR BIPOLAR ICs

2.1.1 Junction Isolation

As mentioned in the introduction, isolation techniques for bipolar ICs were the first to be developed. To see why isolation is needed in bipolar ICs consider Fig. 2-1a, which shows several non-isolated bipolar transistors on a monolithic substrate. It can be seen that all the collector regions are electrically connected through the substrate - obviously, an unacceptable circumstance for all bipolar IC structures (except those few in which common collectors are desired). To circumvent this problem, several approaches to isolating the devices have been developed.

The most important isolation technique used in early bipolar ICs is *junction isolation*, and it has been incorporated as part of the *standard buried collector* (SBC) process (described in this section as well as in chap. 7), the *triple diffused* (3D) process

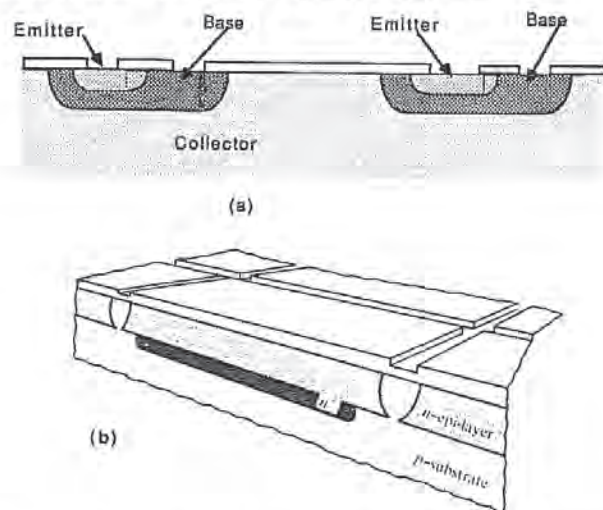


Fig. 2-1 (a) Two non-isolated bipolar transistors on a monolithic substrate. It can be seen that all the collector regions are electrically connected through the substrate - obviously, an unacceptable circumstance for all bipolar IC structures (except those few in which common collectors are desired). (b) A p -type "moat" is diffused into an n -type epitaxial layer on a p -type substrate. The p diffusion extends through the epitaxial layer, joining with the p -type substrate to completely surround the n -type epi islands.

(described further in chap. 7), and the collector diffused process (described in this section).

2.1.1.1 Junction Isolation in the SBC Process. The SBC process is still widely implemented in bipolar analog-circuit and power-circuit applications where power-supply voltages are in the 15-30 V range. Most digital bipolar technologies (which operate with supply voltages smaller than 10 V), however, have come to adopt oxide-isolated structures.

When junction isolation is used for npn devices, a p -type "moat" is diffused into an n -type epitaxial layer on a p -type substrate (Fig. 2-1b). The p diffusion extends through the epitaxial layer, joining with the p -type substrate to completely surround the n -type epi islands. (The n -type epitaxial layer is typically less than $10\text{-}\mu\text{m}$ thick, which is thick enough to accommodate the subsequent base and emitter regions of the bipolar transistors.) Hence, each device is contained in an island of n -type material surrounded by a pn junction; good isolation exists if the substrate p material is held at the most negative potential of the circuit. In effect, each component is surrounded by a reverse-biased pn junction, and room-temperature leakage current for reverse-biased silicon

junctions can be held within the nanoampere range. With the exception of capacitance effects at high frequencies, excellent isolation between circuit elements can be provided.

The isolation structure is formed after the n^+ buried layer and the epitaxial layer have been grown in the p -type substrate (see chap. 7 for more details on bipolar device structures). An oxide is grown on the epi layer and is then patterned and etched. Alignment of the isolation diffusion now becomes an issue. That is, the isolation structure should not touch the buried layer because not only would the high doping on both sides of the resulting pn junction drastically decrease the breakdown voltage, but the capacitance of the collector/substrate junction would also increase. Thus, a minimum spacing between the n^+ buried layer and the isolation diffusion must be specified among the layout design rules. This dimension is determined by taking into account the lateral diffusion distances of both the isolation diffusion and buried layer, with an allowance made for alignment tolerance (Fig. 2-2).

Because of the depth to which the p -region must be driven, the boron isolation diffusion typically takes the longest of any furnace step in the bipolar process. During this step, the lateral diffusion of the boron atoms is about 75 to 80% of the vertical diffusion, and thus the isolation region width is approximately twice the thickness of the epi-layer. Following diffusion, a low-temperature oxidation-and-etch sequence is performed to remove the boron-rich layer on the silicon surface, and a new oxide layer is grown over the isolation diffusion.

Successful fabrication of the isolation structures can be confirmed by stripping the oxide over the isolation diffusion and probing two adjacent structures. If complete isolation has been achieved, no current will flow unless the applied voltage exceeds the breakdown voltage of the isolation junction. If a small gap exists at the bottom of the isolation, leakage current will occur. Poor isolation results in catastrophic device failure since all transistor collector regions are shorted together.

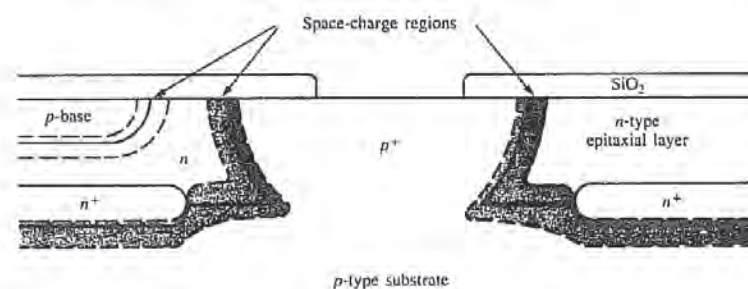


Fig. 2-2 Isolation region between two bipolar transistors. The spacing must be large enough to ensure that the two space-charge regions do not merge together. From R. C. Jaeger, *Introduction to Microelectronic Fabrication*. Copyright, 1988, Addison-Wesley. Reprinted with permission.

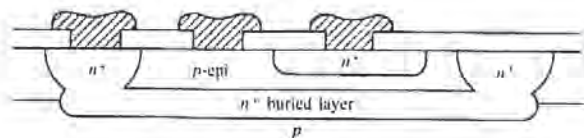


Fig. 2-3 Cross section of a transistor fabricated in the CDI process.¹²⁹ (© 1969 IEEE).

Junction-isolation, while quite simple, exhibits two major drawbacks. First, the wide isolation regions represent large inactive areas on the silicon surface, and this adversely impacts IC packing-densities. Second, the isolation diffusions cause large parasitic collector-to-substrate and collector-to-base capacitance values, which degrade the speed performance of the circuits (see chap. 7).

2.1.1.2 Collector-Diffusion Isolation. The collector-diffusion isolation (CDI) structure¹ allows isolation between adjacent bipolar devices to be achieved without the use of a p -type isolation diffusion (Fig. 2-3). This reduces device area as well as process complexity, as compared to the SBC process. The CDI process is used to best advantage in digital applications in which high speed is an important priority and low power-supply voltages are utilized (e.g., +5 V, as used in digital circuits).

The process is begun by selectively diffusing low-sheet-resistance (15-25 Ω/sq) arsenic buried layers 3-5 μm deep, onto a lightly doped p -type wafer. These buried layers act as the collectors of the transistor and as isolation regions for the diffused resistors. This is followed by formation of a thin p -type (0.2 $\Omega\text{-cm}$) epitaxial layer (1-2 μm thick) which serves as a part of the base. Device isolation and definition of the base are accomplished through an n^+ phosphorus diffusion, which surrounds the transistor peripherally and also provides contact to the collector layer. A p -diffusion is then carried out to form the remainder of the base region. Since the surface of the n^+ regions are heavily doped, this p -diffusion can be done non-selectively, without the need for another mask. This diffusion prevents surface inversion, minimizes edge injection from the emitter, provides a built-in field in the base region, and provides control of the composite base-layer resistance (see chap. 7 for more details in bipolar device operation). A shallow emitter is next implanted or diffused; this is followed by contacts, metallization.

The process in this basic form utilizes only five masks (or six, if the pad mask is counted), and the transistor is considerably smaller than a junction-isolated bipolar transistor (though not as small as an oxide-isolated device). Two-level metallization adds two additional masking steps. At the 2.5- μm geometry level, no plasma etching or ion implantation is required, and no LPCVD of polysilicon or nitride is needed. The flat surface topography of CDI is also well suited to further device scaling.

The primary disadvantages of the CDI process are the large collector-base and collector-substrate capacitances (as well as relatively low collector-base junction breakdown voltages, which limits them to applications which use small power-supply voltages). Consequently, advanced oxide-isolated structures that minimize these capacitances have largely replaced CDI. However, the benefits of oxide isolation are gained only at the cost of significantly increased process complexity. Work is therefore continuing to increase the performance of CDI so that its compact structure and process simplicity can be exploited for additional applications.²

2.2 BASIC ISOLATION PROCESS FOR MOS ICs (LOCOS ISOLATION)

The isolation requirements of MOS ICs are somewhat different than those of bipolar ICs. Because MOS transistors are self-isolated, as long as the source-substrate and drain-substrate pn junctions are held at reverse bias, the drain current (I_D) should only be due to current flow from source to drain through a channel under the gate. This also implies that no significant current between adjacent MOS devices should exist if no channel exists between them (Fig. 2-4). The buried n^+p junction needed to isolate bipolar transistors is thus not necessary for MOS circuits.

The self-isolation property of MOS devices represents a substantial area savings for NMOS (and PMOS) circuits as compared to junction-isolated bipolar circuits. Hence, ICs having the highest component densities are fabricated with MOS technologies.

The method by which the components of an integrated circuit are interconnected involves the fabrication of metal stripes that run across the oxide in the regions between the transistors (the field regions, Fig. 2-4). However, these metal stripes form the gates of parasitic MOS transistors, with the oxide beneath them forming a gate oxide and the diffused regions (2) and (3) acting as the source and drain, respectively. The threshold voltage of such parasitic transistors must be kept higher than any possible operating voltage so that spurious channels will not be inadvertently formed between devices (see chap. 5 for a discussion of inversion in MOS transistors).

In order to isolate MOS transistors, then, it is necessary to prevent the formation of channels in the field regions, implying that a large value of V_T is needed in the field regions. In practice, the V_T of the field regions needs to be 3-4 V above the supply voltage to ensure that less than 1 pA of current flows between isolated MOS devices¹⁰⁸ (e.g., for 5 V circuit operation, the minimum field-region V_T must be 8-9 V). This condition must be maintained at the minimum isolation spacing used in a given technology, since decreases in field-region V_T arise for same reasons as in short-channel MOS devices (see section 5.5.1.1). Furthermore, V_T also decreases as the temperature rises; reductions of as much as 2 V in field- V_T values have been observed as the temperature rises from 25°C to 125°C.¹⁰⁹

As will be discussed in chapter 5, several methods can be used to raise the threshold voltage. The two utilized to increase field-region V_T 's involve increasing the field-

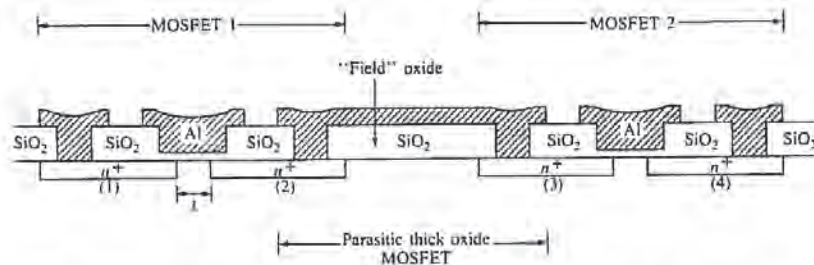


Fig. 2-4 No current should flow between source and drain regions of adjacent MOS transistors. This figure shows the parasitic field transistor with a possible channel under the field oxide if the substrate under this oxide is inverted.¹²⁸ From R. C. Jaeger, *Introduction to Microelectronic Fabrication*. Copyright, 1988, Addison-Wesley. Reprinted with permission.

oxide thickness and raising the doping beneath the field oxide. If the field oxide were made sufficiently thick, it alone could cause a high enough threshold voltage in the field parasitic device. Unfortunately, the large oxide steps produced by this approach would give rise to step coverage problems, and reduced field-oxide thicknesses are thus preferred. To achieve a sufficiently large field threshold voltage with such thinner oxides, the doping under the field oxide must be increased (see chap. 5, Example 5-5).

The field oxide is nevertheless typically made seven to ten times thicker than the gate oxide in the active regions. (This thick oxide also reduces the parasitic capacitance between the interconnect runners and the substrate, improving the speed characteristics of the circuits.) Normally, ion implantation is used to increase the doping under the field oxide. This step is called a *channel-stop implant*. The combination of thick oxide and channel-stop implant can provide adequate isolation for PMOS and NMOS (and as we shall see, for oxide-isolated bipolar) ICs. Additional isolation considerations that arise in the case of CMOS circuits will be described in chapter 6.

The thick field oxide can be fabricated in a number of ways. In one approach (used until about 1970), the oxide is grown to the desired thickness on a flat silicon surface and then etched in the active regions (leaving thick oxide in the field regions, Fig. 2-5b). Although this *grow-oxide-and-etch* approach also allows the isolation regions to be most sharply defined, two disadvantages have prevented it from being utilized in VLSI applications: (1) the field-oxide steps are high and have sharp upper corners, making them very difficult to cover with subsequent metal-interconnect lines; and (2) the channel-stop implant must be performed before the oxide is grown, and thus the active regions must subsequently be aligned to the channel-stop regions in the silicon with a lithography step. This imposes a severe packing density penalty.

In another approach, the oxide is selectively grown over the desired field regions. This is done by covering the active regions with a thin layer of silicon nitride that

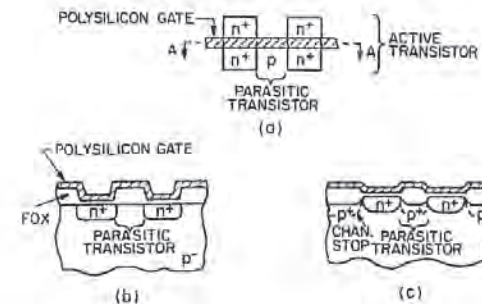


Fig. 2-5 MOSFET isolation. (a) Top view of adjacent MOS transistors with common polysilicon gate illustrating active and parasitic transistor conduction paths. Cross section through A-A for (b) Grow-oxide-and-etch isolation, and (c) LOCOS isolation structures.¹³⁰ From S. M. Sze, Ed., *VLSI Technology*, 2nd Ed., Copyright 1988 McGraw-Hill. Reprinted with permission.

prevents oxidation from occurring beneath them. After the nitride layer has been etched away in the field regions (and prior to field-oxide growth), the silicon in those regions can also be selectively implanted with the channel-stop dopant. Thus, the channel-stop region becomes self-aligned to the field oxide, overcoming one of the drawbacks of the grow-oxide-and-etch method.

The more serious step-coverage limitation of the grow-oxide-and-etch isolation method is also overcome to some degree by the selective-oxidation approach. If the silicon is etched (to a depth of about half the desired field oxide thickness) after the oxide-preventing layer is patterned, the field oxide can be grown until it forms a planar surface with the silicon substrate. (Note: this occurs because the growing oxide film is about twice as thick as the silicon layer it consumes; see Vol. 1, chap. 6). This is known as a *fully recessed isolation oxide process*.

If the field oxide is selectively grown without etching the silicon (Fig. 2-5c) the resulting field oxide will be *partially (or semi) recessed*. In the semirecessed process, the oxide step height is larger than in the fully recessed process, but it is smaller than in the grow-oxide-and-etch process. In addition, the step has a gentle slope that is more easily covered by subsequent polysilicon and metal layers. Semirecessed LOCOS is also less complex than fully recessed LOCOS and results in fewer process-induced defects in the silicon substrate. The semirecessed oxide has become the workhorse isolation technology for MOS devices down to about 1.5- μm geometries, while the fully recessed oxide has primarily been used in bipolar circuits.

The selective oxidation process described above was introduced by Appels and Kooi in 1970,³ and it has become the most widely used approach for forming the thick field oxide. We will first describe this process in detail, discussing its limitations for

submicron devices. Subsequently, alternative isolation schemes that have been proposed to overcome these limitations will be considered.

2.2.1 Punchthrough Prevention between Adjacent Devices in MOS Circuits

Parasitic conduction between adjacent devices due to *punchthrough* (Fig. 2-4c) must also be prevented (see chap. 5 for additional information on punchthrough). In MOS circuits, the source and drain regions of each transistor must be kept far enough from the source and drain regions of any neighboring devices so that the depletion regions do not merge together (i.e., this distance must be greater than twice the maximum depletion-region width). Substrate doping must also be considered, because lighter doping allows wider depletion-region widths, and punchthrough is more likely to occur at lower voltages in lightly doped substrates.

2.2.2 Details of the Semirecessed-Oxide LOCOS Process

Each of the steps used to fabricate conventional semirecessed LOCOS structures will now be discussed in additional detail (Fig. 2-6).

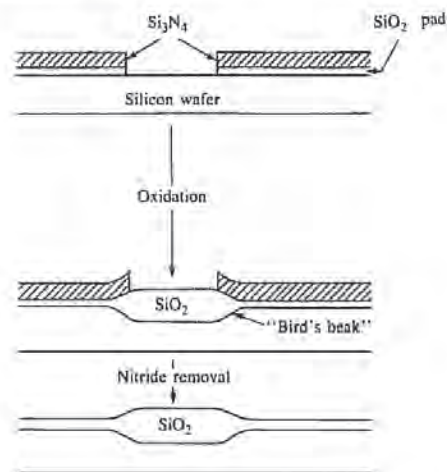


Fig. 2-6 Cross section depicting process sequence for semirecessed oxidations of silicon.¹²⁸ From R. C. Jaeger, *Introduction to Microelectronic Fabrication*. Copyright, 1988, Addison-Wesley. Reprinted with permission.

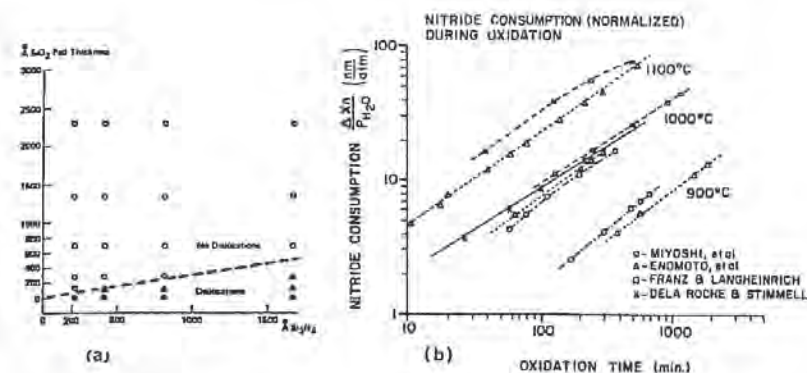


Fig. 2-7 (a) Dislocation generation at Si_3N_4 film edges versus CVD pad-oxide thickness, after 6 h of dry-wet-dry thermal oxidation at 950°C .⁷ Copyright 1978, reprinted with permission of the AIOP. (b) Nitride consumption during oxidation.¹³¹ Reprinted with permission of Semiconductor International.

2.2.2.1 Pad-Oxide Layer. A wafer with a bare silicon surface is cleaned, and a 20-60 nm SiO_2 layer is thermally grown on the surface. The function of this layer, called a *pad* or *buffer* oxide, is to cushion the transition of stresses between the silicon substrate and the subsequently deposited nitride. In general, the thicker the pad oxide, the less edge force is transmitted to the silicon from the nitride (Fig. 2-7a).¹¹⁰ On the other hand, a thick pad-oxide layer will render the nitride layer ineffective as an oxidation mask by allowing lateral oxidation to take place. Therefore, the minimum pad-oxide thickness that will avoid the formation of dislocations should be used. Figure 2-7b indicates that the minimum thickness of a thermally-grown pad oxide should be at least one third the thickness of the nitride layer.

Alternative techniques to allow the use of thinner pad oxides have also been reported. One method involves the use of CVD SiO_2 in place of thermal SiO_2 .⁴ Because CVD SiO_2 is more effective than thermal SiO_2 for avoiding edge defects, such pad layers can be about 25% as thick as thermal-oxide pad layers. The second method utilizes a pad layer consisting of a thin thermal SiO_2 layer and a buffer polysilicon layer.^{5,6,34,35}

2.2.2.2 CVD of Silicon Nitride Layer. Next, a 100-200 nm thick layer of CVD silicon nitride which functions as an oxidation mask is deposited. Silicon nitride is effective in this role because oxygen and water vapor diffuse very slowly through it, preventing oxidizing species from reaching the silicon surface under the nitride. In addition, the nitride itself oxidizes very slowly as the field oxide is grown (typically,

only a few tens of nm of nitride are converted to SiO_2 during the field-oxide growth process). Thus, the nitride should remain as an integral oxidation-barrier layer during the entire field-oxide-growth step. Figure 2-7b shows the thickness of the nitride converted to oxide during a wet silicon oxidation step. Silicon oxidizes approximately 25 times faster than silicon nitride. One criterion for selecting the nitride thickness is that it should be greater than the thickness that will be converted to SiO_2 during the subsequent field oxidation step.

Silicon-nitride films, however, have the well-known drawback of exhibiting a very high tensile stress when deposited by CVD on silicon (on the order of 10^{10} dynes/cm²). The termination of intrinsic stresses at the edge of a nitride film gives rise to a horizontal force that acts on the substrate. Under some circumstances, this stress can exceed the critical stress for dislocation generation in silicon, and will thus become a source of fabrication-induced defects. For example, at 1000°C in oxidizing ambients, defects can be generated at the edges of nitride films as thin as 21 nm if they are deposited directly on silicon.⁷ Pad oxides are used to combat these stresses and avoid dislocation generation. The effect of the pad layer is to reduce the force transmitted to the silicon at the nitride edge, and to relieve the stress of the nitride via the viscous flow of the pad oxide.

2.2.2.3 Mask and Etch Pad-Oxide/Nitride Layer to Define Active Regions. The active regions are now defined with a photolithographic step. A resist pattern is normally used to protect all of the areas where active devices will be formed. The nitride layer is then dry etched, and the pad oxide is etched by means of either a dry- or wet-chemical process. After the pad oxide has been etched, the resist is not removed but instead is left in place to serve as a masking layer during the channel-stop implant step.

2.2.2.4 Channel-Stop Implant. An implant is next performed in the field regions to create a channel-stop doping layer under the field oxide. In NMOS circuits, a p^+ implant of boron is used, while in PMOS (and in the n -tubs of CMOS circuits) an n^+ implant of arsenic is utilized. Although this normally requires two masking steps in CMOS circuits, a single mask process for implanting both p and n channel stops has been reported.⁸ After the implant has been completed, the masking resist is stripped.

2.2.2.5 Problems Arising from the Channel-Stop Implants. During field oxidation, the channel-stop boron experiences both segregation and oxidation-enhanced diffusion. Thus, relatively high boron doses are needed (mid 10^{12} - 10^{13} atoms/cm²) in order for acceptable field threshold voltages to be achieved. This also implies that the peak of the boron implant must be deep enough that it is not absorbed by the growing field-oxide interface (implant energies in the 60-100 keV range are used). If the channel-stop doping is too heavy, it will cause high source/drain-to-substrate capacitances and will reduce source/drain-to-substrate pn junction breakdown voltages.

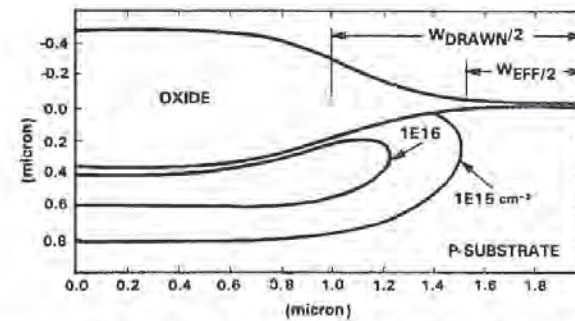


Fig. 2-8 Field oxide and boron encroachment in LOCOS.⁹⁶ From K. M. Cham et al., *Computer-Aided Design and VLSI Device Development*. Copyright 1986, Kluwer Academic Publishers. Reprinted with permission.

The lateral diffusion of the boron also causes it to encroach into the NMOS active areas (Fig. 2-8). Such redistribution raises the boron surface concentration near the edge of the field oxide, causing the threshold voltage to increase in that region of the active device. As a result, the edge of the device will not conduct as much current as the interior portion, and the transistor will behave as if it were a narrower device. (This effect is also enhanced as the dose of the channel-stop implant is increased.)

Finally, dislocations generated during the channel-stop implant step (which would ordinarily climb out to the surface during appropriate annealing conditions) can be caused to glide under the nitride-covered regions by the stresses at the nitride edges. The penetration of such dislocations through nitride-edge-defined junctions can cause increased emitter-base leakage in bipolar devices.¹⁰

The extent of the channel-stop dopant diffusion can be reduced by using high-pressure oxidation (HIPOX) to grow the field oxide,⁸ by use of a germanium-boron co-implant,⁹ or by use of a chlorine implant.⁵¹ HIPOX allows the oxide-growth temperature to be reduced, which reduces the diffusion length of the boron. The germanium-boron co-implant exploits the fact that boron diffuses with a lower diffusivity in the presence of implanted germanium, and boron segregation effects are also reduced (see Vol. 1, chap. 7). The result is a 40% increase in field threshold voltage for the same dose of implanted boron, and a corresponding decrease in lateral-boron encroachment. The chlorine implant is performed in the field regions prior to field oxidation, and this causes the oxide to grow at a faster rate. Consequently, the field oxide can be grown in less time at the same temperature.

2.2.2.6 Grow Field Oxide. After the channel-stop implant has been performed, the field oxide is thermally grown by means of wet oxidation, at temperatures of around 1000°C for 2-4 hours (to thicknesses of 0.3-1.0 μm). The oxide grows where there is

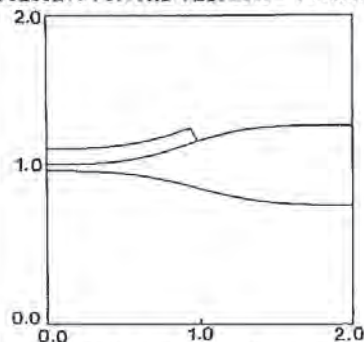


Fig. 2-9 Schematic of the bird's beak that occurs during semirecessed LOCOS.

no masking nitride, but at the edges of the nitride, some oxidant also diffuses laterally. This causes the oxide to grow under and lift the nitride edges. Because the shape of the oxide at the nitride edges is that of a slowly tapering oxide wedge that merges into the pad oxide, it has been named a *bird's beak* (Fig. 2-9). The bird's beak is a lateral extension of the field oxide into the active area of the devices. Although the length of the bird's beak depends upon a number of parameters – including the thicknesses of the buffer oxide, nitride, and isolation oxide (as well as the oxidation temperature and oxygen partial pressure) – the length for a typical $0.5\text{--}0.6\text{-}\mu\text{m}$ field oxide is $\sim 0.5\text{-}\mu\text{m}/\text{side}$. This would make a lithographically defined $1\text{-}\mu\text{m}$ feature disappear on the chip following the field oxidation step (Fig. 2-10). Although this effect led to predictions that LOCOS isolation would have to be replaced for device dimensions smaller than $2\text{-}\mu\text{m}$, optimization of process steps has allowed conventional LOCOS to continue to be used for device-isolation spacings as narrow as $1.25\text{--}1.5\text{-}\mu\text{m}$.^{11,14}

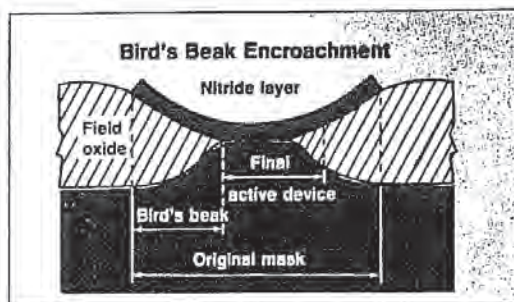


Fig. 2-10 Bird's beak encroachment limits the scaling of channel widths to about $1.2\text{--}1.5\text{-}\mu\text{m}$.¹¹ Reprinted with permission of Semiconductor International.

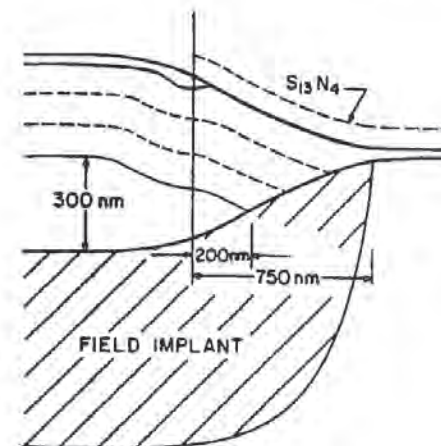


Fig. 2-11 Eventual exposure of the p -substrate region when etching back the LOCOS field oxide in an NMOS device (in an attempt to regain some of the active device area lost by bird's beak encroachment).

The LOCOS bird's beak also creates another problem during the later process step of contacting metal to the source and drain regions of an MOS device. Because of the shape of the bird's beak, any overetching that must be performed during the contact-window-opening step will etch away part of the bird's-beak oxide. This may expose the substrate region under the source or drain region (Fig. 2-11). If this occurs, the source will become shorted to the well region when the metal interconnect film is deposited, impairing or destroying device operation.

The problem worsens in CMOS as shallower junctions are used, due to exposure of the well region. (If deeper junctions are used, the lateral diffusion distance is also longer and etching back part of the bird's beak may not expose the well region.) In NMOS processes, this problem was counteracted by redoping of the contacts with phosphorus after they were opened (and before the metal was deposited). This created a deep junction in the contact areas only. The source/drain junction near the gate oxide was kept shallow in order for good device performance to be retained. In CMOS processing, not only are shallow junctions used, but redoping of the contacts presents an additional problem because both n^+ and p^+ contact openings are defined at the same masking step. As a result, in CMOS processes enclosed contacts are commonly employed. This solution, however, reduces packing density, as area must be provided in the source and drain regions to keep the contact from overlapping the bird's beak. Another solution would be to use an alternative isolation process in which the contacts are able to coincide with (or overlap) the active areas.

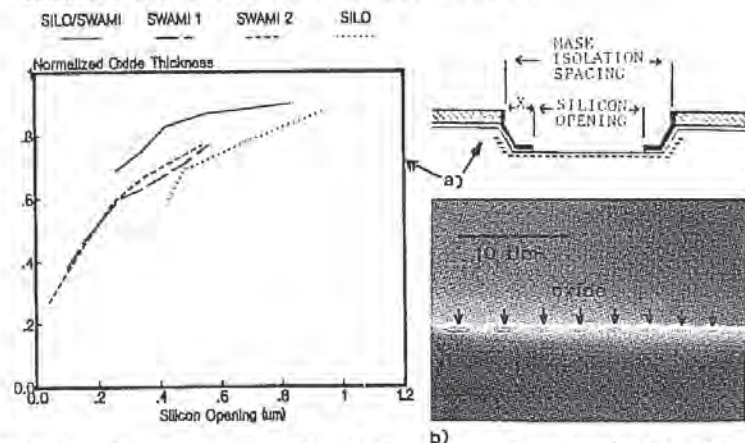


Fig. 2-12 (a) Normalized field oxide thickness versus silicon opening for various advanced LOCOS processes.¹⁴ (© 1985 IEEE). (b) SEM photo of the thinning of the field oxide as the dimension of the exposed silicon gets smaller.¹²¹ This paper was originally presented at the Spring 1989 Electrochemical Society Meeting held in Los Angeles, CA.

Modifications to the basic process have allowed the bird's beak to be reduced in length. These include etching back a portion of the field oxide after it is grown; using silicon nitride without a pad oxide; and using a thin pad oxide covered with polysilicon under the masking nitride layer. These advances will be described in more detail in a later section.

One last limitation of LOCOS-based isolation schemes for submicron structures is the oxide field-thinning effect.¹⁴ The field-oxide thickness in submicron-isolation spacings is significantly less than the thickness of field-oxides grown in wider spacings (Fig. 2-12).^{14,121} The narrower the width of the exposed substrate silicon region, the thinner the field oxide. For example, a field oxide that is grown to a thickness of 400 nm above a region of exposed Si that is 1.5- μm wide would be only about 290 nm thick if grown above a 0.8- μm -wide region of Si.

This effect is believed to be caused by the reduction in the oxidants available in the submicron opening compared to those available in wider openings.¹⁵ Thin field oxides that result from this effect can have a great impact on field-threshold voltages and on the interconnect capacitances to substrate. It is predicted that as a result of this field-thinning effect and the need to maintain defect-free isolation structures (see next section), the minimum space that must be allowed for a LOCOS-type isolation structure with a field-oxide thickness of 550 nm will be 0.75 μm .

2.2.2.7 Strip the Masking Nitride/Pad-Oxide Layer. Following field oxidation, the masking layer is removed. Since 20-30 nm of the top of the nitride is converted to SiO_2 during the field oxidation, this layer must be etched off first. The

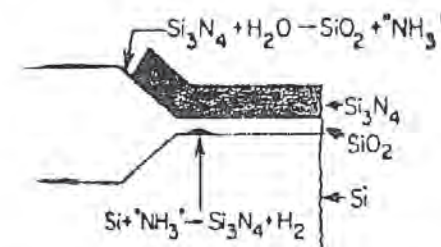


Fig. 2-13 Kooi's model for explaining nitride growth under the gate oxide.¹⁶ Reprinted by permission of the publisher, The Electrochemical Society, Inc.

remaining nitride and pad oxide are then etched. These steps can be carried out by means of wet-chemical etching, since there is no need to maintain dimensions or to otherwise tightly control the etching. Since chemical methods that remove nitride offer excellent selectivity with respect to the underlying oxide, considerable overetching can be used.

2.2.2.8 Regrow Sacrificial Pad Oxide and Strip (Kooi Effect).

During the growth of the field oxide, another phenomenon occurs that causes defects in when the gate oxide is grown. Kooi et al., discovered that a thin layer of silicon nitride can form on the silicon surface (i.e., at the pad-oxide/silicon interface) as a result of the reaction of NH_3 and silicon at that interface (Fig. 2-13).¹⁶ The NH_3 is generated from the reaction between H_2O and the masking nitride during the field-oxidation step. This NH_3 then diffuses through the pad oxide and reacts with the silicon substrate to form silicon-nitride spots or ribbon (these regions are sometimes called *white ribbon*). When the gate oxide is grown, the growth rate becomes impeded at the locations where the silicon nitride has formed. The gate oxide is thus thinner at these locations than elsewhere, causing low-voltage breakdown of the gate oxide. One way to eliminate this problem is to grow a "sacrificial" gate oxide after stripping the masking nitride and pad oxide, and removing it before growing the final gate oxide.^{17,18}

2.2.3 Limitations of Conventional Semi-Recessed Oxide LOCOS for Small-Geometry ICs

The limitations of conventional LOCOS for submicron technologies can be summarized as follows:

- The bird's-beak structure causes unacceptably large encroachment of the field oxide into the device active regions (Fig. 2-10).

- Boron from the channel-stop implant of n -channel MOSFETs is excessively redistributed during the field-oxide growth and other high-temperature steps, leading to unacceptable narrow-width effects (see chap. 5).
- The planarity of the surface topography is inadequate for submicron lithography needs (see section 2.9.2).
- The thickness of the field oxide in submicron regions of exposed silicon is significantly thinner than that grown in wider spacings (Fig. 2-12). This field-oxide-thinning effect can produce problems with respect to field threshold voltages, interconnect-to-substrate capacitance, and field-edge leakage. Finally, it introduces additional nonplanarity across the surface of the wafer.

2.3 FULLY RECESSED OXIDE LOCOS PROCESSES

Several conventional, fully recessed oxide LOCOS methods have been developed for bipolar integrated-circuit applications, including the ISOPANAR,¹⁹ ROI,²⁰ and OXIS²¹ processes. Such fully oxide-isolation processes allow higher performance bipolar ICs to be fabricated than if junction isolation is used. The collar of SiO_2 that forms the sidewall isolation eliminates the sidewall contribution to collector-to-

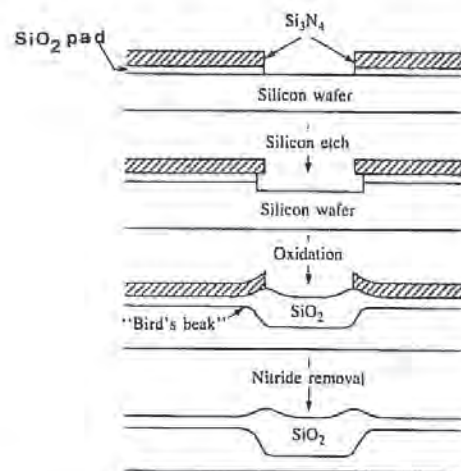


Fig. 2-14 Cross section depicting process sequence for fully-recessed oxidation of silicon.¹²⁸ From R. C. Jaeger, *Introduction to Microelectronic Fabrication*. Copyright, 1988, Addison-Wesley. Reprinted with permission.

substrate and base-to-collector capacitances, thereby increasing the cutoff frequency of oxide-isolated transistors to 3-5 GHz – well above the 1-GHz limit of junction-isolated devices. In addition, the area of the transistor is reduced by about 70%.

In all of these approaches, the basic semirecessed LOCOS process is still followed, with the following modification: After the nitride and pad oxide have been defined, the silicon-epitaxial layer of the standard buried-collector (SBC) n pn bipolar process is etched (Fig. 2-14). About one-half to two-thirds of the epi layer is removed, which creates grooves in the silicon. This step can be performed using wet-isotropic etching (using an HNO_3/HF solution); wet-anisotropic etching, to yield slope-walled grooves (using a KOH solution – see Vol. 1, chap. 15); or dry etching. Boron is diffused or implanted into the bottom of the etched areas as a channel stop. The patterned nitride/oxide regions again mask the implant so that the active areas are not implanted.

Next the field oxide is grown (typically to a thickness of 1-2 μm). About 45% of the oxide growth is downward (see Vol. 1, chap. 6) and extends completely through the epi layer. Since 55% of the oxide growth is upward, it expands to fill the etched space, and a relatively planar device surface can be produced. (We will see in the next section that the surface of the fully-recessed-oxide process is actually not completely planar.)

This field oxide performs a somewhat different isolation function in bipolar ICs than in MOS circuits. That is, it primarily serves to replace the p -diffusion at the device sidewalls, thereby preventing adjacent collector regions from being directly electrically connected. Whereas in MOSFETs the minimum oxide thickness is determined by the required V_T in the field regions, in the bipolar device the oxide must penetrate all the way through the epi layer, and so must normally be thicker. To achieve this without producing intolerably high steps at the surface, it is necessary to etch grooves into the silicon initially. Since the growth of field oxides with thicknesses greater than 2 μm is impractical (even when grooves are etched in the Si), oxide isolation is not used in bipolar ICs that require power supply voltages exceeding 15 V (i.e., such technologies require epi layers greater than 2- μm thick).

The oxide-isolated bipolar transistor is smaller than the junction-isolated transistor because the lateral dimension of the oxide region is somewhat smaller than that of the p -diffusion region, but its size is also reduced for several other reasons (see chap. 7, Fig. for a comparison of bipolar device sizes with various isolation processes). First, the SiO_2 can be butted directly against the buried-collector n^+ region since the problem of depletion-region merging (inherent in junction isolation) does not exist. Second, the collector-metal stripe and the base-metal stripe can overlap the sidewall oxide. Finally,

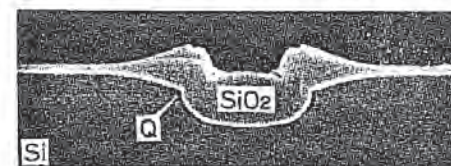


Fig. 2-15 SEM micrograph showing bird's head and beak.

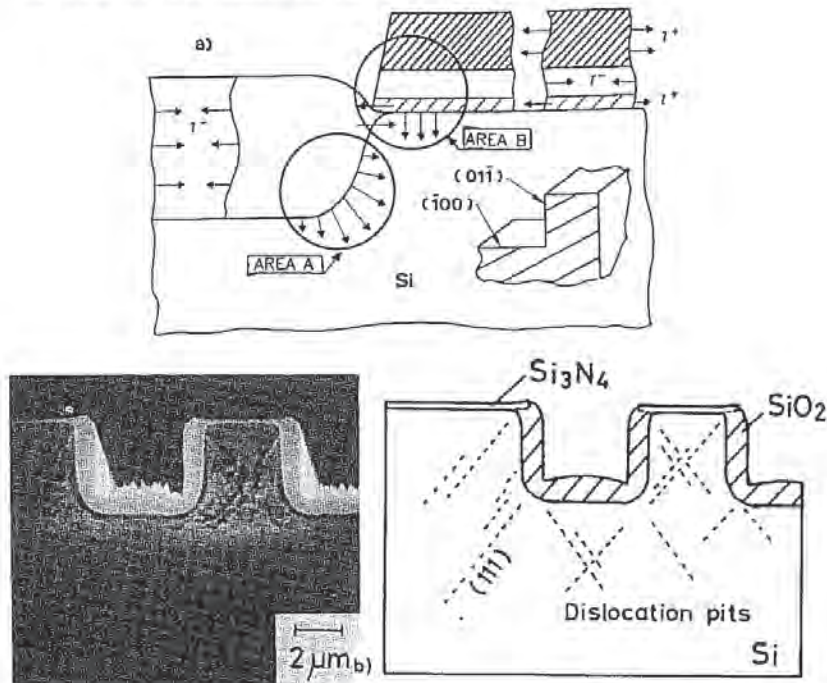


Fig. 2-16 (a) Schematic representation of the stress field induced in the silicon substrate by the SILO process.³² (© 1985 IEEE). (b) Dislocations caused by stress in growing oxide in trench structures as revealed by Secco etching.⁴⁹ Reprinted by permission of the publisher, The Electrochemical Society, Inc.

the collector n^+ contact, the p -type base, and even the emitter region can extend to the isolation region. Since such smaller-area devices further reduce parasitic capacitances, oxide isolation has become the technology of choice for advanced, high-performance, 5-V digital circuits.

There are, however, several problems that arise from the etching of the grooves in the silicon. First, the fully recessed field oxide is not completely planar - that is, a bulge (or *bird's head*) occurs in the oxide surrounding the nitride pads (Fig. 2-15) which does not occur in the semirecessed LOCOS field oxide. The bird's head can cause step-coverage problems in the metallization layers due to uneven surface topology. Etchback procedures have been proposed to minimize the problem in multimetal processes.²² Second, as described in the next paragraphs, the growth of the field oxide in the grooves causes stresses that can lead to defects in the silicon.

The most important defect-generating stresses arise at the bottom corners of the etched grooves during field oxide growth and are due to the volume misfit of the growing oxide. First, as the oxide grows, it pushes against the sidewalls of the groove. In addition, as the oxide in the corner grows, its top surface is pushed outward by the newly forming oxide layer, and thus becomes tangentially stressed. Finally, as the laterally encroaching oxide grows under the nitride, it is restrained from growing upward by the stiffness of the nitride layer, causing further stress downward against the silicon in the corner of the recess. These stresses, if not relieved, will generate dislocations in the silicon (Fig. 2-16).

The two mechanisms that can prevent these stresses from producing defects as the oxide is grown in the silicon recesses are *stress reduction* and *stress relief*. The stress produced at the corners of the recessed silicon can be reduced by ensuring that the corners of the etched structures are rounded, and by using thinner silicon-nitride masking layers. Stress can be relieved through the viscoelastic flow of the oxide using growth above 960°C.²³ Enhanced viscous flow also occurs under high-pressure oxidation at reduced temperatures. It has been reported that high-pressure oxidation can be used to drastically reduce field-oxidation time, and thereby to also reduce the lateral diffusion of the channel-stop implants and the updiffusion of the buried-collector layer in bipolar ICs.²⁴ Fewer defects occur if the sidewalls of the recessed regions are inclined away from the vertical at angles of 75° or more. This type of structure appears to facilitate stress relief by enhancing the viscoelastic-flow effect.

2.3.1 Modeling the LOCOS Process

Among the many aspects of the LOCOS process that would be useful to simulate are the oxide thickness and morphology, the shape and dimensions of the bird's beak and bird's head, the redistribution of the channel-stop dopant, and the stresses induced in the silicon during the LOCOS processing. Models for simulating these features have been developed, and they are described more fully in chapter 9.^{26,27,28}

2.4 ADVANCED SEMIRECESSED-OXIDE LOCOS ISOLATION PROCESSES

The conventional LOCOS process described earlier is limited by oxide and boron encroachment, as well as by nonplanar surface topographies and the field-thinning effect. These limitations must be overcome to allow the LOCOS process to be used for submicron IC designs.

2.4.1 Etched-Back LOCOS

The simplest way to reduce the bird's beak and to obtain a more planar surface is to etch back a portion of the field oxide after it has been grown.^{12, 13} In this procedure, the upper region of the field oxide is removed to recover the active-area loss and to thin the

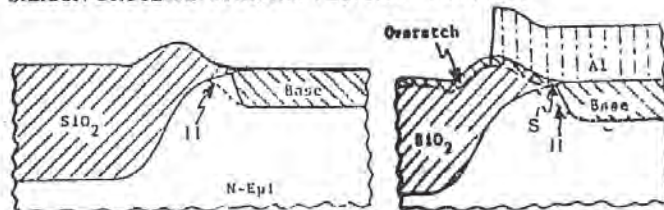


Fig. 2-17 Implanting the base of a bipolar transistor allows the bird's head and beak to be etched back to some degree without exposing the substrate region of opposite doping polarity.

field oxide so that a more planar topography is produced (Fig. 2-11). What keeps this from being useful for submicron processes is that in order for the bird's beak to be significantly reduced, the p -substrate must become exposed, which will cause the drain-to-substrate regions of a MOSFET to be shorted. Implantation of source and drain regions (or base regions of a bipolar transistor), instead of forming them by diffusion, extended the use of this approach, as shown in Fig. 2-17. However, the problem of boron encroachment is not alleviated by this approach.

2.4.2 Polybuffered LOCOS

Another semirecessed-LOCOS bird's-beak compaction scheme exploits the fact that a thinner pad oxide results in a shorter bird's beak. Therefore, the usual pad-oxide layer is replaced with a polybuffered pad layer (poly [50-nm]/oxide [5-10-nm]) and a thicker-than-usual nitride (100-240 nm).^{5,6,34} A bird's beak of only 0.1-0.2 $\mu\text{m}/\text{side}$ is produced for a field oxide of 400-nm thickness. Reported leakage currents of the n^+-p junction were as low as with a conventional LOCOS process. The basic polybuffered LOCOS process reduces bird's beak length but does not solve the boron-encroachment or planarity difficulties of conventional LOCOS. Reference 125 provides an extensive characterization of the polybuffered LOCOS method, including the effects of varying the pad oxide, polysilicon, and nitride layer thicknesses on the bird's beak length and defect formation.

An more successful advanced polybuffered LOCOS process has recently been reported (Fig. 2-18).³⁵ In this approach, a thinner field oxide is used, and the channel-stop implant is performed through this 220-nm-thick field oxide. The 100-keV boron implant that forms the channel stop also serves as the punchthrough implant in the active regions. Since the channel-stop implant is done after the field-oxide growth, much less boron encroachment occurs. In addition, a smaller field-oxide step is produced, thereby helping the planarity requirement. A thinner field-oxide region can be used because the process is designed for 0.7- μm (and smaller) device applications. At these short channel lengths, the doping of the substrate must be increased to prevent punchthrough (see chap. 5). As a result, the dopant concentration of the channel stop also ends up being high enough to prevent field inversion, even with the use of such a thin field oxide.

2.4.3 SILO (Sealed-Interface Local Oxidation)

The SILO isolation process is another modification of conventional LOCOS. It reduces bird's-beak length to about 0.2 μm by forming a layer of silicon nitride directly on the silicon surface before depositing an oxide pad layer over it.³² In this way the interface between the silicon surface and the nitride is sealed, and lateral diffusion of oxidants is prevented (Fig. 2-19). As a result, a shorter bird's beak is produced (Fig. 2-20).¹²³

In SILO, a thin (10-20-nm) nitride layer is formed on the silicon surface by thermal nitridation of the silicon, or CVD. (The use of such a very thin reduces the defect-causing nitride-edge stress.) Next, a 25-30-nm CVD SiO_2 film and a 150-200-nm nitride film are sequentially deposited. After the three-layer-masking film has been patterned by RIE, the field oxide can be grown on the unetched silicon surface to create a semirecessed field oxide, or the silicon surface can be etched so that a fully recessed field oxide can be grown. The channel stop is implanted before the field oxide is grown. After field oxidation, the masking layer is etched away.

There are some problems with SILO. If the silicon is not etched, the semirecessed field oxide will cause steps on the wafer surface. While these are no higher than those

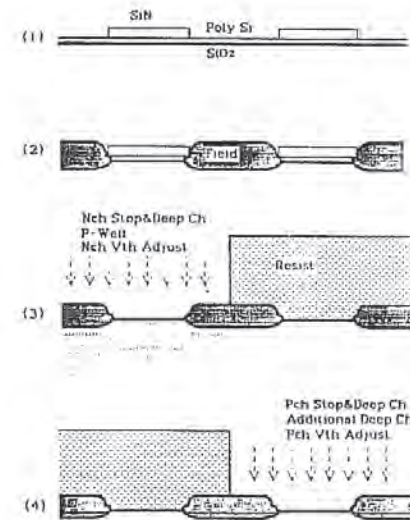


Fig. 2-18 Process sequence of APPL isolation structure.³⁵ (© 1988 IEEE).

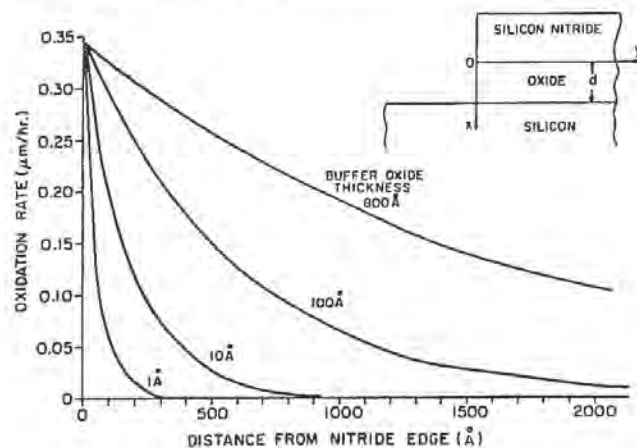


Fig. 2-19 Oxidation rates versus distance from the edge of the nitride edge with the buffer (pad) oxide thickness as a parameter.¹¹⁰ (© 1982 IEEE).

of conventional LOCOS, they are steeper, since the bird's beak has been nearly eliminated (Fig. 2-21).³³ If the silicon is etched prior to oxide growth, many dislocations will be produced if the etched-silicon sidewalls are vertical. In order to

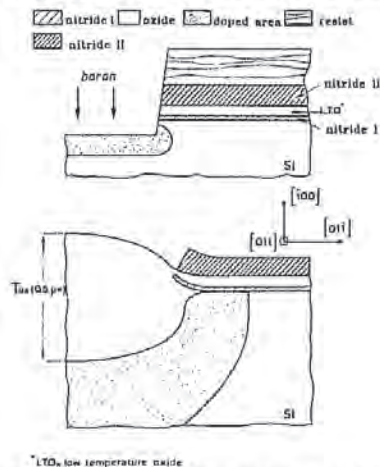


Fig. 2-20 Example of a shortened bird's beak achievable with a SILO process.³² (© 1985 IEEE).

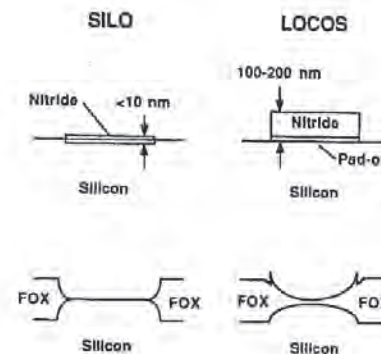


Fig. 2-21 Schematic of SILO and LOCOS structures before and after field oxidation, showing the steeper field oxide sidewalls with SILO.³³ (© 1988 IEEE).

avoid this, the bottom corner of the recess should have a gentle, rounded slope, best produced through isotropic etching. Unfortunately, this leads to lateral oxide encroachment, as illustrated by a 1.2- μm CMOS process described by Motorola in 1986. The SILO process oxide encroachment was 0.2 $\mu\text{m}/\text{side}$.³⁶ Finally, it has been reported that the stress associated with the abrupt transition from field oxide to thin gate oxide leads to increased sensitivity with respect to localized interface-trap generation.³⁷

2.4.4 Laterally Sealed LOCOS Isolation

Ghezzi et al. developed another approach for sealing the silicon interface during field-oxide growth, called *laterally-sealed-LOCOS isolation*.³¹ This process seeks to maintain the bird's-beak reduction of SILO while reducing the defects that are induced when a nitride is in direct contact with the silicon. The active-device area is covered with a pad-oxide layer, while the silicon interface is sealed with nitride. This achieves the low-defect densities provided by the pad oxide, together with the reduced bird's beak of SILO.

This method employs an initial masking layer (called the *active-area stack*) consisting of a 20-nm pad oxide, an 80-nm CVD nitride, and a 200-nm plasma-enhanced (PE) CVD oxide. The purpose of the top PECVD oxide is to provide vertical support for the subsequently deposited second nitride layer which will serve to seal the sidewalls of the active-area stack. In the next stack patterning step, overetching ensures

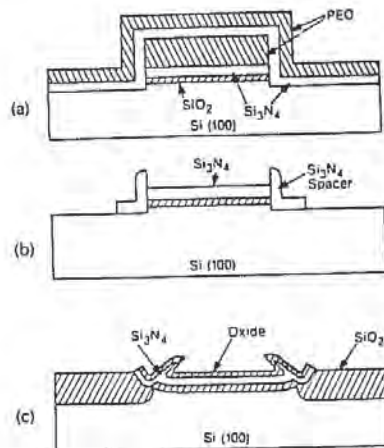


Fig. 2-22 Process sequence for the laterally sealed LOCOS isolation. (a) Active area stack consisting of thin SiO₂ layer covered with Si₃N₄ and a thicker SiO₂ layer (PEO) and "sealed" with a double layer of Si₃N₄ and SiO₂. (b) After RIE to produce a Si₃N₄ spacer around the edges of active area. (c) Active area and field structure after thermal oxidation to form thick field oxide.³¹ Reprinted by permission of the publisher, The Electrochemical Society, Inc.

complete film removal in the field regions. It also removes ~10 nm of Si, but this extra silicon etch actually helps to seal the edge of the original pad oxide by forming a thin inlay for the nitride spacer that is subsequently deposited (see Fig. 2-22a).

After the resist has been stripped, a second 80-nm CVD-nitride film and a second 220-nm PECVD-oxide film is deposited to provide conformal coverage of the active-area stack. An anisotropic RIE step is then carried out to remove these second films from the field, along with the first PECVD oxide. This forms an L-shaped nitride spacer that seals the sidewalls of the stack (Fig. 2-22b). The field oxide can now be grown. The foot of the L-shaped nitride spacer is quite thin and easily bends back toward the sidewall under the volumetric expansion of the growing field oxide (Fig. 2-22c).

A thin layer of nitride is converted to oxide during the field oxidation. The process is completed by removing this oxide and the residual nitride. Figure 2-23a shows an active-area stack and the nitride spacer following spacer formation and sacrificial-oxide removal, while Figure 2-23b shows a cross-section after a 800-nm-thick field oxide has been grown and the nitride is still in place (the bird's-beak is 0.2 μm long). Diode leakage similar to that seen in conventional LOCOS (with a bird's-beak length of 0.6 μm) has been observed in this structure, which has a bird's-beak length of 0.45 μm.

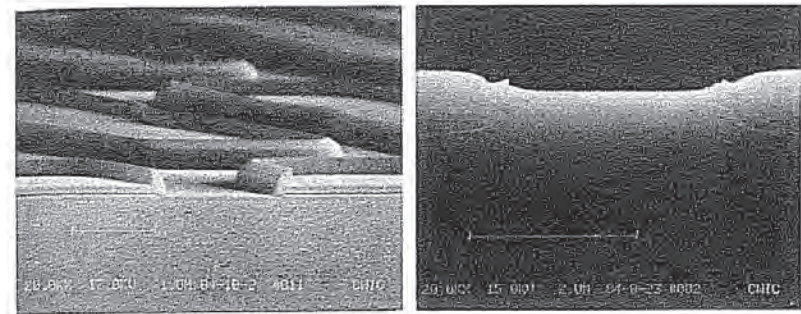


Fig. 2-23 SEM micrographs of isolation structures with laterally sealed LOCOS.³¹ Reprinted by permission of the publisher, The Electrochemical Society, Inc.

In a variation of the laterally sealed LOCOS process (Fig. 2-24) that incorporates a protective-oxide pad and nitride spacer that seals the masking-layer sidewall (named POP-SILO),³⁸ a 12.5-nm pad oxide, an 80-nm CVD nitride, and a 150-nm LPCVD oxide are sequentially deposited. The thickness of the oxide layer is chosen to provide

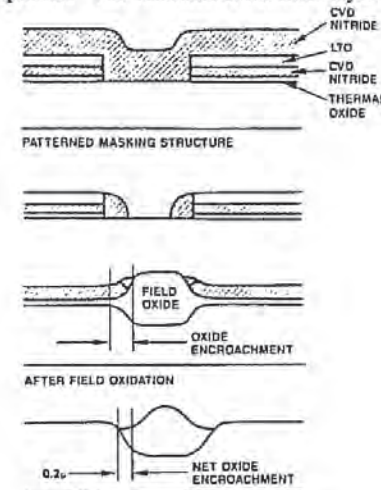


Fig. 2-24 POP-SILO process flow.³⁸ This paper was originally presented at the Spring 1989 Meeting of The Electrochemical Society, Inc. held in Los Angeles, CA.

the necessary step height for the spacer formation. After the previously described layers have been etched to the silicon substrate to define the active areas, a second CVD-nitride film is deposited to achieve a spacer length of 250 nm. The base of the spacer defines the length of the sealed interface. Field oxide is grown under and beyond the spacer so that the area of silicon on which subsequent gate oxide will be grown is not exposed to silicon nitride. After field oxide growth, the masking layer is removed. The typical isolation encroachment for this process is 0.2 μm larger than the originally masked pattern. However, the problems of planarity, boron encroachment, and thinning of the field oxide (when the windows to the exposed silicon are of submicron dimensions) do not appear to be addressed by these two approaches.

2.4.5 Bird's Beak Suppression in LOCOS by Mask-Stack Engineering

When isolation spacings approach 1 μm in dimension some other effects can be exploited to suppress lateral-oxide encroachment in a semirecessed LOCOS process. First, the stiffness of the mask stack (which consists of the pad-oxide and nitride films) is dramatically increased, and as a result it exhibits much more resistance to bending as the oxide film grows laterally under it, than when the mask is wider. If the deformation of the mask stack can be decreased, lateral oxidation is also reduced since the oxide growth requires volumetric expansion. The magnitude of the lateral encroachment can be reduced by as much as 30% for a 0.5 μm mask length, compared to a "long" mask length. Through optimization of the relative film thicknesses of the mask stack, along with the mask length and the degree of interface sealing (i.e., by using nitride, oxynitride, or polybuffered pad layers), an acceptable degree of suppression of the bird's beak can be produced for a given technology. This process-development task has been referred to as *mask-stack engineering*.¹²¹

2.4.6 Planarized SILO with High-Energy Channel-Stop Implant

An advanced SILO approach has also been reported that attacks the problems of bird's-beak encroachment, boron channel-stop encroachment, and planarity.^{39,111,112} In this process (Fig. 2-25), a sandwich film consisting of a silicon-oxynitride pad layer and a silicon-nitride masking layer is used to suppress bird's-beak formation. The semirecessed field oxide is then planarized with an etchback process (see chap. 4 for more information on etchback processes). Finally, the channel-stop implants are implanted through the field oxide; hence, very little boron encroachment occurs. With the exception of the field-thinning problem, this approach is predicted to allow such modified LOCOS techniques to be extended to 0.5- μm CMOS technology.

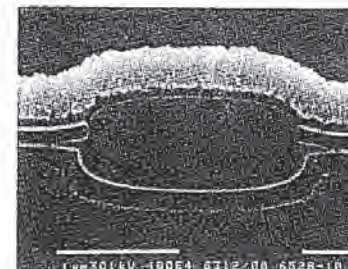


Fig. 2-25 SEM micrograph of 700 nm SLOCOS.¹¹¹ This paper was originally presented at the Spring 1989 Meeting of The Electrochemical Society, Inc. held in Los Angeles, CA.

2.5 ADVANCED FULLY RECESSED OXIDE LOCOS ISOLATION PROCESSES

Work has also been done to overcome the limitations exhibited by fully recessed oxide LOCOS isolation technologies. Although the conventional, fully recessed LOCOS offers the benefit of a planarized surface, it also possesses all of the drawbacks of the semirecessed LOCOS approach (as well as those of increased process complexity and stress-induced defect generation). The advanced fully recessed LOCOS isolation techniques were developed to overcome these drawbacks.

2.5.1 SWAMI (Sidewall-Masked Isolation Technique)

The *sidewall-masked isolation* (SWAMI) technique has received much attention since it was introduced as a bird's-beak-free isolation scheme by Chiu et al. in 1982.²⁹ Its main advantages over conventional LOCOS are an increase in packing density (due to limited lateral encroachment), no severe restrictions on the field-oxide thickness, improvement of the surface planarity, and complete elimination of the gate-oxide-thinning phenomenon (Kooi effect).

SWAMI pad-oxide and CVD-nitride layers are formed and etched in the same manner as in conventional LOCOS. Grooves are then etched in the silicon to a depth of approximately half the desired field-oxide thickness. This is most often carried out by means of an orientation-dependent silicon etch, such as KOH. This produces recesses having sidewalls that are inclined about 60° when formed on a <100> surface. During field oxidation these sloping sidewalls help to reduce the stresses that contribute to the generation of edge defects.

A second stress-relief oxide layer is then grown, followed by the deposition of a second CVD nitride (which provides conformal coverage of the entire surface, including the sidewalls of the silicon recesses) and a CVD oxide (Fig. 2-26a). This composite CVD SiO_2 -CVD Si_3N_4 -thermal SiO_2 layer is anisotropically etched in the field region so that it remains only on the sidewalls of the recessed silicon. Because the CVD oxide forms a spacer that protects part of the second nitride layer, this nitride forms a structure with a foot that extends partway into the exposed silicon at the bottom of the recess. After the oxide spacer has been etched away, the final structure is a silicon mesa whose sidewalls are surrounded by the second nitride and oxide. At this point the channel-stop implant is performed, and the field oxide is grown.

In the latter step, the thin sidewall nitride is bent upward due to the expansion of the converted SiO_2 . The length of the foot of the nitride on the floor of the recessed silicon is selected to minimize the growth of the bird's beak into the active area (the active-area boundary is defined by the edge of the first nitride). Finally the masking nitride/oxide layers are removed.

A perspective view of a SWAMI isolation structure is shown in Fig. 2-26b. The bird's-beak is reduced, and a relatively planar surface topography is achieved.

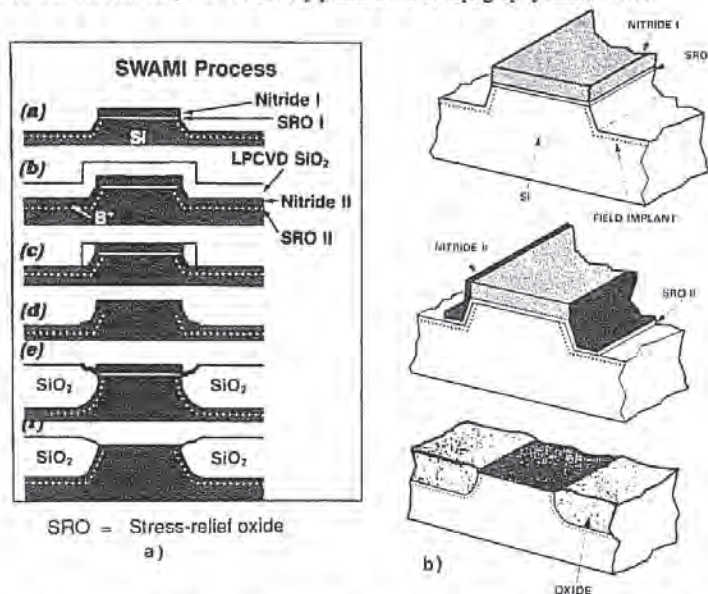


Fig. 2-26 (a) The SWAMI process developed by Hewlett-Packard²⁹ (© 1982 IEEE). (b) Perspective views of the SWAMI process.⁹⁶ From K. M. Cham et al., *Computer-Aided Design and VLSI Device Development*. Copyright 1986, Kluwer Academic Publishers. Reprinted with permission.

Several problems, however, have been reported with the SWAMI method. First, this process is more complex than conventional LOCOS. Second, the depth of the KOH-formed silicon recesses may vary depending on the position on the wafer, especially since no etch stops are employed. This can lead to differences in the corner structure at the field-oxide/active-area border, which in turn might widen the distribution of electrical-device parameters (such as threshold voltage and punchthrough leakage current). Third, the benefits of reduced edge defects obtained by using a KOH etch step (i.e., due to the 60°-inclined sidewalls that the etch produces) are offset by the characteristic property of this type of etch to delineate surface-crystallographic imperfections. If such imperfections are present, the field-oxide quality and uniformity are affected.

A double-silicon plasma etch that still produces sloped sidewalls has therefore been proposed as a replacement for the KOH etch process.^{30,124} The electrical characteristics of SWAMI structures fabricated by this dry-etch process have been reported to be as good as those obtained with standard LOCOS. Although the double-silicon etch SWAMI method involves a more complex manufacturing sequence than LOCOS or single-etch SWAMI, it also provides a greater degree of planarization and a reduction in the lateral field oxide encroachment than the former, and significantly decreased leakage current levels than the latter.

2.5.2 SPOT (Self-Aligned Planar-Oxidation Technology)

Another modified SWAMI-like fully recessed LOCOS process that eliminates the bird's head and produces a highly planar surface with no observed dislocation generation has been reported by Sakuma et al.²⁵ In this isolation process, called *self-aligned planar-oxidation technology* (SPOT), a conventional semirecessed field oxide is first grown using high-pressure oxidation. This field oxide is then removed with a buffered HF solution. Next, a second pad oxide (100 nm thick) is grown on the exposed recessed silicon, followed by a 100-nm CVD nitride that conformally covers all surfaces. These two layers are then anisotropically etched, which removes them from the bottom of the recessed silicon areas but not from under the overhanging nitride film (Fig. 2-27a). The second field oxide (2 μm thick) is grown using high-pressure oxidation at 900°C.

Figure 2-27b shows the SPOT isolation structure compared to the structures obtained with semirecessed and fully recessed LOCOS. Because the shape of the exposed silicon following removal of the first field oxide is very smooth, only small stresses are created between the silicon and SiO_2 as the second field oxide is grown. Hence, dislocation generation is avoided.

2.5.3 FUROX (Fully Recessed Oxide)

The fully recessed-oxide (FUROX)³⁹ isolation technology is very similar to the SPOT process just described, except that a nitridized oxide is used instead of using pad oxide to

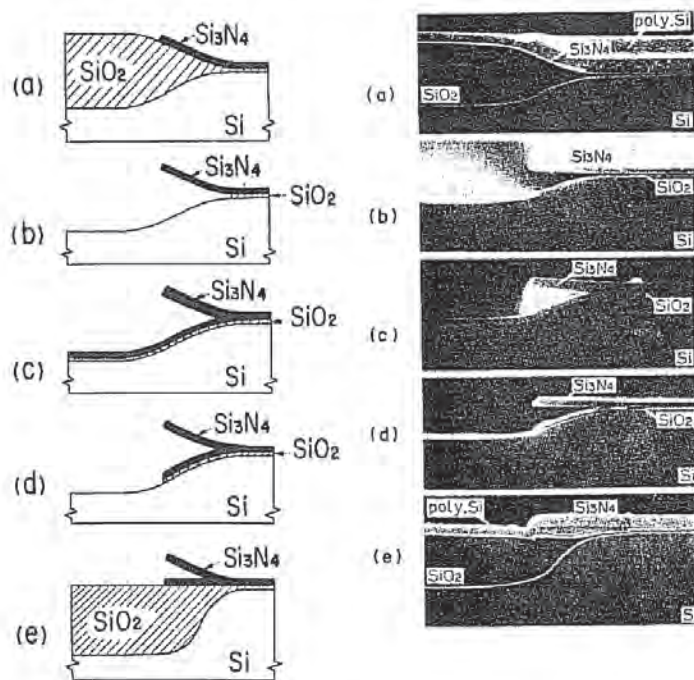


Fig. 2-27 Schematic of the SPOT isolation process.²⁵ Reprinted by permission of the publisher, The Electrochemical Society, Inc.

seal the silicon surface during growth of the oxide that forms the recess in the silicon substrate. Thus, there is less oxide encroachment into the active-device regions.

A thermal-oxide layer of ~20-nm thickness is grown and is then nitridized in an NH_3 ambient at 1200°C for seven hours. This provides a pad layer consisting of nitridized oxide that is capable of shutting off the diffusion path along the Si_3N_4 - SiO_2 -Si interface, with only a slight loss of the pad-oxide viscosity. The first LPCVD-nitride layer of 80-nm thickness and an undoped CVD oxide of 20-nm thickness are then deposited and patterned by means of plasma etching (Fig. 2-28a). The exposed nitridized oxide under the CVD nitride is etched away in a hot phosphoric bath. At this point the first field oxide is grown (450 nm thick), with the bird's beak largely suppressed by the nitridation-enhanced sealing ability. This field oxide is then chemically etched away to provide a moat of well-controlled depth and sidewall angle. The silicon surface is free of dry-etch damage and has a short nitride overhang (Fig. 2-28b).

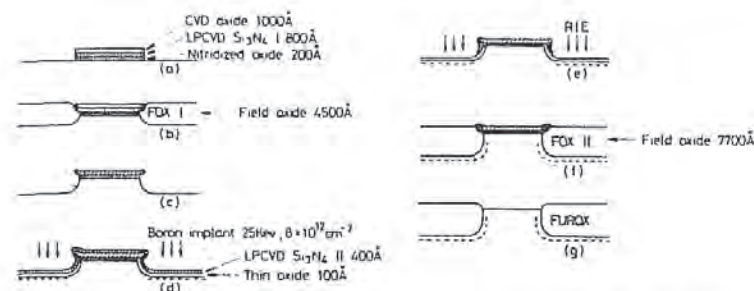


Fig. 2-28 Major process steps of the FURIX isolation technology.³⁹ (© 1988 IEEE).

A second pad-oxide layer (10 nm thick) is then grown and covered with a second CVD nitride (40 nm thick). After the self-aligned boron channel-stop implant is performed, the second nitride layer is etched anisotropically, so that it still remains on the sidewall for oxidation masking (Fig. 2-28c). After the second field oxide has been grown, a defect-free, near-zero bird's-beak, fully recessed oxide is produced with fairly good planarity. The narrow-width effects observed in conventional LOCOS are also reported to be significantly reduced. Figure 2-29 shows a successfully fabricated 770-nm-thick field oxide using FURIX, with a 1.1- μm isolation width and a bird's beak of about a 0.15 μm . A similar process, named PROXIS, has also been reported.¹²⁷

2.5.4 OSELO II

Another recessed selective-oxidation-based isolation process that reduces the bird's-beak length through lateral sealing of the silicon surface with a nitride layer is called OSELO II.⁴⁰ This process is designed to achieve isolation at spacings as small as 0.7 μm with low parasitic-junction capacitance. Conventional patterning of the masking layer, which here consists of a 50-nm pad oxide and a 240-nm CVD-nitride film, is followed by a surface boron implant (Fig. 2-30a), and a second CVD nitride layer (30 nm thick)

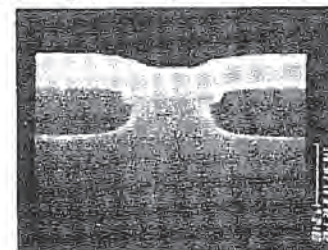


Fig. 2-29 SEM cross sectional photograph of the FURIX structure.³⁹ (© 1988 IEEE).

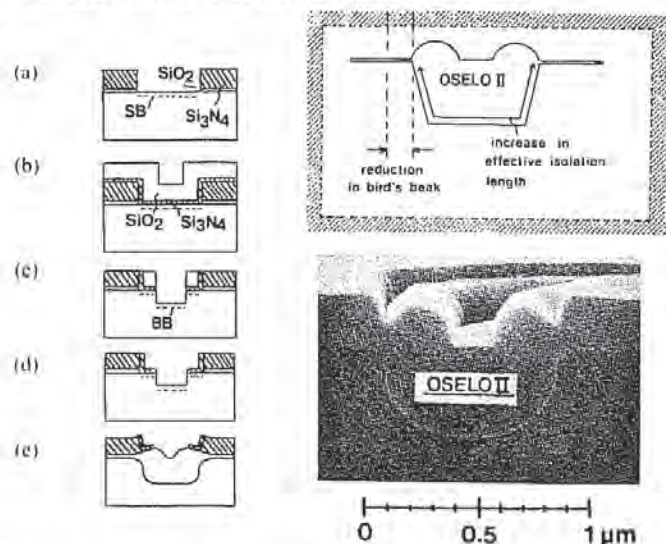


Fig. 2-30 (a) Fabrication process steps for OSELO II. (b) SEM photograph of a cross section of OSELO II.⁴⁰ (© 1988 IEEE).

– which prevents bird's-beak intrusion into the active region – and a 200-nm CVD SiO₂ film (Fig. 2-30b).

Dry etching removes this second nitride/oxide film everywhere except along the sidewalls of the first nitride layer, creating an offset structure of sidewall spacers. With the spacers used as a mask, 0.2 μm deep grooves are etched into the exposed Si with a dry etch that minimizes etching of the spacer materials (Fig. 2-30c). This is then followed by a second channel-stop boron implant that penetrates the silicon at the bottom of the groove (Fig. 2-30d), but is prevented from reaching the area that received the surface implant by the nitride/oxide sidewall spacer. Following the *bottom* implant, the CVD oxide remaining on the spacer is removed. Finally, a 550-nm-thick oxide is grown over the field region by means of wet-oxidation step for 120 minutes at 1000°C (Fig. 2-30e).

Since dry etching is used to form the grooves in the OSELO II process, depth control is better than in the SWAMI process, which uses KOH etching. The surface implant keeps the sidewalls of the etched groove from being inverted, while the bottom implant prevents inversion of the bottom of the groove. The effective isolation length of the OSELO II structure (defined as the length along the SiO₂-Si interface under the thick oxide) is 1.3 μm, even though the isolation-mask dimension is only 0.8 μm. Thus, a long, effective-isolation length is achieved. However, OSELO II was also shown to generate dislocation defects during oxide growth.

2.6 NON-LOCOS ISOLATION TECHNOLOGIES, I: (TRENCH ETCH AND REFILL)

Refilled trench structures have been developed for use in four primary VLSI and ULSI applications:

- Replacement of LOCOS as the isolation technology for like devices within the same tub in CMOS.
- Isolation of bipolar devices.
- Isolation of *n*-channel from *p*-channel devices and prevention of latchup in CMOS.
- Use as trench-capacitor structures in DRAMs.

Trench/refill approaches for isolation applications fall into the following three categories: shallow trenches (<1 μm), moderate depth trenches (1-3 μm), and deep, narrow trenches (>3 μm deep, <2 μm wide). *Trench-capacitor* structures are normally fabricated with narrow, deep trenches. The next sections will describe the fabrication details of the three trench types, as well as the chief uses of each type. (Trench-capacitor structures will be considered in detail in chap. 8.)

2.6.1 Shallow Trench and Refill Isolation

Shallow, refilled trenches are used primarily for isolating devices of the same type, and hence they can be considered as replacements for LOCOS isolation.

2.6.1.1 BOX Isolation. The *buried-oxide* (or BOX) isolation technology uses shallow trenches refilled with a CVD-SiO₂ layer that is then etched back to yield a planar surface. While the use of shallow trenches makes this process less difficult to develop than the deep-trench etch processes, the shallow trenches are also much less effective in preventing latchup and isolating *n*-channel from *p*-channel devices in CMOS. BOX isolation can be considered as a replacement isolation technology for LOCOS, but one that must be supplemented with another isolation method in CMOS for latchup protection.

In the basic BOX technique (Fig. 2-31),⁴¹ shallow trenches (0.5-0.8 μm deep) are anisotropically etched into the silicon substrate through dry etching. Next, a CVD oxide is deposited onto the wafer surface and is then etched back so that it remains only in the recesses, its top surface level with the original silicon surface. Etchback is performed by etching photoresist and SiO₂ at the same rate. The top surface of the resist layer is highly planarized prior to etchback through the application of two layers of resist, and flowing the first of these layers. Active regions are those that were protected from etch when the trenches were created. This technique has the advantages of eliminating the bird's beak of LOCOS and of providing a planar surface.

The basic BOX process has several drawbacks, including the following:

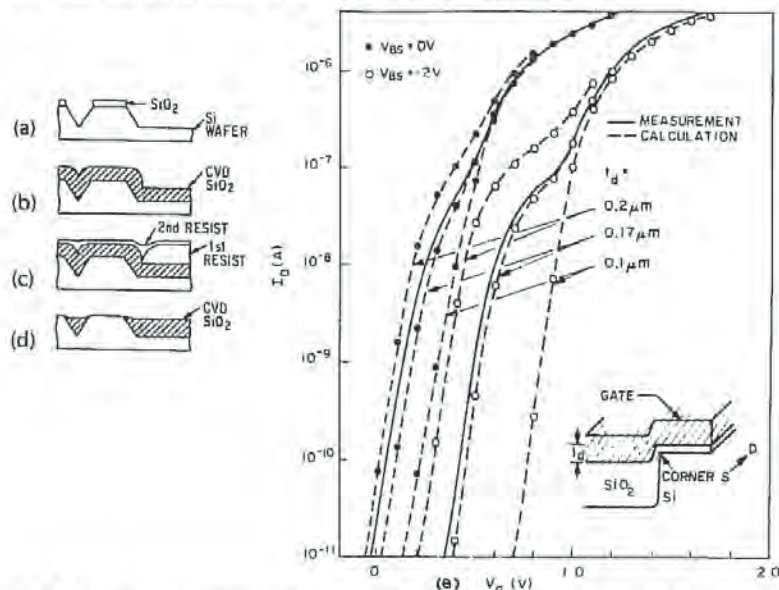


Fig. 2-31. The BOX isolation technique. (a) Si is etched using an oxide transfer mask. (b) Oxide is deposited by CVD. (c) Two layers of resist are applied, with the first being flowed. (d) The oxide is planarized using an RIE etchback technique which etches oxide and resist at the same rate.⁴⁶ (© 1983 IEEE). (e) Measured and simulated drain current versus gate potential for devices having a downward step in the field oxide (t_d).⁴³ (© 1983 IEEE).

- Void formation can occur if the CVD oxide is deposited into trenches that are narrower than about $2 \mu m$ (see chap. 4 for additional details about this CVD-deposition effect). A high-temperature-SiO₂ CVD-deposition step that gives conformal coverage without void formation has been reported to prevent this problem.⁴²

- Inversion of the silicon at the sidewalls of p -type active regions may occur (see the following section which defines isolation in CMOS and gives further details of this problem).

- The field-oxide profile following the planarization etch varies among different areas. (Various approaches to eliminate this problem have been reported, see the following discussion on improvements to the basic BOX approach).

- In practice, the resist layer ends up being not perfectly planarized, even when a two-layer-resist technique is used. This is especially true if there is a wide range of active-area sizes and spacings. To ensure that the CVD SiO₂ is

removed from all of the active areas at the conclusion of the etchback step, one must overetch of the field oxide by about 100 nm. The field oxide is thus etched below the active-area surface, exposing a portion of the active-area sidewall (Fig. 2-31b). This can lead to sidewall and edge-parasitic conduction (observable in the subthreshold-device characteristics), as well as to a higher electric field in the gate oxide at the active-area edge.

The electric-field concentration at the sharp-corner region of the active transistor reduces the threshold voltage of the corner region, and this part of the device turns on at a lower voltage than does the interior portion of the device.⁴³ The problem becomes worse if there is a downward step in the field oxide, as can occur in BOX isolation. The larger the step, the lower the threshold voltage, and unwanted subthreshold conduction begins at progressively lower values of V_{GS} .

2.6.1.2 Modifications to Improve BOX Isolation. Modifications to improve the basic BOX process include eliminating sidewall inversion of p -active-device regions through a shallow-angle implant (7° to the sidewall) of boron (see Fig. 2-32a).⁴⁴ This increases the threshold voltage of the parasitic n -channel that could occur along the sidewalls. Alternatively, a CVD borosilicate glass (BSG) layer may be deposited into the trench as a boron-diffusion source.⁴⁵ The BSG is etched away from the regions of the trench where the boron doping is not desired (Fig. 2-32b).

Several methods have been proposed to solve the problem of varying field-oxide depths. The first involves the use of an additional masking step.⁴⁶ The second employs LOCOS in the large-area field regions together with BOX-isolation between closely spaced devices.⁴² The third uses a thin, sputtered SiO₂ layer between double photoresists (as an etchback layer) and a polysilicon film above active regions (as an etchback buffer). This process is called *planarization with resist/oxide and polysilicon* (PROPS). This process offers a uniform field-oxide thickness in both narrow and wide field regions.⁴⁵

To avoid exposing the sidewalls of the active region, a *buried-oxide with etch-stop process* (BOXES) has been proposed (Fig. 2-33). The process uses a Mo/nitride etch-stop layer and passivates the recessed-silicon surfaces with a thin thermally-grown oxide before the CVD oxide is deposited. The oxide is grown by means of wet-hydrogen oxidation (i.e., oxidation in an ambient of a small percentage of O₂ in H₂) since Mo is not oxidized in this ambient.⁴⁷

Advanced BOX processes are reported to be capable of isolating MOS and bipolar transistors of the same type (i.e., n -channel devices) with submicron spacings. It appears, however, that their use will still require either a large spacing, a deeper trench, or an SEG-based technique between n^+ and p^+ regions in CMOS circuits to avoid latchup and the inversion of parasitic field devices between n^+ and p^+ structures (see the following sections on deep-trench isolation and SEG, and see chap. 6 [the material on CMOS isolation issues] for reasons that support this latter statement).

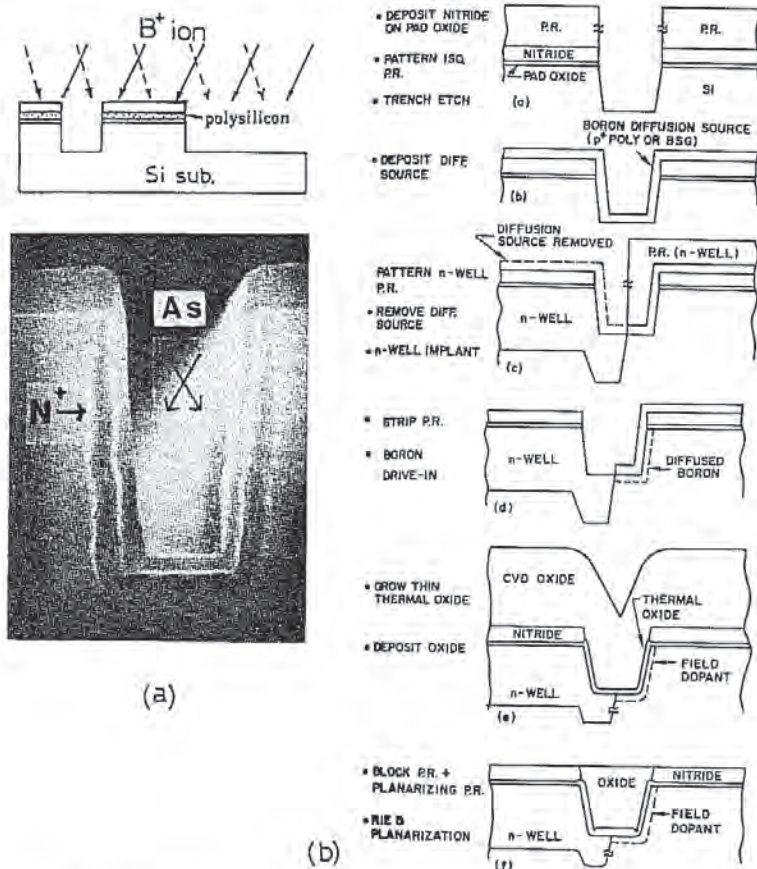


Fig. 2-32 (a) Schematic of oblique ion implantation and SEM cross section result of this procedure.⁴⁴ (© 1987 IEEE). (b) Shallow trench isolation process flow using a doped oxide as diffusion source of trench sidewall doping.⁴⁵ (© 1988 IEEE).

2.6.2 Moderate-Depth Trench and Refill Isolation

Two moderate-depth trench and refill processes will be described: *U-groove isolation* for bipolar ICs, and a *moderate-depth trench process to prevent latchup in CMOS*. Moderate-depth trenches are still easier to fabricate and refill than deep, narrow trenches.

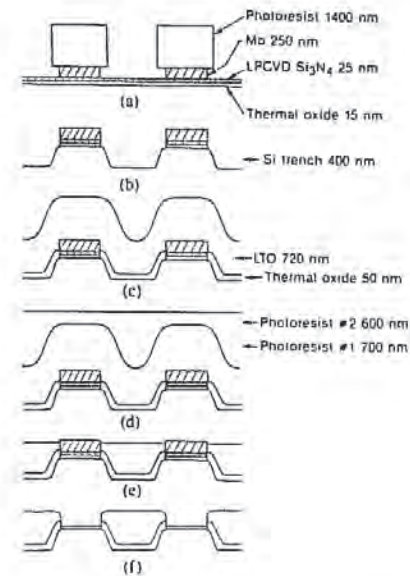


Fig. 2-33 Schematic process flow of BOXES process.⁴⁷ (© 1988 IEEE).

2.6.2.1 U-Groove Isolation.⁴⁸ U-groove isolation is a method of isolating bipolar devices that was introduced in 1981, and has reportedly found use in bipolar SRAMs. It actually combines the techniques of selective oxidation and trench refill of a unique, moderately deep groove to create the isolation structure. It requires less isolation spacing than fully recessed LOCOS and generates fewer dislocations under certain processing conditions.⁴⁹

A U-shaped groove is etched in the silicon field regions. The upper 0.5- μ m is etched anisotropically with a KOH/isopropyl-alcohol solution to yield sidewalls angled at 60°, while the remaining trench depth is dry etched to yield less sloped sidewalls and rounded bottom corners (Fig. 2-34a). With the nitride mask that was used to define the trenched areas still in place, 0.4 μ m of SiO_2 is thermally grown in the silicon trench. The trench is refilled with polysilicon, and the structure is etched back and thermally oxidized to yield a planar surface.

When the field oxidation is conducted at 1100°C, fewer dislocations are generated than in a comparable fully recessed LOCOS structure.⁴⁹ The shape of the U groove, which has rounded upper and lower corners and sloped sidewalls is credited with reducing the dislocation generation. Reduced bird's-beak and thinner SiO_2 field oxide allow this structure to be fabricated with smaller bipolar isolation spacing than does

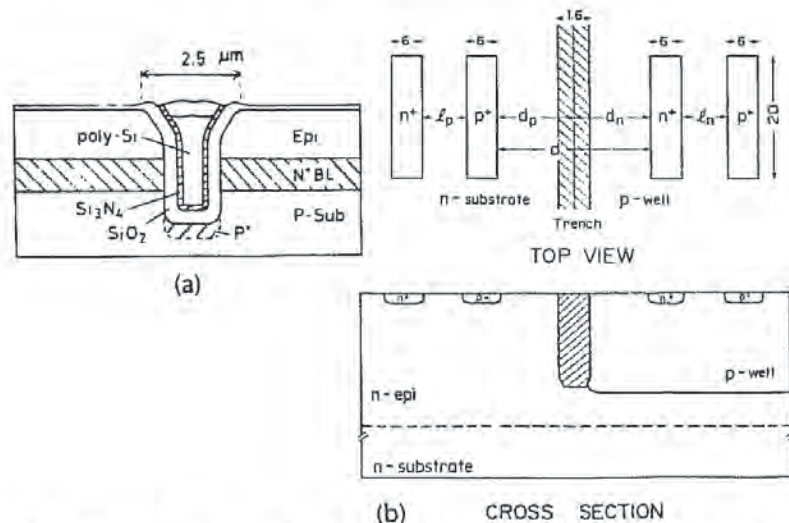


Fig. 2-34 (a) Cross sectional view of U-groove isolation structure.⁴⁸ Reprinted by permission of the publisher, The Electrochemical Society, Inc. (b) Schematic view of moderate-depth Toshiba trench process.⁵⁰ (© 1985 IEEE).

fully recessed LOCOS - for example, 3 μm -deep U grooves can provide a 2.5- μm -wide isolation spacing.

2.6.2.2 Toshiba Moderate-Depth Trench Isolation for CMOS. A relatively simple trench technology for preventing latchup and isolating n^+ and p^+ regions in CMOS circuits was described by Nitsu et al. in 1985.⁵⁰ It utilizes a shallow epi layer and shallow trenches (both can be less than 3 μm deep). This reduces the complexity of trench etch-and-refill.

The process begins with the formation of p -well regions in a thin (2-3 μm thick) n -epi layer on a heavily doped n^+ substrate. Moderately deep trenches at the borders of the n and p regions are created through dry-etching, refilled with polysilicon, and are etched back to give a planar surface (Fig. 2-34b).

Trenches are 2.5 μm deep and 1.6 μm wide, and have n^+ to p^+ spacings of 2 μm . Shallower trenches may be used (e.g., 1.4 μm deep) if the n^+ to p^+ spacings are increased to 5.8 μm . CMOS circuits having this isolation approach exhibit a holding voltage > 10 V, ensuring latchup-free operation in circuits operated with a 5-V power supply. Because the poly trench-refill process cannot be used to fill trenches of varying widths, CMOS circuits using this latchup-prevention isolation technology must still utilize another method (e.g., LOCOS) for isolating like devices in a common well.

2.6.3 Deep, Narrow Trench and Refill

Deep (greater than 3 μm) and narrow (less than 2 μm) trench structures have been utilized for three major purposes in advanced-IC fabrication: (1) to prevent latchup and to isolate n -channel from p -channel devices in CMOS circuits; (2) to isolate the transistors of bipolar circuits; and (3) to serve as storage-capacitor structures in DRAMs. Such trench structures are attractive for several reasons. First, trench isolation, because of its vertical wall, exhibits a significantly smaller width than a LOCOS isolation field oxide of the same depth. Second, if the trench is used together with a lightly doped epi layer on a heavily doped substrate, latchup in CMOS will be eliminated if the trench extends completely through the epi. Finally, without a deep trench, n^+ to p^+ spacing must be ≥ 8 μm to avoid latchup in most CMOS technologies, but since the trench width can be made very small (less than 1 μm), the packing density of the CMOS devices is significantly increased.

Despite its advantages, deep-trench technology has been hindered by fabrication complexity. Deep trenches are not only difficult to reliably manufacture from run-to-run, and over an entire wafer, but they also makes it difficult to meet the throughput demands of full production. In addition, inversion is possible because of positive charges at the trench surface or in the trench-refill material.⁹³ The problem is more severe for n -well CMOS because the n -well is biased at a positive voltage (see chap. 6). In this case the n -well acts like an electrode and tends to invert the p -type substrate opposite the n -well. If the trench surface is inverted, a conducting path will connect the n -well and the n^+ region, causing leakage between isolated n^+ regions and between the source and drain of the n -channel transistor. Latchup sensitivity will also be increased. In order for this problem to be avoided, the n^+ regions must be kept about 1 μm away from the trench (which reduces the packing density), or the substrate must be doped more heavily (which increases the junction capacitance and the body-effect factor, see chap. 5).

In this section we describe the sequence of process steps used to fabricate and refill such deep and narrow trenches. In subsequent sections we will discuss how these trenches are applied for isolation purposes. Finally, in chapter 8 (which deals with memory-device fabrication), we will examine trench storage-capacitor structures.

There are three steps in the trench fabrication sequence: reactive ion etching of the substrate, refilling with dielectric materials, and planarization (Fig. 2-35).

2.6.3.1 Reactive Ion Etching of the Substrate. The deep, narrow trenches that are etched into the silicon should have the following characteristics:

- The sidewalls should be smooth and slightly tapered, with an angle of $\sim 87^\circ$. There should be no undercutting of the mask used to define the trench width, no bowing, or kinks along the trench walls, nor should the etch process damage the sidewalls. Control of the sidewall profile is needed so that the trench can be refilled by CVD without the formation of voids. Undamaged sidewalls are necessary to avoid leakage along the trench. The angle of the sidewall cannot be

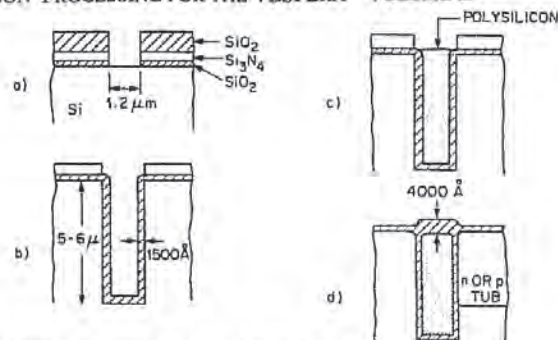


Fig. 2-35 Process sequence for forming a deep, narrow trench isolation structure. (a) Trench mask patterning. (b) Trench etching and oxide growth. (c) Polysilicon deposition to fill trench.⁶² (© 1984 IEEE).

too great, since this would enlarge the dimension of the trench width at the surface, negating the benefit of increased packing density.

- The bottom of the trench should be smooth, with rounded corners (ideally U-shaped, Fig. 2-36a), in order to maintain the integrity of the oxide that is grown on the trench surface after etch, and for keeping stress-induced defects from being generated during this oxide growth.⁵²
- The trench depth should be uniform both across the wafer and from wafer to wafer, especially in capacitor structures where capacitance values must exhibit a tight distribution.

A material must first be deposited on the silicon to act as the masking material that will define the trench width. This layer must remain intact during the dry etching of the trench. Various masking layers have been reported, but a combination layer of thermal oxide, CVD silicon nitride, and CVD oxide seems to be the most popular.⁵³

After patterning of the mask layer and removal of the resist, the exposed silicon-substrate regions must be cleaned to remove any native oxide that has formed during the resist-removal process. This is a key step, since any remaining native oxide will result in "grass" or "black silicon" at the bottom of the trench (Fig. 2-36b). A separate, preliminary dry-etch step (performed *in situ* with the trench-etch step) can be used to effect this cleaning.⁵³ It should also be noted that two other causes of "black silicon" have been identified: (1) sputtered material from the mask layer can redeposit on the Si and interfere with the Si etching at local spots, thereby causing the towers seen in Fig. 2-36b; and (2) the presence of oxygen in the Si can produce micromasks that lead to this condition.¹⁰⁵

Taper of the trench sidewalls is controlled by the deposition of a carbonaceous polymer layer during the etch process. This protects the sidewalls from the etchant species, thus preventing lateral etching. The polymer deposit is continually removed

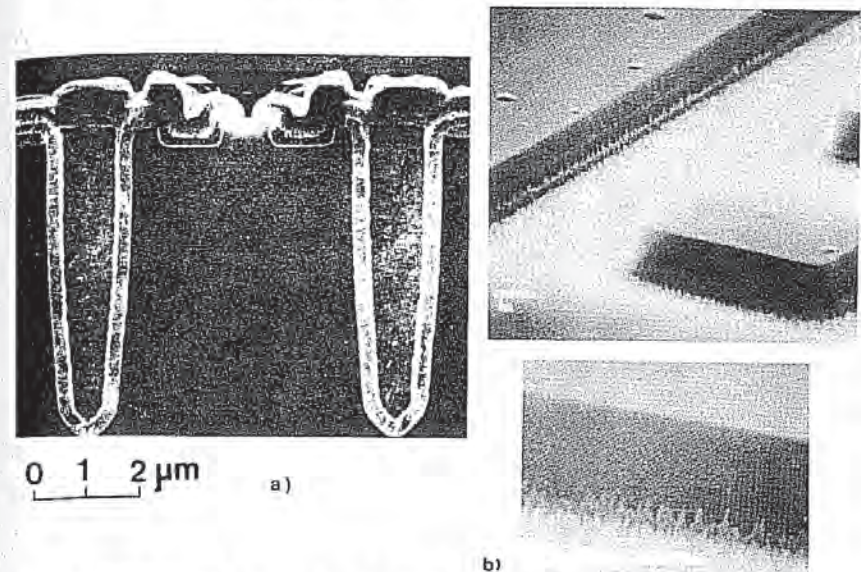


Fig. 2-36 (a) SEM photograph of a deep trench with smooth sidewalls and a rounded bottom. (b) SEM photograph of the bottom of a trench that has "black grass" or "towers", resulting from a residual oxide on the silicon surface before trench etching is begun.⁵³ Reprinted with permission of Solid State Technology.

from the bottom of the trench by energetic ions so that the etching reaction can continue downward. In one process, the polymer layer on the sidewalls builds up in a uniform fashion during the etching, causing the pattern geometry to shrink slowly (Fig. 2-37).⁵⁴ This results in a linear, positive slope on the sidewall without enlarging of the pattern geometry. Another report indicates that the temperature during the etch process must be kept constant to maintain smooth sidewalls with a uniform slope (implying that the polymer deposition rate may be strongly dependent on temperature).⁵⁶ Nevertheless, the deposited polymer layer must be removed prior to refilling of the trench, and this may be done using a wet-chemical procedure.^{53,54}

Different Cl-based etch gases have reportedly produced good trench-etch profiles. These include BCl_3/Cl_2 ,⁵³ $\text{H}_2/\text{Cl}_2/\text{SiCl}_4$,⁵⁵ and $\text{CHCl}_3/\text{O}_2/\text{N}_2$.⁵⁶ Good trench-wall profiles have also been obtained through the use of a triode-type dry-etch reactor and $\text{Cl}_2/\text{O}_2/\text{SiCl}_4$ ¹¹⁶ or CBrF_3/N_2 ¹¹⁷ gas mixtures.

The etch rate must be uniform throughout the entire etching step so that the same depth of trench can be created from wafer to wafer and from run to run. This is apparently possible, provided only one width of trench is being etched. Because etch

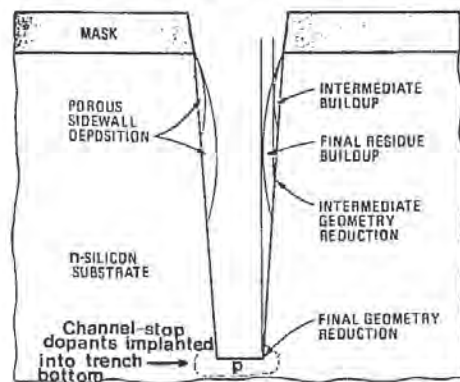


Fig. 2-37 The etch products in one trench etch technique are selectively deposited on the trench sidewalls to protect them during deep etching and to produce a controlled slope on the trench sidewalls.⁵⁴ Reprinted with permission of Electronics.

rate varies depending on trench width, only one width can thus be used in a given trench process. This may be a limitation from the point of view of circuit or device design.

In order for a rounded-bottom profile to be produced, the dry-etch process conditions must change to isotropic near the end of the etch step.⁵³ Increasing Cl_2 flow is one way to do this. No barreling of the trench profile occurs because the sidewalls are still protected by the polymer film deposited previously. Two other reports that review all of the details of trench etching are found in references 106 and 107. Note that channel-stop dopants are implanted into the trench bottom after etching and prior to refilling.

2.6.3.2 Refilling the Trench. After trench etching, refilling of the trench begins by removing the carbonaceous sidewall polymer, as well as some of the silicon from the surface of the trench.⁵² S. Samukawa et al. have reported that the leakage current of bipolar transistors isolated by trenches is significantly reduced when this



Fig. 2-38, Cross-sectional SEM photograph of a convex corner after oxidation in dry O_2 ambient.⁶⁰ (© 1987 IEEE).

deposited layer is removed with a combination of a wet HF solution etch and a dry O_2 etch.⁵⁷

Next, a sacrificial-oxide layer (e.g., 50 nm thick) may be thermally grown on the sidewalls. The growth and etching of this oxide can remove silicon damage caused by high-energy-ion bombardment during the trench-etch step.⁵⁸ In addition, the growth and removal of the oxide can round any sharp, concave corners of the bottom of the trench, as well as the sharp convex corner at the trench-oxide/gate-oxide point of connection (Fig. 2-38).^{59,60,61} Trench capacitors were reported to exhibit much lower leakage currents when fabricated with such a rounded-off oxidation. The removal of the sacrificial oxide is normally done with a dilute HF solution, but removal in a gas-phase HF ambient has also been suggested.

If the trench is to act as an isolation structure, a thicker thermal oxide (~100-150 nm) is usually used than if a storage capacitor is to be formed.⁶² In trench capacitors, a 15-25-nm thin thermal oxide,⁵⁴ a combination thermal-oxide/CVD-nitride,^{53,63} or even a combination thermal-oxide/nitride/CVD-oxide layer⁶⁴ of about the same thickness is used. Herb et al. reported that the combination oxide/CVD-nitride layer produced capacitor structures with a higher breakdown voltage, as well as a tighter breakdown-voltage distribution than those produced by the oxide layer alone.⁵³ Use of silicon nitride is advantageous because its higher dielectric constant (6-7) produces a trench structure with increased capacitance.

The thermal-oxide growth conditions that generate the fewest number of defects in fully recessed LOCOS isolation structures also generally produce oxide films in deep trenches with fewest defects. Trenches with rounded bottom corners, and higher-temperature oxidation conditions (>1050°C) have given the best results.⁵² The growth rate is reduced on concave surfaces (Fig. 2-39), an effect that becomes more pronounced as the degree of concavity is increased and the radius decreased.⁶⁵ This accounts for the greatly reduced oxide growth at the sharp corners of a trench bottom.

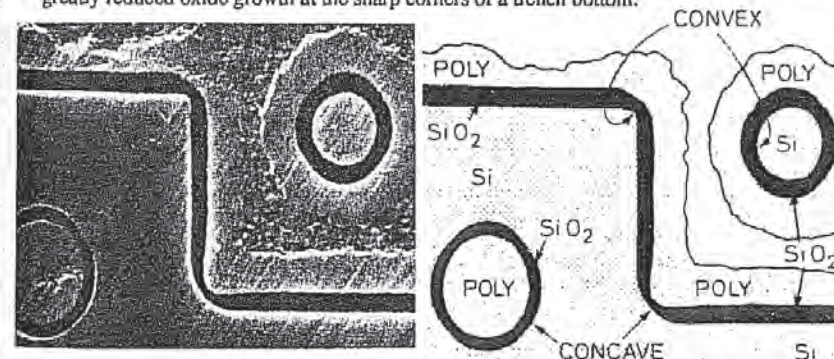


Fig. 2-39 A thermally grown oxide on a concave surface is thinner than that on a convex surface. Similarly, the concave corner also grows thinner oxide than the convex corner.⁶⁵ (© 1985 IEEE).

For trench-refill, the next step, various materials have been used, depending on the application. In *trench-isolation structures*, atmospheric-CVD polysilicon,⁶⁶ LPCVD polysilicon, CVD SiO₂, and selective, epitaxially-grown (SEG) silicon⁶⁷ have reportedly been used. In *trench-capacitor* applications, the charge is stored either on the edge of the trench facing the substrate or on the edge of the material inside the trench, and thus the one of the capacitor plates must reside within the trench. Consequently, polysilicon (typically heavily doped with phosphorus) is used to refill trench capacitors. (Additional information on trench-capacitor structures is presented in chap. 8.)

With the exception of the SEG process, the refill process proceeds by means of a conformal covering of the trench surfaces and the original substrate surface (see Vol. 1, chap. 6 for a discussion of conformal-CVD deposition). In deep, narrow trenches, the deposition occurs in four distinct stages (Fig. 2-40):⁶⁶

1. Initial conformal stage.
2. Seam-formation stage.
3. Seam-closure stage.
4. Cusp-planarization stage.

At the initial cusp-formation stage, voids can begin to form as the surfaces of the two opposite sidewalls meet centrally to close and fill the trench. Voids that form are of the order of the roughness of the two meeting surfaces. In some cases there can be "healing" or refill of these small voids with continued trench fill, since reactant gas apparently has access through the seams. This helps to explain why slightly angled trench sidewalls aid in filling high-aspect-ratio trenches in a void-free manner. The cusp that forms at the central regions of the trench is associated with the closure of the seam. Self-planarization of the cusp occurs with extended overfill. If the refill step is ended at the point of exact fill, the thickness of the film on the top surface will be about half that of the trench width.

If trenches of widely varying widths are filled, the narrow ones must be well overfilled in order for the wider ones to be filled completely. Thus, the thickness of the top-surface-deposited film will vary, making planarization very difficult. This is one reason why deep trenches refilled with CVD polysilicon or SiO₂ cannot replace LOCOS as an isolation technology for field regions of varying widths. SEG refill, however, does not suffer this limitation.⁶⁷

2.6.3.3 Planarization after Refill. After trench refill, the dielectric material deposited on the horizontal surface of the wafer must be etched back until the nitride/oxide mask film is exposed. This can be done in one of two ways: by directly removing the layer with a fully anisotropic etch step, or by applying a resist film over the wafer surface, and etching to remove the resist and the dielectric film at the same rate. In either case, the etch step must be carefully controlled to prevent the notch (associated with the trench at the top of the filled layer) from being greatly increased or exaggerated due to removal of some of the material from the filled trench.

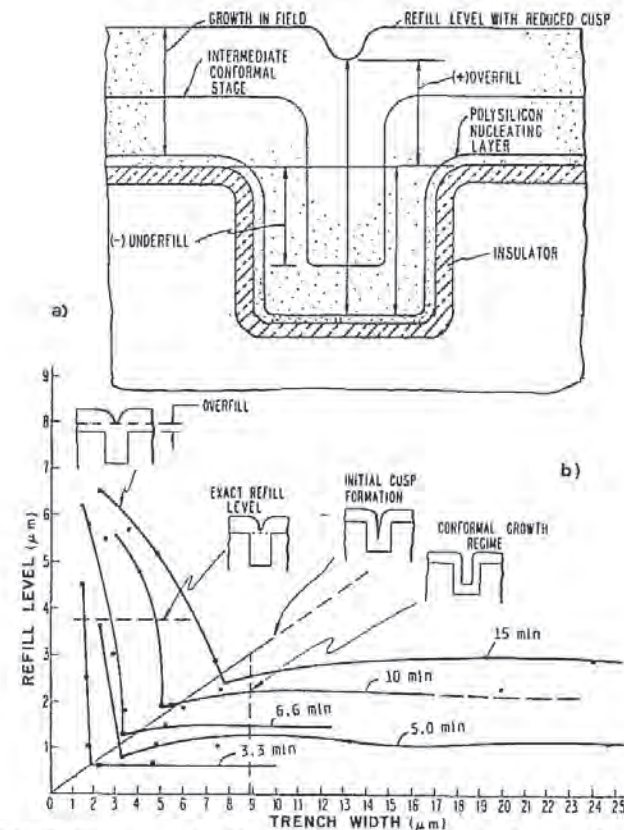


Fig. 2-40 Refilling a trench with polysilicon. (a) Schematic diagram of trench cross section showing experimental quantities measured. (b) Refill level versus trench width plotted for experimental refill times. This figure also shows the four phases of the deposition process.⁶⁶ Reprinted by permission of the publisher, The Electrochemical Society, Inc.

One novel way to improve the control of planarization is to deposit a buffer layer of CVD SiO₂ on top of the nitride mask before etching the trench. Later, during poly etchback, this CVD SiO₂ acts as an etch-stop. In addition, if a buffered-HF solution is used to strip the CVD SiO₂, the nitride mask is completely preserved. Thus, it can be ensured that all poly residue is removed and an undamaged nitride surface remains. Note that if polysilicon is used for refill, its top surface must be oxidized to complete the

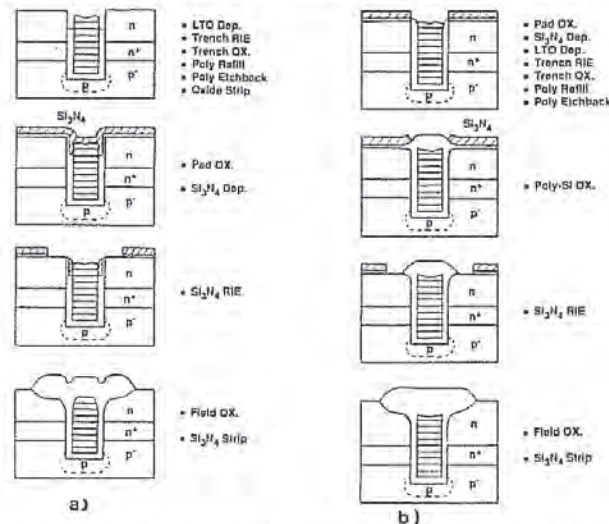


Fig. 2-41 Trench isolation process with LOCOS isolation. (a) Original deep trench process. (b) New planarized deep-trench process.⁶⁸ This paper was originally presented at the Fall 1988 Meeting of The Electrochemical Society, Inc. held in Chicago, IL.

process sequence. An example of a deep-trench isolation process combined with LOCOS for bipolar circuits is shown in Fig. 2-41.⁶⁸

2.7 NON-LOCOS ISOLATION TECHNOLOGIES, II: SELECTIVE EPITAXIAL GROWTH (SEG)

The use of *selective epitaxial growth* (SEG) for isolating bipolar, CMOS, and BiCMOS devices is a relatively new development. The basic concepts of SEG (see Vol. 1, chap. 5) were established in 1962, but the full potential of the concept could not be implemented until reduced-pressure epitaxial reactors were developed in the late 1970s. The effort to apply reduced-pressure SEG to isolation structures began shortly thereafter.^{69,70}

Low-pressure epitaxy has overcome some of the problems of atmospheric SEG, such as:

- Attack of the sidewalls of oxide openings, which takes place as these openings are refilled with SEG.
- Local loading effects that produce poor thickness uniformity in the SEG layers.

- Loss of selectivity, which deposits epitaxial Si on single-crystal Si surfaces, and polysilicon on other surfaces, such as oxide.

But remaining problems with SEG still include junction leakage, sidewall inversion, and corner faceting. *Junction leakage* is caused by sidewall-interface defects. Their generation is reduced by a lowering of the deposition temperature.⁷⁰ *Sidewall inversion* can be avoided by proper doping control.⁷¹ It is also argued that the doping levels in the channel regions of submicron MOS devices will be so high that sidewall inversion will disappear or be minimal. *Corner faceting* along the $\langle 100 \rangle$ direction (Fig. 2-42) can be minimized by optimizing the SEG growth conditions and device-layout geometry.⁷⁰

A number of different methods have been proposed for accomplishing isolation by SEG. These may be categorized as follows.

2.7.1 Refill by SEG in Windows Cut into Surface Oxide

Openings are etched in a surface oxide down to the silicon substrate and then filled by means of SEG, until a planar surface is achieved (Fig. 2-43).^{71,72,73} This produces an isolation structure with no bird's beak, and with a field-oxide thickness that does not depend on the width of the oxide space.⁷⁴ In addition SEG can provide a planar surface (except for facets), and the channel-stop region is removed from the source and drain regions (so that boron encroachment into the active regions is avoided). Furthermore, isolation structures with various widths can be planarized. Electrical isolation exists not only between wells, but also between devices (i.e., there is no need to use LOCOS to provide device-to-device isolation). Another bonus of this method is that it allows oversize contacts to be made to the source and drain regions of CMOS devices. It was noted in an earlier section that such contacts cannot be made when conventional



Fig. 2-42 Corner faceting problem that occurs during selective epitaxial growth in oxide windows.⁷⁰ Reprinted with permission of Solid State Technology.

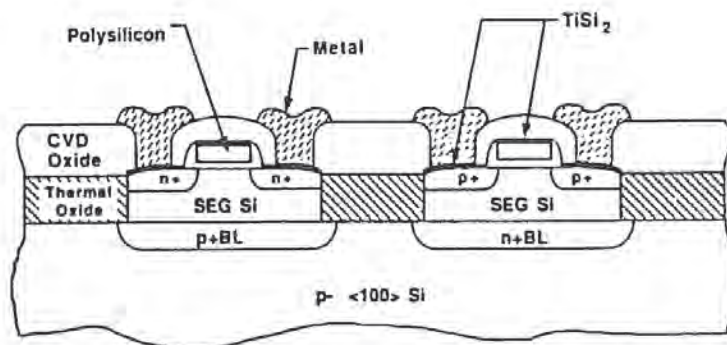


Fig. 2-43 Schematic cross section of a twin-well SEG CMOS technology.⁷¹ (© 1987 IEEE).

LOCOS is used with CMOS (because of the bird's-beak structure), and that this restriction imposes a packing-density penalty.

The SEG isolation process has several drawbacks:⁷⁴ (1) facets still occur (even under conditions that minimize their depth), which produce depressions in the silicon surface that may be very hard to clean out when filled with overlying film material. Chemical-mechanical polishing has been proposed for facet removal from the top silicon layer; (2) the sharp transition between the field oxide and the gate oxide at the edges of the device causes a penetration of the electric field from the field region to the active-channel area (Fig. 2-44). This lowers the V_T of the device (a phenomenon known as the *reverse narrow-width effect*). The problem may also be reduced as the doping is increased in the channel regions in submicron devices; (3) channel stops along the oxide sidewalls are needed to prevent device leakage caused by inversion along the sidewalls.

2.7.2 Simultaneous Single-Crystal/Poly Deposition (SSPD)

Here again, openings in the surface oxide are etched down to the silicon substrate. Silicon is then deposited non-selectively in an epi reactor.^{75,76} This deposits polysilicon over the oxide regions, but produces single-crystal silicon in the exposed single-crystal areas. This approach is known as *simultaneous single-crystal/poly deposition* (or SSPD).

2.7.3 Etching of Silicon Trenches and Refilling with SEG to Form Active Device Regions

Recesses with vertical sides are etched into the silicon substrate by means of an anisotropic dry etch and an oxide transfer mask. An insulating layer is then formed on the recessed-silicon surfaces (e.g., through thermal oxidation⁷⁷ or through a thermal oxide and CVD nitride⁷⁸), and the insulating layer is removed from the bottom of the recessed silicon by means of another anisotropic dry etch (Fig. 2-45). This provides a nucleation site for epitaxial growth, while simultaneously allowing the insulating layer to remain on the sidewalls. The recesses are then refilled with SEG until they are level with the original silicon surface.

2.7.4 Selective-Epitaxial-Layer Field Oxidation (SELFOX)

With this approach, SEG refills trenches ($\sim 1 \mu\text{m}$ deep) etched in silicon.⁷⁹ Yu et al. reported that the trench sidewalls and the surfaces of the active-device regions are masked with a CVD-nitride layer before the SEG step. SEG fills the trenches about halfway full and selectively oxidation forms $1 \mu\text{m}$ -thick field-oxide regions. No bird's beak is formed, no defect generation during field-oxide growth is observed, and isolation spacings as small as $1 \mu\text{m}$ wide are fabricated. Device regions are formed only in areas of the original single-crystal substrate (not in any epi-layer regions), and a planar surface is produced. No mention has been made of possible sidewall-inversion problems in n -channel active-device regions, nor has the structure been evaluated for its

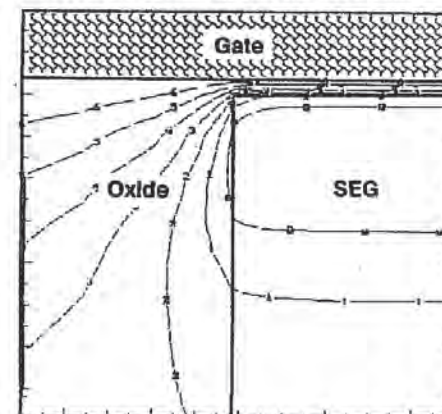


Fig. 2-44 The sharp transition between the field oxide and the gate oxide at the edges of the device causes a penetration of the electric field from the field region to the active-channel area. This lowers the V_T of the device (a phenomenon known as the *reverse narrow-width effect*).

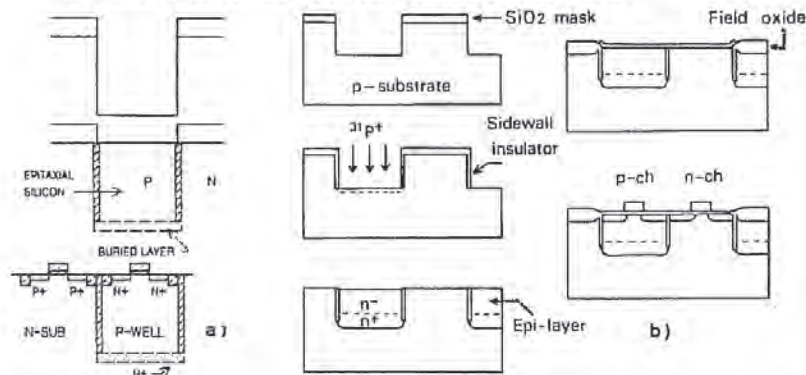


Fig. 2-45 (a) Schematic diagram showing SEREPI fabrication process, culminating in a SEREPI CMOS structure.⁷⁷ (© 1987 IEEE). (b) Selective epi and LOCOS isolation.¹³² (© 1985 IEEE).

effectiveness in preventing latchup in CMOS circuits. This approach appears to be another possible LOCOS replacement for isolating like-type MOS devices in a common tub.

2.7.5 SEG Refill of Trenches (as an Alternative to Poly Refill)

Trenches are again etched into the silicon substrate and are refilled by means of SEG.⁶⁷ In this case, however, the SEG material is not used to form device regions, but instead acts merely as filler for the trenches that are the isolation structures between active regions. Note that the sidewall oxide of the trench must again be removed from the trench bottom to provide a site for epitaxial nucleation. Epitaxial growth is terminated when planarity is achieved.

For refill of trenches the advantages of SEG over CVD SiO₂ or CVD poly are that trenches of widely varying widths can be filled simultaneously, there is no void formation, and the refill process is insensitive to trench profile. Trenches can also be refilled to the top without the need for overfill and etchback, which must be done with polysilicon refill.

2.7.6 Epitaxial Lateral Overgrowth (ELO)

In this approach, totally isolated silicon-on-insulator structures are fabricated by means of epitaxial-silicon overgrowth (Fig. 2-46).⁸⁰

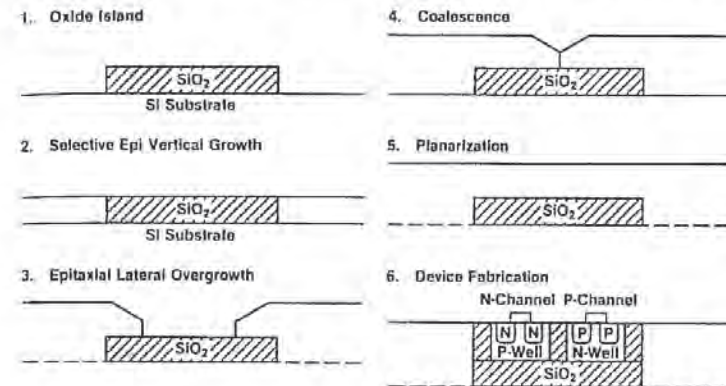


Fig. 2-46 Silicon on insulator by epitaxial lateral overgrowth.

2.8 MISCELLANEOUS NON-LOCOS ISOLATION TECHNOLOGIES

2.8.1 Field-Shield Isolation

An alternative approach for forming submicron isolation structures that does not require the growth of a thick field oxide is the *field-shield* technology.¹⁰¹ A thin (50-nm-thick) thermal oxide known as the *field-shield gate oxide* is grown and a doped-polysilicon film (200 nm thick) called the *field-shield plate*, is deposited on it and is patterned to remain over the field regions. A CVD oxide is deposited over the field-shield plate (Fig. 2-47). With a 5-V gate voltage applied to the field shield, the silicon surface in the field regions does not invert.

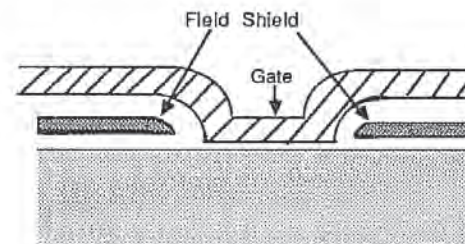


Fig. 2-47 Cross sectional view of a field-shield isolation structure.¹⁰¹ (© 1988 IEEE).

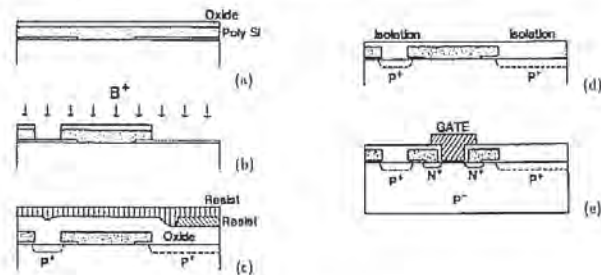


Fig. 2-48 Process sequence of BIPS isolation.¹⁰² (© 1988 IEEE).

Since there is no channel-stop implant and no thick field oxide, there is no lateral bird's beak encroachment or redistributed boron. The threshold voltage of the narrow-width active devices is not degraded, and in addition, this isolation structure exhibits good resistance to punchthrough, even for field-shield-transistor gate lengths of $0.6 \mu\text{m}$. The method's drawbacks are the nonplanarity of the isolation structure, and the need to provide a contact for each isolated field-shield plate. The latter problem could reduce the packing density of circuits and provide additional interconnection difficulties.

Another report which provides more extensive details concerning the fabrication and the isolation characteristics of field-shield structures for VLSI has recently been published.¹²⁶ The structure is self-aligned without any alteration in circuit design, and is also a self-aligned directly insertable structure. There is no penalty in the layout area and junction capacitance. In addition, the field shield protects the silicon surface from process-induced radiation during the active device fabrication and multilevel metal processing.

2.8.2 Buried Insulator between Source/Drain Polysilicon (BIPS)

Another novel non-LOCOS isolation structure is the so-called *buried insulator between source/drain polysilicon* (BIPS).¹⁰² The isolation structure is created by the refilling of CVD oxides in openings between source/drain polysilicon patterns through double-photoresist etchback (Fig. 2-48). In this process, $0.5\text{-}\mu\text{m}$ isolation spaces are achieved with no bird's beak, and narrow-width effects are almost entirely eliminated.

There are two concerns associated with this approach: first, when the polysilicon is etched to define the channel region, the etch must be stopped at the polysilicon/single-crystal interface; and second, there is no apparent way to easily create an LDD structure in the polysilicon source/drain transistors.

2.9 SUMMARY: CANDIDATE ISOLATION TECHNOLOGIES FOR SUBMICRON DEVICES

2.9.1 Basic Requirements of VLSI and ULSI Isolation Technologies

Isolation technologies for submicron CMOS and BiCMOS ICs will be selected based on how well they can meet the following requirements:

1. **Device-to-device isolation within a well.** The threshold and punch-through voltages of the field parasitic transistors must be large enough to prevent leakage between active-device regions through the parasitic field devices.
2. **Well-to-well isolation.** This type of isolation is needed to prevent leakage between parasitic MOS devices (see section 6.5 on CMOS isolation), as well as to prevent latchup of the parasitic bipolar-devices that exist in CMOS structures (see section 6.4 on latchup in CMOS).
3. **Minimum area.** Isolation structures need to be as small as possible, and they must be potentially scalable to even smaller sizes.
4. **Process simplicity.** Processes should be simple, transparent to the existing device-process flow, and resistant to contamination. Further, it should require a minimal number of additional masking steps.
5. **Active-device characteristics should not be degraded by parasitic effects.** For example, the isolation structure should not increase the device junction capacitance and body-effect sensitivity, nor should it decrease the drain breakdown voltage. Narrow-width effects caused by boron encroachment should be alleviated.

2.9.2 The Need for Planarity

A major limiting factor in the processing of submicron-feature-sized ICs will be the very small depth of focus of the optical steppers used to pattern the circuit features.^{83,104} In order to be able to obtain the maximum resolution from these steppers, everything within the field of view will have to be kept within a depth of focus of about $0.5 \mu\text{m}$. Accordingly, the imaging surface will need to be very flat. Loss of flatness in the field of view arises from several factors: resist-thickness differences at different locations, residual lens aberrations, autofocus errors, wafer nonflatness, and wafer-surface topography. To overcome non-planarity due to the last group of effects, use of device structures that result in a planar topography will be mandatory.

2.9.3 How the Various Isolation Technologies Meet the Requirements

• **LOCOS-based processes.** If a planarized semirecessed SILO or polysilicon-buffered LOCOS process is used, along with a channel-stop implant performed after the field-oxide growth, then most of the requirements can be met. Only the problem of field thinning remains, and the same limitation applies to fully recessed LOCOS processes. Advanced LOCOS-based processes, however, involve increased process complexity.

• **Trench-based isolation processes.** The modified BOX technique with selectively-doped sidewalls,^{42,45} has the potential of providing excellent planarity, as well as freedom from encroachment and field thinning. However, it requires two more masking steps than does a conventional LOCOS process, as well as three or four dry-etch steps (including one etchback step). Moderate- and deep-trench processes offer the benefit of producing latchup-immune or even latchup-free CMOS circuits, as well as dense isolation structures for bipolar devices. Deep-trench structures generally require significantly increased process complexity, and consequently result in potentially lower productivity (i.e., lower throughput and yield).

• **Selective-epitaxial-growth isolation processes.** SEG processes appear to have the potential of providing planar topology and small isolation spacing. In addition, compared to the other isolation approaches, SEG processes offer the benefits of process simplicity. However, they require the use of a relatively new technology, which still has some problems; further, a history of use in a production-line environment has not yet been established.

• **Miscellaneous processes.** The field-shield process creates a nonplanar surface, and every field-area polysilicon must be electrically connected to V_{DD} (which requires an area for contact, as well as an additional interconnection). The BIPs process looks interesting, but requires a polysilicon etch stop at the single-crystal silicon surface. This appears to be challenging for production applications.

2.10 SILICON-ON-INSULATOR (SOI) ISOLATION TECHNOLOGIES

Junction isolation is not suitable for high-voltage applications* because at supply voltages of ± 30 V junction breakdown occurs under reasonable doping levels and device-structure dimensions. Junction isolation is also ineffective in high-radiation environments due to transient photocurrents produced in pn junctions by gamma rays. For these applications, a preferred isolation is one that depends on completely surrounding devices with an insulator, rather than with a pn junction. Technologies that

* For this discussion the term *junction isolation* includes structures that are isolated by an oxide along the sidewalls and by a junction at the bottom.

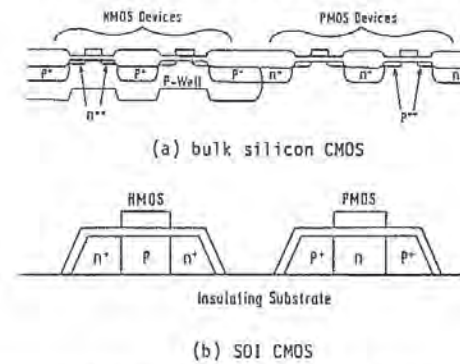


Fig. 2-49 Comparison of (a) structure cross sections and (b) process flows of bulk and SOI CMOS.¹²² (© 1986 IEEE).

provide this full isolation are collectively referred to as *silicon-on-insulator* (SOI) technologies.

SOI offers many additional advantages. In some cases it uses a simpler fabrication sequence and resultant cross-section compared to circuits fabricated on bulk silicon. This is clearly demonstrated in Fig. 2-49,¹²² which compares a mesa-isolated SOI CMOS process with a p -well bulk CMOS process. SOI also provides reduced capacitive coupling between various circuit elements over the entire IC chip, and in CMOS circuits latchup is eliminated. SOI reduces chip size and/or increases packing density, and minimum device separation is determined only by the limitations of lithography. Finally, it provides increased circuit speed, due to reductions in parasitic capacitance and chip size.

When an SOI technology based on a thin silicon film is used, two other important advantages can also be gained. A relatively benign surface topography (for step-coverage) is produced if device isolation can be achieved by a complete island, sloped-wall etch process of the thin silicon film. Second, since SOI isolation eliminates the parasitic field FET between adjacent devices, LOCOS processes are not needed.

As with all technologies SOI has its own disadvantages. To date, active-device regions in SOI technologies are poorer in crystalline quality than their counterparts in bulk silicon. In addition, the presence of an insulating substrate, or insulating layer, may complicate or prevent the adoption of effective defect- and impurity-gettering processes. Nevertheless, SOI's potential advantages are sufficiently attractive that development work in this area remains quite active.

2.10.1 Dielectric Isolation

The first technique developed for fully isolating the collectors of bipolar transistors with a layer of SiO_2 rather than with a pn junction was given the name *dielectric*

isolation (DI). This has been a production process for at least ten years, and circuits are available on DI wafers in both CMOS and bipolar technologies.

The fabrication steps used to produce DI wafers are shown in Fig. 2-50. The starting material is a bare n -type silicon wafer with a resistivity appropriate for the collector region of the transistor. The first step etches V-grooves in the back side of the wafer; these will become the isolation regions in the finished circuit. Using oxide as an etching mask, an orientation-dependent etching step is used to delineate these grooves ($\sim 20 \mu\text{m}$ deep for typical analog-circuit applications). Next, an oxide is grown on the grooved surface (to a thickness of $\sim 2\text{--}4 \mu\text{m}$), and this is followed by a layer of thick CVD polysilicon. This poly layer will become the mechanical support for the finished DI wafer and must therefore be $\geq 200 \mu\text{m}$ thick.

The starting wafer is then lapped from the top side until the silicon dioxide in the V-grooves is exposed. The only remaining single-crystal silicon material is that which is left in the isolated active-device islands. The active-device surface is then mechanically and chemically polished using a polishing procedure very similar to the final step used by vendors of silicon wafers. The wafer is then turned over, yielding regions of single-crystal silicon $\sim 8\text{--}10 \mu\text{m}$ thick that are completely isolated from one another by an oxide layer. Standard processing can then be used to fabricate bipolar transistors in these islands.

Although the DI process provides structures that will allow high-speed, high-voltage, and/or radiation-hardened circuits to be built on them, the process has several drawbacks. The fabrication sequence used to produce DI wafers is expensive. For example, even in 1984 the cost of the largest DI wafers was $\sim \$60\text{--}\$100$; high compared to the cost of processing an entire CMOS wafer on a conventional wafer which then was $\sim \$150\text{--}\$250$. Next, precise mechanical alignment of the wafer during the lapping

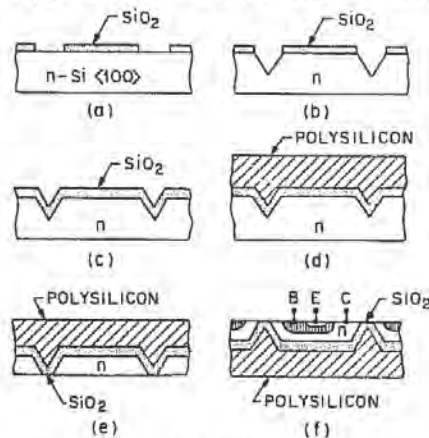


Fig. 2-50 Process sequence for dielectric isolation.¹³³ Reprinted by permission of the publisher, The Electrochemical Society, Inc.

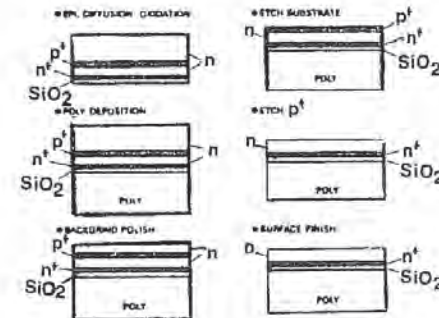


Fig. 2-51 FELT/SOI process sequence.¹²⁰ This paper was originally presented at the Spring 1989 Meeting of The Electrochemical Society, Inc. held in Los Angeles, CA.

process must be maintained to ensure that this step removes exactly the correct thickness of silicon everywhere on the wafer. Third, the polysilicon-deposition processes cause warping of the wafer due to the thermal-expansion coefficient differences between the polysilicon and the single-crystal silicon. Thus, as the wafer cools down from the polysilicon-deposition temperature, bowing can occur. During lapping, this produces some isolation islands in which the silicon is too thin, or some "islands" that are not completely isolated. Finally, the formation of the V-grooves at the front end of the process places limitations on the density achievable with DI technology.

A recent improvement on the conventional dielectric-isolation process is called *full-etchstop layer transfer/silicon-on-insulator* (or FELT/SOI).¹²⁰ A p^+ etchstop layer is formed on a lightly doped wafer, and a layer of lightly doped n -type material is epitaxially deposited to the desired thickness and doping concentration (Fig. 2-51). This n -epi layer will be the device layer when the substrate processing is completed. A CVD insulator is then applied and a thick polysilicon layer is deposited (as in the conventional DI process).

The original single-crystal silicon wafer is then lapped back to within a few mils of the p^+ etchstop layer. A doping-concentration-dependent etch removes the remaining silicon but stops at the p^+ etchstop layer (which is subsequently removed with another doping-concentration-dependent etch). The end result is a layer of silicon of known thickness and doping concentration fabricated on an insulating layer. In addition, since device isolation is not built into the material, this method offers more design flexibility than does the DI process — that is, device isolation does not have to be predesigned into the substrate, allowing reduction in the overall cycle time for new circuit designs.

2.10.2 Wafer Bonding

A planar process similar to the conventional DI process can be implemented with wafer bonding. In this method, two oxidized silicon wafers are fused together through a high-temperature furnace step. The following approaches have been developed to consummate the bonding:

- Two oxidized-silicon wafers are pressed together and subjected to an oxidizing ambient at 700°C.⁸¹ Bonding under these conditions is believed to occur as a result of the conversion of the gaseous oxygen between the wafers to SiO₂. This conversion is thought to create a partial vacuum that forces the wafers into intimate contact, bringing about the bonding chemical reaction. (This is believed to be a polymerization of silanol bonds to form a siloxane network).
- When a moderate voltage is applied between two silicon wafers, bonding is induced at temperatures of 1100-1200°C (Fig. 2-52).⁸² One advantage of electrostatic bonding is that since no mechanical force needs to be applied to press the wafers together initially, it does not introduce crystal defects into the silicon.
- One oxidized silicon wafer and one bare silicon wafer are cleaned with an H₂O₂-H₂SO₄ mixture. Treatment with an acid solution is then used to form an O-H group on the wafer surfaces. After drying, the wafers are placed face-to-face at room temperature in clean air. A self-adhesive contact is formed, and thereafter no weight or pressure need be applied to the contacted wafers. The bonding is then carried out by a heat treatment of 1100°C in a nitrogen ambient for four hours.⁸³

The SOI structure is also formed in a number of different ways. With one method, a lightly doped epitaxial layer is grown on a heavily doped substrate, and this layer is then thermally oxidized (i.e., to a thickness of 1-2 μm). Following bonding, a preferential etch is used to remove the heavily doped substrate, leaving a thin, lightly doped epitaxial layer above the thermally grown oxide.⁸¹ The active-device layer is

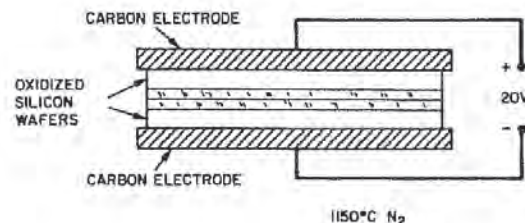


Fig. 2-52 Electrostatic technique for bonding silicon wafers.⁸² Reprinted by permission of the publisher, The Electrochemical Society, Inc.

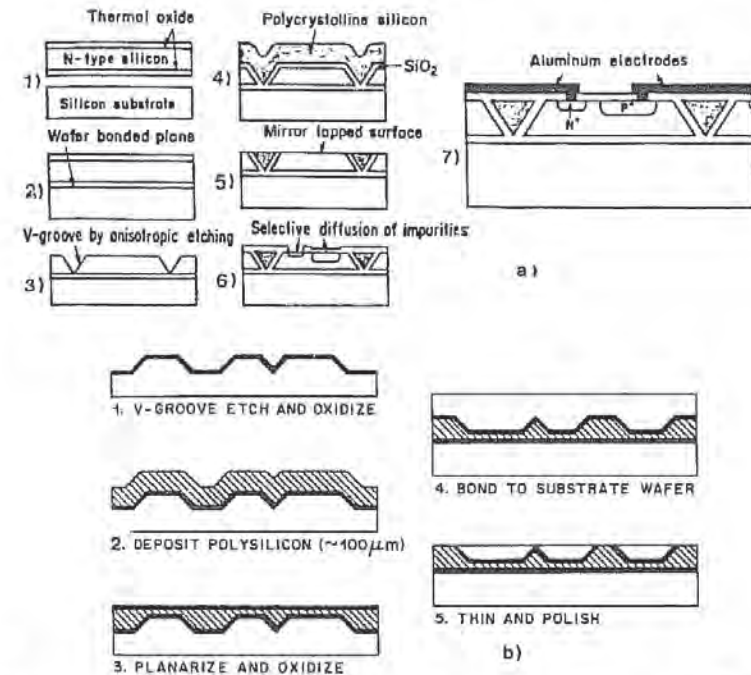


Fig. 2-53 Isolation structure formation in bonded silicon wafers. (a) SDB technique,⁸³ (© 1986 IEEE). (b) Electrostatically bonded technique.⁸² Reprinted by permission of the publisher, The Electrochemical Society, Inc.

determined by the thickness of the epitaxial layer. Alternatively, the heavily doped substrate is removed by lapping, and the remaining epi layer is polished to the desired thickness.⁸⁴

In a second method, the surface of the bonded silicon wafer is V-groove etched and oxidized. A polysilicon film, much thinner than that used in conventional DI, is then deposited, planarized, and oxidized (Fig. 2-53a).⁸³ This wafer is then bonded to another wafer. Finally, the single-crystal silicon of the wafer with the deposited polysilicon is thinned and polished to form dielectrically-isolated active regions.

In yet another approach, the upper part of one of the wafers is chemically polished until a 70-μm-thick Si layer remains (Fig. 2-53b). V-grooves are formed, and SiO₂ is regrown over the surface. Polysilicon is deposited to a thickness of 100 μm and the wafer is again polished until most of the polysilicon is removed and the 70-μm-thick region has been reduced to a thickness of 50 μm. The remaining poly is that which

remains in the grooves. An SiO_2 layer is then grown on the surface, and the wafers are ready for device processing.

In a report comparing the various SOI approaches, it was stated that bonded-wafer technologies exhibit very low leakage currents.⁸⁴ On the other hand, they are limited by the minimum thickness that can be achieved in the silicon and the minimum variation of that thickness across the wafer. This characteristic might still make SOI-by-wafer-bonding an attractive choice for advanced bipolar applications, as these require somewhat thicker silicon-device regions. In addition, it could be useful in some CMOS applications, such as those used for very low-power or high-voltage circuits, in which film thickness is not required to be minimized. A 1-kbit ECL SRAM fabricated on a bonded SOI wafer is described in reference 85.

A recently reported problem associated with the wafer-bonding process is that voids can occur at the interface between the wafers if a particle exists on the surface of either wafer (Fig. 2-53b). Such particles can keep two mechanically rigid surfaces from mating and bonding in the vicinity of the particle. The acute susceptibility of this bonding process to particles is reported to cause low production yield.¹²⁰

2.10.3 Silicon-on-Sapphire (SOS)

Another early SOI technology, begun in the 1960s, uses sapphire for both substrate and insulator (hence the name *silicon-on-sapphire*, or SOS). Sapphire has the advantage of being fairly well lattice-matched to silicon. Although there have been notable successes with SOS over the years, technology and material issues have prevented it from becoming a mainstream circuit technology. Recent advances in SOS, however, continue to preserve the viability of this material, and in the late 1980s it remained the only commercial, non-DI, SOI technology.⁸⁶ (More information about SOS can also be found in Vol. 1, chap. 5.)

2.10.4 Separation by Implanted Oxygen (SIMOX)

The creation of a buried layer of SiO_2 by implanting oxygen into silicon, a process known as SIMOX (for *separation by implanted oxygen*), is one of the major contenders for creating SOI structures.⁸⁷ The technique requires a high dose ($\sim 2 \times 10^{18} \text{ cm}^{-2}$) of oxygen (O^+) ions, as this provides the minimum concentration necessary to ensure that a continuous layer of stoichiometric SiO_2 will be formed by reaction of the oxygen with the silicon during the annealing process (Fig. 2-54).⁸⁸ The energy of the implant must also be high enough (150-180 keV) so that the peak of the implant is sufficiently deep within the silicon (0.3-0.5 μm). The wafer is normally heated to more than 400°C during the implantation process to ensure that the surface maintains its crystallinity during the high-dose implantation step (see Vol. 1, chap. 9 for more details on this topic).

A post-implant anneal is performed in a neutral ambient (N_2) for a sufficient time (3-5 hours) and at a high enough temperature (1100-1175°C) to form a buried layer of

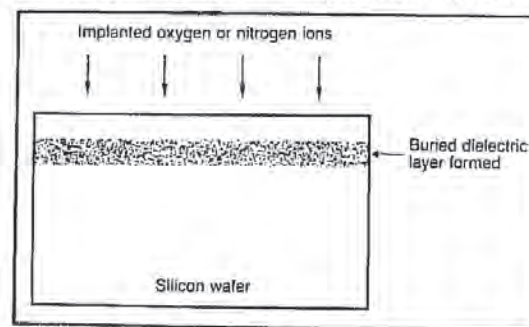


Fig. 2-54 Implanted oxygen ion process for the formation of a SIMOX wafer.⁸⁷ (© 1987 IEEE).

SiO_2 . The anneal step also allows excess oxygen in the surface silicon to out-diffuse, thereby increasing the dielectric strength of the buried-oxide layer.

After the anneal step, the crystalline-silicon surface layer is typically only 100-300 nm thick. As a result, an additional layer of epitaxial silicon is usually deposited so that single-crystal device regions $\geq 0.5 \mu\text{m}$ thick are available for fabricating devices. CMOS devices produced in such regions on SIMOX substrates have exhibited satisfactory operating characteristics, especially in cases in which submicron-CMOS transistors were evaluated.^{89,90}

SIMOX exhibits some advantages over other SOI technologies. The most important is that the technology is transparent to the manufacturing line since the fabrication of SIMOX-based circuits uses processing steps similar to those used in conventional IC manufacturing. Even the novel high dose implant of oxygen is a procedure that is understood by process engineers.

SIMOX does, however, have a number of disadvantages, including the following:

- The process requires the availability of special oxygen implanters. Beam accelerators with the ability to implant oxygen at 150 keV with a 10-mA beam current first became commercially available with the introduction of the NV-10 implanter by the Eaton Corp. The beam current of this machine puts it into the medium-current-implanter category. However, with such an implanter, an entire day is needed for implanting of a 50-wafer SIMOX lot. High-beam-current machines (100 mA) have recently been introduced (e.g., the Eaton/Nova-100), making the high-volume production of SIMOX wafers more feasible.
- The thickness of the buried-oxide layer is limited to about 0.5 μm . For some high-voltage applications, a greater thickness may be required. Furthermore, the relatively thin buried-oxide layer will increase parasitic capacitance, and can make SIMOX-based circuits somewhat slower than those based on an SOI process with a thicker oxide.

▪ The thickened device-silicon layer may not be free of defects, since the SOI epitaxial-seeding layer may still contain some residual damage from the implant step. (SIMOX substrates have been found to contain larger densities of threading dislocations than conventional bulk wafers. These defects considerably reduce the minority-carrier lifetime and increase the surface-recombination velocity.)¹¹⁸

Defect generation in the surface-silicon film and the deposited epi layer can occur from metal contamination sputtered from the beam-line and end-station by the primary oxygen ion-beam. The contamination can be as high as 1% of the primary dose. To combat this problem, some metallic parts in oxygen implanters have been replaced with graphite parts; however, this results in carbon contamination. During the anneal step, carbon precipitates out of solid solution, forming microdefects that degrade the subsequently deposited epi layer. For removal of the carbon precipitates, a shallow plasma etch step is suggested.⁹¹

▪ Implantation parameters and anneal schedules must also be appropriately chosen to provide optimum IC performance because the microstructure of the surface-silicon film is very sensitive to the oxygen dose and the post-oxygen-implant annealing temperature.⁸⁹ For example, a lower dose of oxygen results in a higher oxygen content in the silicon film and a higher density of oxygen precipitates at the silicon-film/buried-oxide interface following an 1150°C anneal. For oxygen doses of $2.25 \times 10^{18} \text{ cm}^{-2}$, thermal annealing at 1275°C annihilates the oxygen precipitates in the silicon film.

SIMOX is attractive for military electronic systems requiring high-speed, low-power, radiation-hardened devices that can operate reliably over a wide temperature range as well as for power IC and telecommunications applications which must integrate low-voltage control circuits with high-voltage devices. SIMOX is also a candidate for CMOS circuits with feature sizes below $0.8 \mu\text{m}$. Finally, SIMOX is well-suited for integrating bipolar and CMOS devices on the same chip. The manufacturability of SIMOX has been demonstrated, but although commercially available, the 1989 cost of SIMOX wafers is still quite high.

Recently, use of wafers on which no epitaxial layer is deposited following oxygen implantation and annealing (so-called *thin-film* SIMOX wafers) has been studied.^{113,114,115} Fabrication of MOS devices in such wafers offers several important advantages. Long-channel behavior (see chap. 5) can be preserved at gate lengths below half a micron because of the limited vertical extension of the depletion region into the thin film, and a reduced electric field at the drain junction in such *fully depleted SOI devices* is produced (eliminating the kink effect observed in thin-film transistors and lowering hot-carrier degradation). In addition, thin-film SIMOX processing is simpler than a bulk process because there is no need to form doped wells, and shallow junctions are automatically formed when the doping impurity reaches the underlying buried oxide.

Implanting a high dose of nitrogen and annealing at 1200°C to form SOI structures has also been investigated. Potentially, this has several advantages over the implanting

of oxygen. Hot-filament, high-current nitrogen-implant sources are easier to construct than are similar oxygen sources. The defect-laden layer that exists at the top Si/insulator interface for implanted oxygen is absent in the case of implanted nitrogen. Annealing of the as-implanted nitrogen atoms results in the formation of a top silicon layer, which is sufficiently thick that a subsequent epitaxial-growth step may not be needed, and a continuous buried silicon-nitride layer may be formed at about 40-50% of the dose necessary to form a buried SiO_2 layer with an abrupt top Si/ SiO_2 interface.⁹² However, the higher dielectric constant of silicon nitride may reduce the speed performance of ICs compared with those fabricated on SIMOX wafers.

2.10.5 Zone-Melting Recrystallization (ZMR)

In the ZMR process, an active-device layer is formed by recrystallizing a polysilicon layer that has been deposited on a thermally oxidized silicon wafer,⁹⁴ i.e., the grain sizes of the as-deposited CVD-poly layer are enlarged by the movement of a narrow molten zone. The most promising method for creating the molten zone is heating the substrate close to the melting point of silicon by an auxiliary means, and then using a graphite heating strip (Fig. 2-55)⁹⁵ or focused lamp⁹⁶ to supply the additional heat necessary to achieve melting. In a typical process, $0.5\text{--}2 \mu\text{m}$ of thermal oxide is grown, and $0.5 \mu\text{m}$ of polysilicon film is deposited over it. A $2 \mu\text{m}$ layer of SiO_2 is deposited to cover the poly. Next, the wafer is heated to 1100-1300°C by a lower heater so that radiative heating from a strip above the wafer can then produce a molten

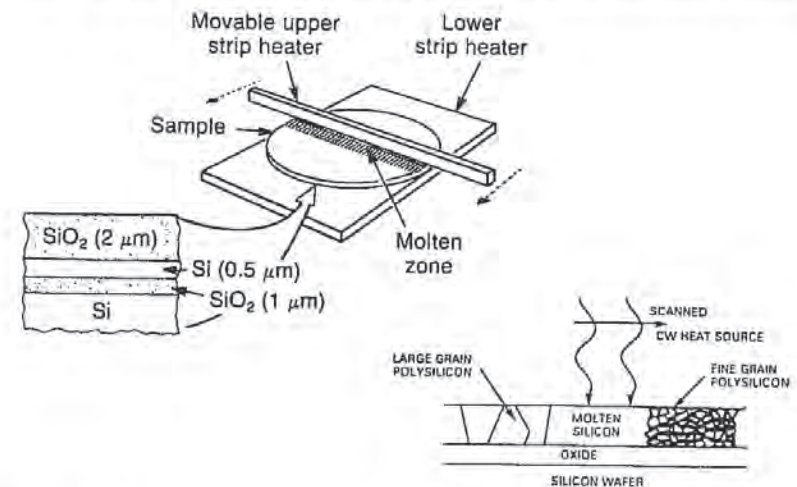


Fig. 2-55 Schematic diagram of sample and graphite strip heaters used in zone melting recrystallization of SOI films.⁹⁴ (© 1987 IEEE).

zone in the silicon film directly below it. The zone is scanned across the sample by moving the upper strip-heater, typically at speeds of ~ 1 mm/sec. The recrystallization process is performed in a flowing Ar gas ambient. Several types of CMOS circuits have been fabricated in ZMR SOI films, including 1-kbit static RAMs and 1.2-k gate arrays.

ZMR can implement greater buried-oxide layer thicknesses than SIMOX (in which the range of buried oxide is only $0.3\text{--}0.5\ \mu\text{m}$). With a thicker oxide, of course, there is lower parasitic capacitance and therefore, higher circuit speed. Thicker buried oxides are also needed for power ICs. ZMR's chief disadvantage is that defects remain in the recrystallized film. The predominant ones are low-angle grain boundaries (see Vol. 1, chap. 2), spaced $20\text{--}50\ \mu\text{m}$ apart. Between these grain boundaries, however, the material is free of crystallographic defects. The grain boundaries in ZMR material have very little effect on majority-carrier transport, and thus there is little mobility degradation. However, they do provide paths for enhanced dopant diffusion. Such diffusion in short-channel MOSFETs can result in excess transistor leakage and even in shorting between source and drain. The threshold-voltage variation is greater in ZMR SOI devices than in bulk MOS devices. Furthermore, the minority-carrier lifetimes in ZMR films are reduced from those in conventional bulk wafers ($2.5\ \mu\text{s}$ vs. $240\ \mu\text{s}$).¹¹⁸

The ZMR process requires high operating temperatures, which make it sensitive to small perturbations. For example, particulate contamination of the starting wafers can result in macroscopic defects in the recrystallized films, and temperature nonuniformity can lead to wafer distortion. Nevertheless, the overall quality of ZMR wafers is good enough for some VLSI applications and efforts are being made to establish a commercial supply of this material.

2.10.6 Full Isolation by Porous Oxidized Silicon (FIPOS)

The SOI technology known as FIPOS is more complicated than other SOI technologies, but offers the potential for essentially defect-free active silicon.⁹⁷ This full-isolation technology is implemented by depositing an epitaxial film on a silicon substrate that has a heavily doped surface layer. Openings are cut through the epi layer, and the entire heavily doped silicon region under the epi layer between the etched recesses is converted to porous silicon (Fig. 2-56). The original process employs anodization of a heavily p -doped layer with an HF solution to preferentially render the silicon porous. More recently an n^+ layer between an n^- island and an n^- substrate has been selectively converted to porous silicon.⁹⁸ This allows the thicknesses of the porous silicon layer and the island to be easily controlled and to be made quite uniform. In addition, it provides an automatic end point for the island formation: as soon as all of the easily-anodized n^+ layer is converted to porous silicon, the anodization current drops, and the conversion reaction stops. Finally, the formation of porous silicon occurs only laterally, and hence large silicon islands can be isolated with this technique.

Porous silicon oxidizes much more readily than single-crystal silicon, and this characteristic is exploited to form silicon islands completely surrounded by SiO_2 . The

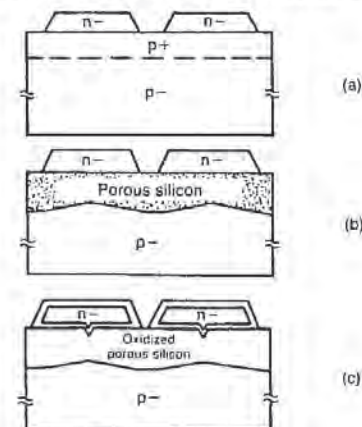


Fig. 2-56 Schematic of the buried p^+ anodization process for forming SIPOS isolation structures.⁹⁷ (© 1987 IEEE).

thermal-oxidation process yields material that has electrical and physical properties similar to thermal oxide. High-pressure oxidation reputedly eliminates the formation of defects in the silicon islands during the oxide growth and significantly reduces wafer warpage (which is one of the drawbacks of the FIPOS process). Oxidized porous silicon etches about an order of magnitude more rapidly than thermal oxide. The etch rate can be slowed by densifying the porous material in steam at about 1100°C . During the oxidation step, a layer of SiO_2 also grows on the surface and sidewalls of the single-crystal islands.

The FIPOS advantages are that the buried-oxide layer can be made much thicker than with SIMOX, and that it can provide arbitrarily thick island regions, without the need to resort to an extra epi step, as in SIMOX or ZMR. As a result, the islands on FIPOS substrates can have far fewer defects than those of SIMOX or ZMR. High-speed, 64-kbit SRAMs have been among the devices reported to have been built on FIPOS substrates. One of the main problems with FIPOS, however, is that a variety of nonconventional silicon-processing steps must be developed in order to implement it. These include the anodization process to form the porous silicon layer and an oxidation and anneal sequence for this layer.

2.10.7 Novel SOI CMOS Processes with Selective Oxidation and Selective Epitaxial Growth

Two novel SOI processes were recently reported that do not need a special oxygen implanter (as does SIMOX) nor a special anodization process (as does FIPOS), and yet

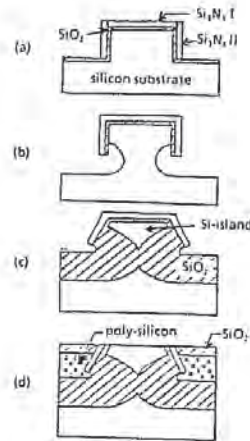


Fig. 2-57 Essential process steps of SOI method described in section 2.10.5.¹⁰⁰ (© 1986 IEEE).

produce active regions without high-angle grain boundaries (as occur in ZMR active regions).

In the first process, a selective oxidation beneath a top silicon layer is used to form the SOI structure.¹⁰⁰ A double silicon-nitride layer (similar to the structure used in SWAMI) is first created (Fig. 2-57a). The bottom of the trenches that were etched in the course of producing the structure just described are isotropically etched to create the structure shown in Fig. 2-57b. This structure is then selectively oxidized until the silicon islands are electrically isolated from the substrate. The trenches are finally filled to form a planar-surface topography with polysilicon and CVD SiO_2 . The process was used to form islands with widths of 1.1-1.7 μm and minimum separations of 1.5 μm . One obvious difficulty with this method is that the width of the islands created is quite small. The possible generation of defects that might result from the growth of the oxide in a confined area was also not discussed in the report.

In the second process, wide trenches are etched 15 μm deep into a Si substrate (15-30 μm wide), and a 2- μm -thick CVD oxide is deposited into the trenches. Next, a seed hole of $\sim 3\text{-}\mu\text{m}$ width is etched into the CVD oxide at the bottom of the trenches. A selective epitaxial silicon region is then grown through the seed hole to fill the trenches. After the epi is deposited, the wafers are oxidized to grow a 2- μm -thick oxide film to fill the gap between the epi islands and the substrate and to remove the seed hole to the silicon substrate.¹¹⁹

REFERENCES

1. B. T. Murphy, V. J. Glinski, P. A. Gary, and R. A. Pederson, "Collector Diffusion Isolated Integrated Circuits," *Proceedings of the IEEE*, 57, 1523 (September, 1969).
2. J. A. Bruchez and P. Pollok "The Philosophy of a Simple Collector Diffusion Isolation Bipolar Process" *Solid State Technology*, p. 93, Aug. 1987.
3. E. Kooi and J. A. Appels, in "Semiconductor Silicon 1973", H. R. Huff and R. Burgess, Eds., p. 860, *The Electrochemical Symp. Series*, Princeton, N.J. (1973).
4. I. Magdo and A. Bogh, *J. Electrochem. Soc.*, 125, 932 (1987).
5. R. Chapman et al., "0.8 μm CMOS Technology" *IEDM Tech. Dig.* 1987, p. 362.
6. Y. Han and B. Ma "Poly Buffer Layer for Scaled MOS", *VLSI Sci. & Tech., Electrochem. Soc.*, 1984, p. 334.
7. A. Bogh and A. K. Gaiind, *Appl. Phys. Lett.* 33, 895 (1978).
8. L. C. Parrillo, *Semiconductor Intl.*, April, 1988, p. 64.
9. J. Pfister and J. Alvis, "Improved CMOS Field Isolation Using Ge/B Ion Implantation", *IEEE Elect. Dev. Lett.*, Aug. 1988, 391.
10. S. M. Hu, "Defects and Device Processing: Achievements and Limitations", *VLSI Sci. & Tech., Electrochem. Soc.*, 1987, p. 722.
11. J. Manoliu, *Semiconductor Intl.*, April, 1988, p. 90.
12. T. Mizuno et al., "Oxidation Rate Reduction in the Submicron LOCOS Process" *IEEE Trans. Elect. Dev.*, Nov. 1987, p. 2255.
13. W. R. Hunter et al., *IEEE Trans. Electron Dev.*, ED-26, 353, (1979).
14. J. Hui, P. Vande Voorde, and J. Moll, "Scaling Limitations of Submicron LOCOS Technology", *Tech. Dig. IEDM*, 1985, p. 392.
15. T. Mizuno et al., *IEEE Trans. Electron Dev.*, Nov. 1987, 2255.
16. E. Kooi et al., *J. Electrochem. Soc.* 123, 1117 (1976).
17. T. A. Shankoff et al., *J. Electrochem. Soc.*, 127, 216 (1980)
18. C. A. Goodwin and J. W. Brossman, *J. Electrochem. Soc.*, 129, 1066 (1982).
19. V.A. Dhaka et al., "Subnanosecond ECL using Isoplanar II", *IEEE J. Solid State Ckts.*, SC-8, Oct. 1973, p. 368.
20. E. Bassous et al., *J. Electrochem. Soc.* 123, 1729, (1976).
21. H. Murrmann, *Siemens Forsch. - Entwicklungsber.* 5, 353, (1976).
22. R. C. Ellwanger et al., *Ext. Abs. Fall Meeting ECS*, 1984, No. 418, p. 603.
23. A. Cosand and S. Prussin, *ECS Meeting*, May, 1976. Ext Abs. No. 109, p. 292.
24. B. O. Kolbesen and H.P. Strunk, "Analysis, Electrical Effects, and Prevention of Defects", *Solid-State Devices, Inst. Phys. Conf. Series*, 57, 1980, p. 21.
25. K. Sakuma et al., "A New Self-Aligned Planar Oxidation Technology", *J. Electrochem. Soc.*, June, 1987, p. 1503.
26. N. Guillemot et al., "A New Analytical Model of the Birds Beak", *IEEE Trans. Elect. Dev.* ED-34, May 1987, p. 1033.
27. P. Sutardja et al., *IEDM Tech. Dig.*, 1986, p. 526.
28. A. Lewis et al., "Device Isolation in High-Density LOCOS CMOS", *IEEE Trans. Elect. Dev.*, June 1987, 1337.
29. K. Y. Chiu et al., *IEEE Trans. Electron Dev.*, ED-29, p. 536, (1982)

30. C. Claeys and J. Vanhellemont "Structural and Electrical Characterization of SWAMI for Submicron Technologies" *Ext. Abs. of Spring Meeting of the ECS*, 1988, Abs. No. 140, p. 211.
31. M. Ghezzi et al., "Laterally Sealed LOCOS Isolation", *J. Electrochem. Soc.*, June 1987, p. 1475.
32. P. Deroux-Dauphin & J.P. Gonchond, *IEEE Trans. Electron Dev.*, Nov. 1985, p. 2392.
33. M. P. Brassington, R. Razouk, and C. Hu, *IEEE Trans. Electron Dev.*, Jan 1988, p. 96.
34. N. Hoshi et al., *Tech. Dig. IEDM*, 1986, p. 300.
35. T. Nishihara et al., *Tech. Dig. IEDM*, 1988, p. 100.
36. L. C. Parrillo et al., *Tech Dig. IEDM*, 1986, p. 244.
37. M. Brassington et al., *IEEE Trans. Electron Dev.*, Jan 1988, p. 96.
38. F. Bryant et al., *Ext. Abs. Spring Meeting Electrochem. Soc.* 1988, Abs. No. 137, p. 205.
39. H. H. Tsai et al., *IEEE Trans. Electron Dev.*, March 1988, p. 275.
40. T. Kaga et al., "Advanced OSELO Isolation with Shallow Grooves" *IEEE Trans. Electron Dev.*, July 1988, p. 893.
41. M. Mikoshiba et al., *IEDM Tech. Dig.*, 1984, p. 578.
42. G. Fuse et al., *IEEE Trans. Electr. Dev.*, Feb. 87, p. 356.
43. T. Iizuka et al., *Tech. Dig. IEEE IEDM*, 1981, p. 384.
44. G. Fuse et al., *IEDM Tech. Dig.*, 1987, p. 732.
45. B. Davari et al., *Tech. Dig. IEDM*, 1988, p. 92.
46. T. Shibata et al., *IEDM Tech. Dig.*, 1983, p. 27.
47. R. Kwasnick et al., "Buried Oxide Isolation with Etch-Stop", *IEEE Elect. Dev. Lett.*, Feb. 1988, p. 62.
48. Y. Tamaki et al., *Jap. J. Appl. Phys.*, 21, Supplement 21-1, p. 37 (1982).
49. Y. Tamaki et al., *J. Electrochem. Soc.*, March 1988, p. 726.
50. Y. Niitsu et al., "Latchup-Free CMOS Structure Using Shallow Trench Isolation", *Tech. Dig. IEDM*, 1985, p. 509.
51. J. R. Pfister and J. R. Alvis, "A Novel CMOS VLSI Isolation Technology Using Selective Cl Implantation", *IEEE Electron Dev. Letts.*, Nov. 1988, p. 561.
52. M. J. Kim et al., *Ext. Abs. Fall Mtg Electrochem. Soc.*, 1988, Abs. No. 241, p. 338.
53. G. K. Herb, D. J. Rieger, and K. Shields, "Silicon Trench Etch in a Hex Reactor", *Solid-State Technol.*, Oct., 1987, p. 109.
54. "TI's Trench Technology Moves into the Factory", *Electronics*, July 9, '87, p. 75.
55. M. Sato and Y. Arita, *J. Electrochem. Soc.*, Nov. 1987, p. 2856.
56. R. N. Carlile et al., "Trench Etches in Silicon with Controllable Sidewall Angles", *J. Electrochem. Soc.*, Aug., 1988, p. 2058.
57. S. Samukawa et al., *Ext. Abs. Fall Meeting Electrochem. Soc.*, 1988, Abs. No. 110, p. 163.
58. R. E. Nowak, *Solid State Tech.*, March 1988, p. 39.
59. K. Yamada et al., *Tech. Dig. IEDM*, 1985, p. 702.
60. K. Yamabe and K. Imai, *IEEE Trans. Electron Dev.*, Aug. 87, p. 1681.
61. K.V. Rao et al., *Tech. Dig. IEDM*, 1986, p. 140.

62. R. D. Rung, "Trench Isolation Prospects for Application in CMOS VLSI", *Tech. Dig. IEDM*, 1984, p. 574.
63. K. Tsukamoto et al., *IEDM Tech. Dig.*, 1987, p. 328.
64. F. Horiguchi et al., *IEDM Tech. Dig.*, 1987, p. 324.
65. D.-B. Kao et al., "Two-Dimensional Silicon Oxidation Experiments and Theory", *Tech. Dig. IEDM*, 1985, p. 388.
66. V. J. Silvestri, "Growth Kinematics of a Polysilicon Trench Refill Process", *J. Electrochem. Soc.*, Nov. 1986, p. 2374.
67. V. J. Silvestri, "Selective Epitaxy Trench", *J. Electrochem. Soc.*, July, 1988, p. 1808.
68. Y.-C. S. Yu, "Planarized Deep-Trench Process for Bipolar Isolation", *Ext. Abs. Fall Meeting 1988, Electrochem. Soc.* Abs. No. 234, p. 326.
69. J. Borland and C. Drowley, *Solid State Tech.*, Aug. 1985, p. 141.
70. J. Borland et al., "Silicon Epitaxial Growth for Advanced Structures", *Solid State Tech.*, Jan 1988, p. 111.
71. J. Manoliu and J. O. Borland, "A Submicron Dual Buried Layer Twin Well CMOS SEG Process", *IEDM Tech. Dig.*, 1987, p. 20.
72. S. Nagao et al., *IEEE Trans. Electron Dev.*, Nov. 1986, p. 1738.
73. N. Endo et al., *IEEE Trans. Electron Dev.*, Nov. 1986, p. 1659.
74. A. Stivers and C.H. Ting, "Selective Epitaxial Growth for CMOS Isolation", *Presented at the Third Annual Applied Materials Inc., "Innovations in Epitaxial Technology for Advanced Device Structures"* Dec. 1988, Palo Alto, CA.
75. J. Borland et al., *Ext. Abs. of Intl. Conf. on Solid State Dev. and Mfg.*, 1986, Tokyo, Japan, p. 53.
76. F. Mieno et al., *Tech. Dig. IEDM*, 1987, p. 16.
77. T. I. Kamins and S. Shiang, *Electron Dev. Letts.*, Dec 1985, p. 617.
78. N. Endo et al., *IEEE Trans. Electron Dev.*, Nov. 1986, p. 1659.
79. Y.-C. Yu and A. Witkowski, *J. Electrochem. Soc.*, October 1988, p. 2562.
80. J. O. Borland, D. Schmidt and A. Stivers, *Ext. Abs. of 18th Conf. on Solid-State Devices and Materials*, 1986, Tokyo, Japan, p. 1659.
81. J.B. Lasky et al., "SOI by Bonding and Etchback" *Tech. Dig. IEDM*, 1985, p. 684.
82. R. C. Frye et al., *J. Electrochem. Soc.*, Aug 1986, p. 1673.
83. H. Ohashi et al., *Tech. Dig. IEDM*, 1986, p. 210.
84. W. A. Krull et al., *IEEE Circuits and Systems Mfg.*, July 1987, p. 20.
85. K. Ueno et al., *Tech. Dig. IEDM*, 1988, p. 870.
86. P. K. Vasudev "Recent Advances in Solid-Phase Epitaxial Recrystallization of SOS for CMOS and Bipolar Devices", *IEEE Circuits and Devices Magazine*, July 1987, p. 17.
87. H. W. Lum, "SIMOX SOI for IC Fabrication", *IEEE Circuits and Sys. Mag.*, July 1987, p. 6.
88. D. J. Foster et al., *IEEE Trans. Electron Dev.*, March 1986, p. 354.
89. B.-Y. Mao et al., *IEEE Electron Dev. Letts.*, July, 1987, p. 306.
90. J. R. Davis et al., *IEEE Electron Dev. Letts.*, July, 1987, p. 291.
91. L. Jastrzebski et al., *J. Electrochem. Soc.*, July, 1988, p. 1746.
92. L. Nesbit et al., *J. Electrochem. Soc.*, Nov. 1985, p. 2713.

93. K. M. Cham, S. Oh, D. Chin, and J. L. Moll, *Computer-Aided Design and VLSI Device Development*, Boston, Kluwer, 1986, Chap. 9.
94. B.-Y. Tsaur, "ZMR SOI Technology", *IEEE Circuits and Sys. Mag.*, July, 1987, p. 12.
95. J. C. C. Fan et al., *J. Cryst. Growth*, 63, No. 3, 1983.
96. T. Stultz et al., in *Laser-Solid Interactions and Transient Thermal Processing of Materials*, Ed. by J. Narayan, North-Holland, New York, p. 463, 1983.
97. S. S. Tsao, "Porous Silicon Techniques for SOI Structures", *IEEE Circuits & Sys. Mag.*, November 1987, p. 3.
98. E. J. Zorinsky et al., *Tech. Dig. IEDM*, 1986, p. 431.
99. C. T. Sah, "The Evolution of the MOS Transistor", *Proceedings of the IEEE*, October 1988, p. 1280.
100. M. Kubota et al., *Tech. Dig. IEDM*, 1986, p. 814.
101. W. Wakayama et al., *Tech. Dig. IEDM*, 1988, p. 246.
102. M. Shimizu et al., *Tech. Dig. IEDM*, 1988, p. 96.
103. M. H. El-Diway et al., *Tech. Dig. IEDM*, 1987, p. 917.
104. H. J. Levinson and W. H. Arnold, *J. Vac. Sci. Tech.* B5(1), p. 293 (1987).
105. D. W. Hess, *Solid State Technol.*, Apr. 1981, p. 192.
106. M. Ameen et al., *Semiconductor Internat.*, September, 1988, p. 122.
107. R. N. Carlile et al., "Trench Etches in Silicon with Controllable Sidewall Angles", *J. Electrochem. Soc.*, Aug. 1988, p. 2058.
108. J. Y. Chen, *CMOS Science and Technology*, Prentice-Hall, Englewood Cliffs, N. J., 1990, p. 237.
109. K. L. Wang et al., *IEEE Trans. Electron Dev.* ED-29, p. 541, 1982.
110. J. Hui et al., *IEEE Trans. Electron Dev.*, ED-29, p. 554, 1982.
111. R. P. Kramer, *Ext. Abs. Spring Meeting Electrochem. Soc.*, 1989, Abs. No. 131, p. 182.
112. P. A. van der Plas et al., *Dig. of 1987 Symp. on VLSI Technology*, III-4, p. 19.
113. J. C. Sturm, *Materials Research Soc. Symp. Proc.*, 107, p. 295, (1988).
114. T. W. MacElwee, *Ext. Abs. Spring Meeting Electrochem. Soc.*, 1989, Abs. No. 192, p. 280.
115. J. -P. Colinge, *Ext. Abs. Spring Meeting Electrochem. Soc.*, 1989, Abs. No. 193, p. 282.
116. G. R. Powell, S. V. Davies, and A. H. Chambers, *Ext. Abs. Electrochem. Soc. Conf.*, Spring, 1989, p. 252.
117. M. Englehardt and S. Schwarzl, *Ext. Abs. Electrochem. Soc. Conf.*, Spring, 1989, p. 254.
118. G. A. Rozgonyi et al., *Ext. Abs. Electrochem. Soc. Conf.*, Spring, 1989, p. 339.
119. D. M. Boisvert and J. D. Plummer, *Ext. Abs. Electrochem. Soc. Conf.*, Spring, 1989, p. 431.
120. S. Mahli et al., *Ext. Abs. Electrochem. Soc. Conf.*, Spring, 1989, p. 427.
121. B. Coulman et al., *Ext. Abs. Electrochem. Soc. Conf.*, Spring, 1989, p. 293.
122. S. L. Partridge, *Tech. Dig. IEDM*, 1986, p. 428.
123. J. Hui et al., *IEEE Trans. Electron Dev.*, vol. ED-29, p. 554, 1982.
124. C. Clacys et al., *J. Electrochemical Society*, September 1989, p. 2619.
125. M. Ghezzi et al., *J. Electrochemical Soc.*, July 1989, p. 1992.

126. I. Manchanda et al., "A High Performance, Self-Aligned, Ultra-Rad-Hard Field Oxide Technology for NMOS/CMOS VLSI", *IEEE Trans. Electron Dev.*, April 1989, p. 651.
127. H. Kabza et al., *IEEE Electron. Dev. Lett.*, August 1989, p. 344.
128. R. C. Jaeger, *Introduction to Microelectronic Fabrication*, Addison-Wesley, Reading MA, 1988.
129. B. T. Murphy et al., *Proc. of the IEEE*, 57, 1523 (1969).
130. S. M. Sze, Ed. *VLSI Technology*, 2nd Ed., New York, McGraw-Hill, 1988, p. 476.
131. D. R. Craven and J. B. Stummell, *Semiconductor International*, June 1981, p. 59.
132. N. Kusai et al., *Tech. Dig. IEDM*, 1985, p. 419.
133. K. E. Bean and W. R. Runyan, *J. Electrochem. Soc.*, 124, 5C (1977).

PROBLEMS

- 2.1 Assume that two n^+ diffused lines are run parallel in a substrate that is doped with 10^{15} boron atoms/cm³, and the substrate is biased to -5 V, and the two lines are each connected to +5 V. Using one-dimensional junction theory, calculate the minimum spacing required between the two lines to keep their depletion regions from merging.
- 2.2 Use one-dimensional theory to estimate the punchthrough voltage between two adjacent MOSFETs whose respective source and drain regions are separated by 1.5 μ m. Assume a substrate doping of 3×10^{16} atoms/cm³ and a substrate bias of 0 V.
- 2.3 What is the maximum gate-to-source voltage that a MOSFET with a 10-nm gate oxide can withstand. Assume the gate oxide breaks down at 5 MV/cm and that a 0 V substrate bias.
- 2.4 Explain why oxide isolation has not found widespread use in analog IC applications.
- 2.5 It was noted that the CDI process is used mainly for digital applications. What characteristics of the structure are responsible for this?
- 2.6 Explain why bird's beak and boron encroachment limit the minimum dimension in which conventional LOCOS isolation can be used.
- 2.7 Discuss each of the following drawbacks of LOCOS isolation technology:
 - (a) bird's beak encroachment
 - (b) boron encroachment
 - (c) stress induced defects
 - (d) oxide thinning in narrow field oxide regions
 - (e) non-planarity
 - (f) process complexity
- 2.8 Compare the various SOI technologies, and list the advantages and drawbacks of each of them.

6.5 If an LDD structure is needed for both n - and p -channel MOSFETs in a submicron CMOS process, list the CMOS masking sequence and associated major processing steps. If the doping concentration of the p -LDD is an order of magnitude larger than that of the n -LDD, can any of these steps be omitted?

6.6 Comparing a retrograde well CMOS process to a conventional well CMOS process, what is the most important difference in the two process sequences and in the completed device structures? How do these differences impact the device isolation?

6.7 In a $(p^+$ diffusion)-(n -well)-(p -on- p^+ substrate) structure, the n -well is uniformly doped at $1 \times 10^{16} \text{ cm}^{-3}$ and is $1 \mu\text{m}$ deep, the p -epi thickness is $4 \mu\text{m}$ and is uniformly doped at $1 \times 10^{15} \text{ cm}^{-3}$, and the p^+ diffusion is $0.4 \mu\text{m}$ deep. If a substrate bias generator is needed and the p^+ diffusion and n -well are both biased at 5 V , what is the maximum (in absolute value) substrate voltage that can be applied before punchthrough occurs?

6.8 Explain why using a lightly doped epitaxial layer on a heavily doped substrate is one of the most effective ways to increase latchup resistance. Why is a thinner epi even more helpful and what limits the continuous scaling of epi thickness?

6.9 For an n -well CMOS structure, draw a lumped equivalent circuit model of the parasitic device structures that give rise to latchup and derive the conditions for latchup to occur.

6.10 Explain why latchup cannot be sustained if the holding voltage is greater than the supply voltage. What is the most effective technique for increasing the holding voltage?

6.11 If the n^+ -to- n -well spacing is $3 \mu\text{m}$, the n -well depth is $1.3 \mu\text{m}$, and the p^+ diffusion depth is $0.3 \mu\text{m}$, estimate the minimum pulse width needed to trigger latchup.

6.12 In an n -well CMOS structure the well depth is $1 \mu\text{m}$ and the well concentration $1 \times 10^{16} \text{ cm}^{-3}$. The p^+ diffusion is $0.3 \mu\text{m}$ deep and it is doped at $1 \times 10^{19} \text{ cm}^{-3}$. Estimate the vertical pnp transistor common-emitter current gain (β) using one-dimensional analysis. (Hint: You will need to calculate the minority-carrier diffusion length in the base of the transistor, and compare this with the base width in order to estimate α , which can then be used to find β .)

CHAPTER 7

BIPOLAR AND BICMOS PROCESS INTEGRATION

This chapter deals with bipolar and BiCMOS process integration. A review of the physics of bipolar transistors is provided first.¹⁻³ Next is a discussion of the relationship between device performance and process technology, followed by a discussion of the evolution of bipolar IC process technology. (A book dealing exclusively with the device physics of bipolar transistors has recently been published.)¹⁵¹

In the second half of the chapter we consider BiCMOS technology, emphasizing process integration. The details of two process sequences are discussed – one for the fabrication of high-performance, digital-BiCMOS ICs, and the other for the fabrication of analog/digital BiCMOS ICs. (Reference 4 provides information on an excellent, recently published book dealing with the device physics, process technology, and circuit design of BiCMOS technology.)

7.1 BIPOLAR-TRANSISTOR STRUCTURES FOR INTEGRATED CIRCUITS

The bipolar transistor is an electronic device with two pn junctions in very close proximity. Idealized one-dimensional representations of bipolar transistors are shown in Figs. 7-1a and 7-1b. It can be seen that there are three device regions: an *emitter*, a *base* (the middle region), and a *collector*. In the ideal structure, the two pn junctions (called the *emitter-base* [E-B] and *collector-base* [C-B] junctions) are in a single bar of semiconductor material, separated by a distance designated as W_B . Modulation of the current flow in one pn junction by means of a change in the bias of the nearby junction is called *bipolar-transistor action*.

External leads can be attached to each of the three regions, as depicted in Fig. 7-1, and external voltages and currents can be applied to the device from these leads. Figure 7-1 also shows the circuit symbols of the devices with each lead labeled. If the emitter and collector are doped n -type and the base is doped p -type, the device is called an *npn* transistor (Fig. 7-1a). If the opposite doping configuration is used, the device is referred to as a *pnp* transistor (Fig. 7-1b).

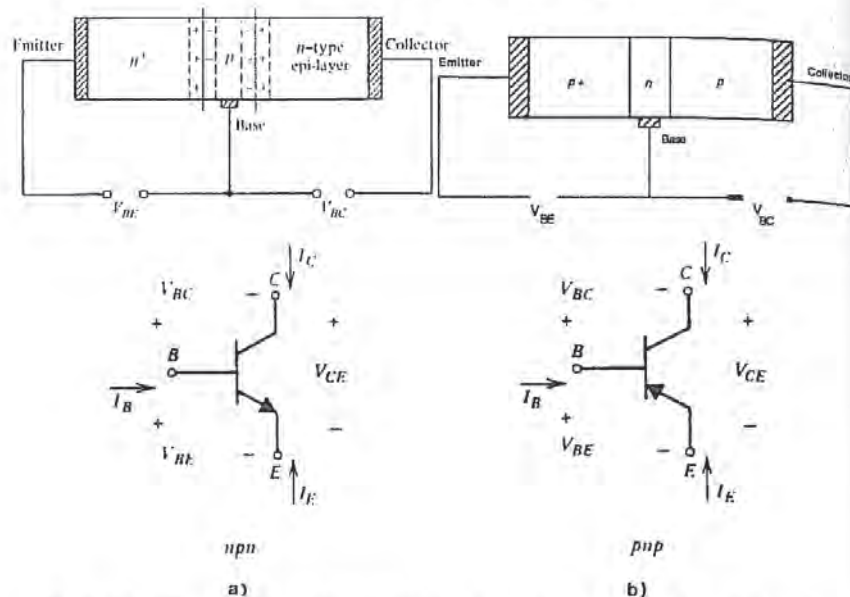


Fig. 7-1 Idealized one-dimensional model and circuit symbol of: (a) an *npn* bipolar transistor, and (b) a *pnp* bipolar transistor.

7.1.1 The Transistor Action

Transistors are used as either amplifying or switching devices. In the first application, the transistor's function is to faithfully amplify small ac signals; in the second, a small current is used to switch the transistor from an *ON* to an *OFF* state and back. These capabilities exist because of *transistor action*, which will be described in detail in this section. (Note that since the mobility of minority carriers – i.e., electrons – in the base region of *npn* transistors, is higher than that of holes in the base of *pnp* transistors, higher-frequency operation and higher-speed performances can be obtained with *npn* devices. As a result, the latter make up the vast majority of bipolar transistors used to build ICs.)

7.1.1.1 Basic Bipolar-Transistor Physics. The energy bands pertaining to the idealized *pnp* transistor at equilibrium are shown in Fig. 7-2a. Under equilibrium, no net current flows across either the E-B or C-B junction, and hence no currents flow in the external leads (e.g., the emitter, base, and collector currents – I_E , I_B , and I_C , respectively – are all equal to zero).

Since external voltages may be applied to the transistor, either junction may be forward- or reverse-biased. If the E-B junction is forward-biased and the C-B junction is

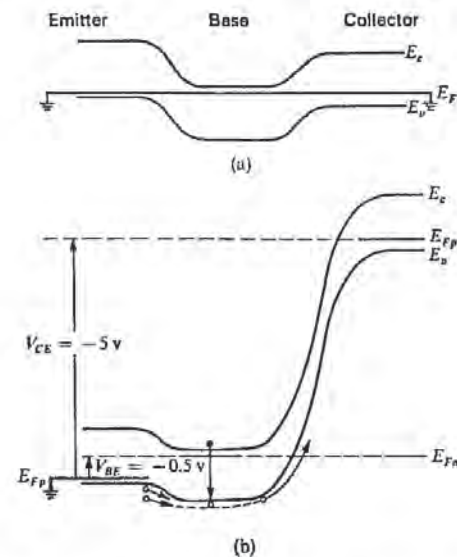


Fig. 7-2 Energy band diagram for a *pnp* transistor: (a) In equilibrium; (b) With the E-B junction forward biased and the C-B junction reverse biased (forward-active mode).

reverse-biased, the energy-band diagram of the device is as shown as in Fig. 7-2b. This mode of operation is called the *forward-active mode*. Forward-biasing of the E-B junction lowers its potential-energy barrier by qV_{BE} , whereas reverse-biasing of the C-B junction increases the potential energy barrier of this junction by qV_{CB} (where q is the charge of one electron – $1.6 \times 10^{-19} C$ – and V_{BE} and V_{CB} are the bias voltages applied to the E-B and C-B junctions, respectively).

In an *npn* transistor operating in the forward-active mode, a large number of electrons are injected into the *p*-type base as a result of the forward bias applied to the E-B junction. If the distance between the two junctions is small enough, most of these electrons will reach the reverse-biased C-B junction. The electric field there will sweep the electrons across the junction, where they will be *collected* in the *n*-type collector, giving rise to an I_C current that is almost as large as the forward-bias current of the E-B junction, I_E . Hence, a large current can flow in a reverse-biased junction due to the existence of a nearby forward-biased junction, but only if the two junctions are close enough physically to interact in the manner described. *This effect represents bipolar-transistor action.**

* The conceptualization of transistor action and the invention of a device in which such action could be implemented (the *bipolar transistor*) together represent one of the most

7.1.1.2 Bipolar-Transistor Current Gain. Not all of the electrons injected into the base will arrive at the collector. Some will recombine in the base region as they move through it. In addition, the forward bias applied to the E-B junction will cause some holes to be injected into the emitter region from the base. These two phenomena give rise to a current that flows in the base lead of the transistor.

The total emitter current (I_E) consists of those electrons that reach the collector (I_C) plus those that flow out of the transistor through the base lead (I_B):

$$I_E = I_C + I_B. \quad (7-1)$$

Two important parameters that characterize the transistor action described above are the *common-base current gain*, α , defined as

$$\alpha \equiv I_C/I_E \quad (7-2)$$

and the *common-emitter current gain*, β , defined as

$$\beta \equiv I_C/I_B. \quad (7-3)$$

From Eqs. 7-1, 7-2, and 7-3 it can be shown that α and β are related to each other by

$$\beta = \frac{\alpha}{1 - \alpha} \quad (7-4)$$

The value of α must be less than unity, but in well-designed transistors it closely approaches this value (e.g., $\alpha > 0.9$). Accordingly, β will be much greater than unity in such transistors (typically, between 30 and 200).

Figure 7-3b shows the current-voltage output characteristics of an *npn* transistor for the common-emitter circuit configuration depicted in Fig. 7-3a. These characteristics imply that the transistor can be operated in modes other than the forward-active mode. When it is operated in that mode, however, a small change in I_B will bring about a β -times-larger change in I_C . Thus, when the transistor is used as an amplifying device (with I_B acting as the input signal to be amplified and I_C as the output signal), its operation is normally restricted to the forward-active mode.

As just noted, however, the E-B and C-B junctions may not always be biased so that the transistor is operating in this mode. Two other operating modes are also normally possible. In the first (the so-called *cutoff mode* of operation), the voltage across each junction is reverse-biased. In cutoff, the amount of leakage current flowing between the collector and emitter is so small that in many applications it can be neglected. The bottom portion of the I-V curves shown in Fig. 7-3b represents the cutoff region.

significant feats in the history of device electronics. As a result of the investigations that led to these accomplishments the Nobel Prize in physics was awarded to the bipolar transistor inventors, William Shockley, John Bardeen, and Walter Brattain.

In the second operating mode, both junctions are forward-biased (the so-called *saturation mode*). As a result, the emitter-collector voltage (V_{CE}) approaches zero (note that if $V_{CE} > 0$, the E-B junction of *nnp* devices is more forward-biased than the C-B junction). The left-hand portion of the characteristic curves of Fig. 7-3b represents this region of operation.

In saturation, however, the forward-biased C-B junction injects charge carriers into the base (as does the emitter), and the collector current thus becomes the net of two opposing flows of charge carriers. Therefore, once the device has entered the saturation mode, I_C decreases as V_{CE} decreases. Furthermore, in saturation the base current no longer controls the collector current, and current varies with power-supply voltage according to Ohm's law.

The transistor is operated in the cutoff and saturation modes when it is being used as a switching device. By controlling a small current (I_B), it becomes possible to turn a larger one (I_C) *ON* and *OFF*. If the base current is made zero, the transistor is in cutoff, $I_C \approx 0$, and V_{CE} is equal to the power-supply voltage. The electrical path between the emitter and collector leads of the transistor thus appears as an *OFF* switch. When the base current is large enough, I_C is increased to the degree that the transistor goes into

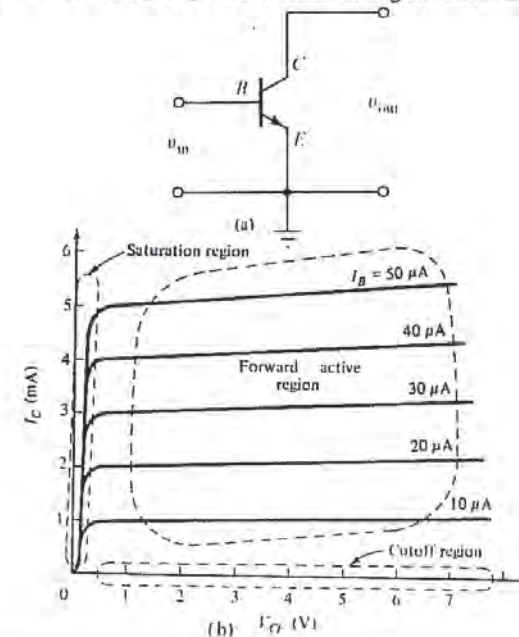


Fig. 7-3 (a) Common-emitter circuit configuration. (b) Current-voltage (I-V) output characteristics of *nnp* transistor in common-emitter (C-E) configuration. Copyright, 1967, John Wiley and Sons, Inc. Reprinted with permission from ref. [11].

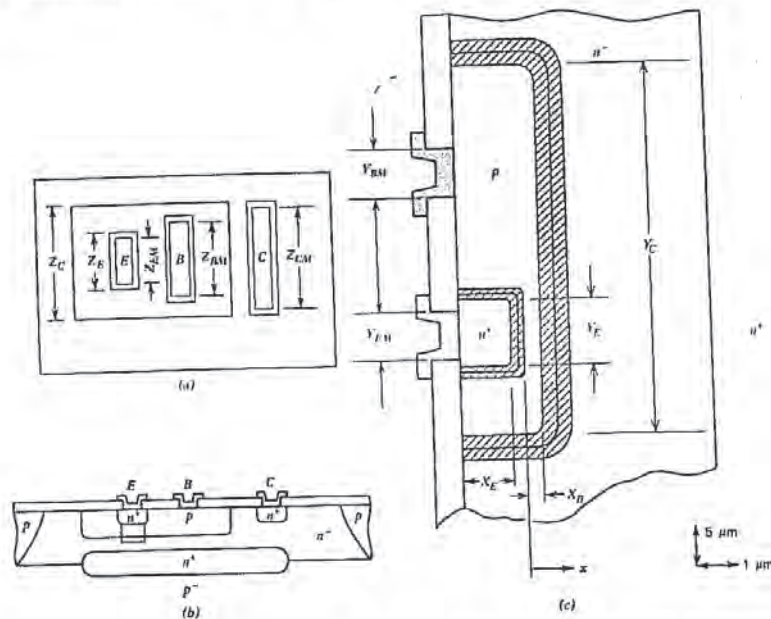


Fig. 7-4 Top view (a) and cross sections [(b) and (c)] of a representative *nnp* transistor. The region dominating the transistor action is shaded in (b) and the area bounded by the base diffusion is rotated 90° and expanded in (c); *x* and *y* scales are indicated for (c) only. Copyright, 1986, John Wiley and Sons, Inc. Reprinted with permission from ref. [1].

saturation. Under this mode of operation, the ratio of the output voltage to the output current is small (i.e., V_{CE} is small and I_C is large). Hence, the current path between the emitter and collector leads of the transistor appears like the current path that exists in an *ON* switch.

7.1.2 Integrated-Circuit Transistor Structures

The physical structure of bipolar transistors used in ICs is quite different from the idealized 1-D structures shown in Fig. 7-1, primarily because the universal method used to fabricate bipolar transistors is the *planar process*. A top view of such a structure with junction isolation is shown in Fig. 7-4a; cross-sectional views are shown in Figs. 7-4b and 7-4c. The specific device dimensions, doping concentrations, and process sequences used to produce such devices will be described later, but it is useful to emphasize here that one of the most critical parameters in transistor fabrication is the width of the base region (W_B in Fig. 7-4c). Transistor action is critically dependent on the proximity of

the two interactive junctions, and in bipolar transistors this dimension is less than 10 μm (in fact, in high-performance bipolar transistors, W_B is less than 0.5 μm).

Because W_B is so much smaller than the other dimensions in the device, transistor action is essentially confined to the shaded region shown in the cross-sectional view in Fig. 7-4b (as shown on a larger scale in Fig. 7-4c, this is the region bounded by the width of the emitter, Y_E). This implies that for many purposes the transistor can still be well approximated by the one-dimensional device structure in Fig. 7-1a. It should also be noted that the collector region in most bipolar *nnp* IC transistors consists of a heavily *n*-doped layer buried beneath a lightly doped epitaxial layer, shown in Fig. 7-4b.

7.2 DIGITAL CIRCUITS USING BIPOLAR TRANSISTORS

A common application of bipolar transistors is in digital logic gates and memory arrays. Several types of circuits have been developed to implement logic gates in bipolar ICs; the most popular are transistor-transistor logic (TTL) and emitter-coupled logic (ECL). Although the detailed circuit behavior of these logic families lies beyond the scope of this text, it is useful to present the information that is relevant to our discussion of processing. For an excellent source of information concerning the analysis and design of digital logic circuits (both bipolar and MOS), see reference 5.

7.2.1 Basic Bipolar-Transistor Inverter Circuits

To define the language of digital bipolar ICs we present a description of a basic bipolar inverter circuit. This inverter consists of a bipolar transistor — connected in a common-emitter configuration — and two external resistors. The transistor is the driver element and is switched *ON* and *OFF* by the input signal; the collector resistor, R_C , is the *load*. A circuit diagram of the inverter is shown in Fig. 7-5a, and a three-dimensional cross-sectional view of the structure is given in Fig. 7-5b.

The load resistor connects the power-supply voltage, V_{CC} , and the output of the inverter. The input signal (V_{in}) is fed to the base of the transistor; the output signal is the voltage level at the output node, V_o . Figure 7-5c depicts the inverter's voltage-transfer characteristic. When there is a logic 0 input signal, V_{in} is less than the turn-on voltage, $V_{BE(on)}$, and the transistor is in the cutoff mode (i.e., the voltage across E-B is less than V_{BE} , and the voltage across C-B is reverse-biased). In this case, I_C is essentially zero; however, since the output is electrically connected to V_{CC} through R_C , V_o rises (or is *pulled up*) to logic 1 (close to V_{CC}).

When the input voltage increases above the value of $V_{BE(on)}$, the driver transistor is turned *ON*. The transistor first enters the forward-active mode, because V_{CB} is reverse-biased. In this mode, the collector current is related to the base current by $I_C \approx \beta I_B$. As the input voltage increases, I_B also increases. This produces a decrease in V_o , given by ($V_o = V_{CC} - \beta I_B R_C$). If the input voltage becomes high enough, V_o will become so small that the transistor will enter the saturation mode (both junctions would have to be forward-biased to allow V_o to drop below V_{BE}). In the saturation, region V_o will

remain at a constant low value (i.e., V_{CEsat} , which is typically between 0.1 and 0.2 V) as the voltage is increased further. This demonstrates how the logic level at the input is inverted at the output of the circuit.

As noted in chapter 5, the desired characteristics of digital logic gates as IC elements are the following: fast switching speed (i.e., short propagation-delay time); low power dissipation; small size (i.e., minimum silicon area); and high noise margin. The basic bipolar inverter circuit just described does not fulfill such requirements as well as do more complex digital bipolar circuits, and hence it is not used in VLSI circuits. Instead, inverters (and other digital logic gates) are implemented with other digital circuits, which will be discussed next. Note, however, that the basic bipolar IC inverter circuit requires the availability of resistors as part of the integrated circuit. Such IC resistors are also required in other bipolar-logic gates. A discussion of the fabrication of IC resistors is given in Appendix A.

7.2.2 Bipolar Digital-Logic-Circuit Families

Many types of bipolar digital logic families have been developed for LSI and VLSI

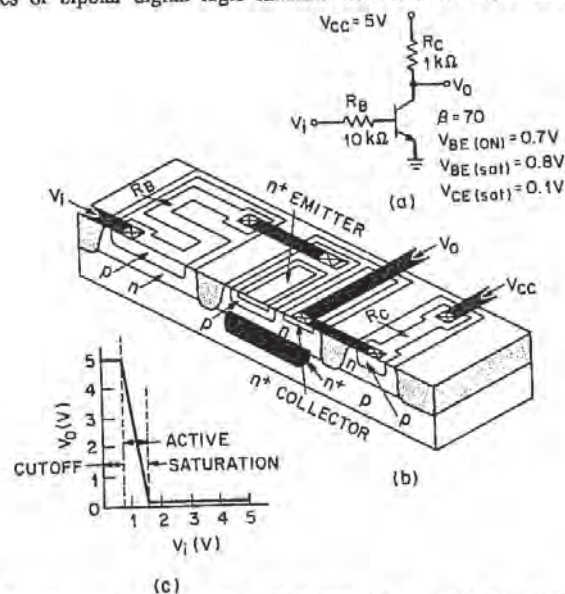


Fig. 7-5 Bipolar transistor inverter. (a) Circuit diagram. (b) Perspective view of the inverter structure. (c) Voltage transfer characteristics. From D. A. Hodges and H. G. Jackson, *Analysis and Design of Digital Integrated Circuits*, Copyright 1983, McGraw-Hill Book Co. Reprinted with permission from ref. [5].

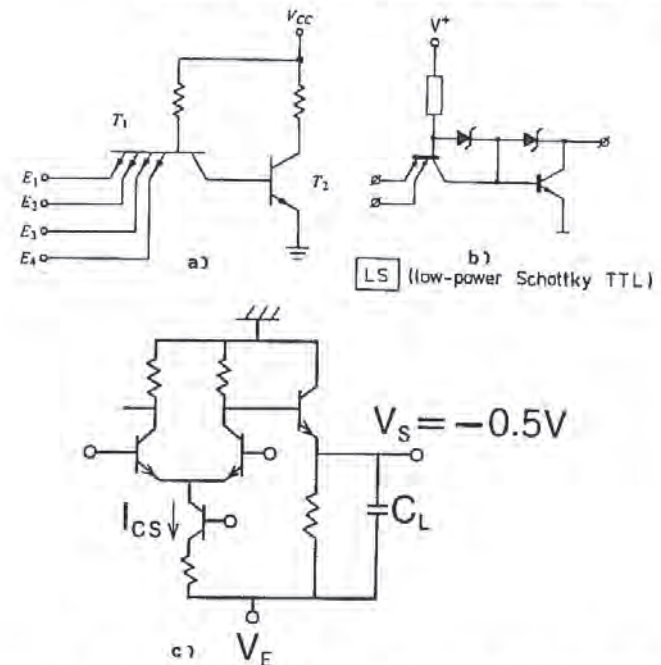


Fig. 7-6 Equivalent circuits of basic bipolar gates in the following logic families: (a) Transistor-transistor logic (TTL), (b) Low-power Schottky TTL, (c) Emitter-coupled logic (ECL).

circuits. A detailed analysis of their circuit characteristics can be found in reference 6 (which also provides a survey of the most important of these families), and in reference 5 (which, as noted earlier, contains a thorough treatment of transistor-transistor-logic [TTL] and emitter-coupled logic [ECL] circuit behavior).

In this section we present the electrical characteristics of TTL and ECL circuits. This background information provides a rationale for some of the device-design and process approaches used to fabricate bipolar and BiCMOS ICs.

Figure 7-6a shows the circuit diagram of a standard TTL gate, while Figs. 7-6b and 7-6c give the circuit diagrams of a low-power Schottky TTL gate and a basic emitter-coupled logic gate (ECL), respectively. The two TTL circuits are operated from a power-supply voltage of 5 V, but the ECL gate uses a -5.2 V supply.

Table 1-1 lists some of the other important properties of these logic families (including the 10K and 100K ECL families, of which the latter is the more advanced). The logic levels of TTL and ECL are quite different (e.g., 2.4 to 0.4 V in standard-TTL, but -0.9 V to -1.7 V in 10K ECL). Thus, if a digital system is built with both

Table 7-1 Typical Electrical Characteristics of Bipolar Logic Families, at $T_A = 25^\circ\text{C}$.

	V_{OL}/V_{OH}	V_{IH}/V_{IL}	Power Dissipation per Gate	Propagation Delay Time
TTL Families (Supply volts = 5.0 V, Logic swing = 2.2 V)				
Series 74 (Standard)	2.4 V/0.4 V	2.0 V/0.8 V	10 mW	10 ns
Series 74S (Schottky)	"	"	20 mW	3 ns
Series 74LS (Low-power Schottky)	"	"	2 mW	10 ns
Advanced Schottky TTL (ALS)				
Series 54/74F (Fast)	"	"	4 mW	2.5 ns
Series 54/74 AS (Adv. Schottky)	"	"	20 mW	1.5 ns
Series 54/74 ALS (Low-power)	"	"	1 mW	4 ns
ECL Families (Supply volts = -5.2 V, Logic swing = 0.8 V)				
10K	-0.9 V/-1.7 V	-1.2 V/-1.4 V	24 mW	2 ns
100K	"	"	40 mW	0.75 ns

TTL and ECL logic gates, the circuits must be interfaced correctly (i.e., it is necessary to voltage-level-shift the output from one of the families to be compatible with the input of the other). This level-shifting function is performed by circuits known as ECL-to-TTL and TTL-to-ECL *translators*.

Some other interesting comparisons between the logic families can be made. First, ECL gates are faster than TTL gates, but they consume more power. (The speed advantage of ECL is primarily due to two factors: the transistors are prevented from entering the saturation mode, and the output logic swings are smaller.) ECL gates are also faster than CMOS gates; thus, for applications where ultimate speed is the most important requirement, ECL has been, and remains, the technology of choice (for example, ECL circuits with gate delays as small as 35 ps have been demonstrated).⁷

However, as shown in Fig. 7-7 (which gives the projected delay as a function of the switch current of the scaled ECL circuit), in order for circuit delay to be in proportion to the lithographic dimension, the circuit's current - and hence its power dissipation - must remain essentially constant in scaling. With the power dissipation per gate in the milliwatt range, the integration of ECL logic chips will remain severely limited. For instance, a 10,000-gate ECL IC, with each gate dissipating 5 mW, would dissipate a total of 50 W!

Low-power Schottky TTL technology was found to provide slower performance than ECL, but to consume much less power. For MSI and LSI chips, Schottky TTL

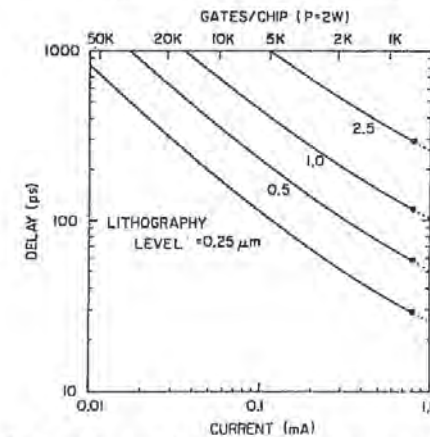


Fig. 7-7 ECL gate delay as a function of switch current and emitter-stripe width, predicted by bipolar scaling theory.¹⁴ (© 1979 IEEE).

offered the compromise of faster speed than CMOS, with lower power dissipation than ECL. In the VLSI era, however, CMOS was eventually able to provide speeds comparable to Schottky TTL, but with much smaller power-dissipation levels (see chap. 6, Fig. 6-7b). CMOS has thus tended to displace Schottky TTL technology. Today, ECL remains an important digital logic family, and BiCMOS-based circuits also incorporate bipolar devices in logic circuits. A relatively new family, *non-threshold logic* (NTL; see Fig. 7-8) has recently been drawing attention as a possible replacement for ECL, since NTL-with-emitter-follower circuits exhibit speeds comparable to those of ECL with lower power dissipation. NTL circuits as fast as 30 ps at 1.48 mW have been demonstrated.⁸

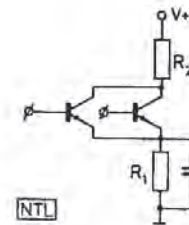


Fig. 7-8 Circuit diagram of non-threshold logic gate (NTL).

7.3 MAXIMIZING BIPOLAR-TRANSISTOR PERFORMANCE THROUGH DEVICE DESIGN AND PROCESSING TECHNOLOGY

The desired device characteristics of bipolar transistors include the following: high current gain (large value of β), high-frequency ac operation, fast switching speed, high device-breakdown voltages, minimum device size (to achieve high functional density), and high reliability of device operation.

In order for high-frequency ac performance and fast switching speed to be achieved, the parasitic resistances of the transistor (i.e., R_E , R_B , and R_C), and the parasitic junction capacitances (i.e., C_{EB} , C_{CB} , and C_{CS}) must be minimized. In addition, high-level injection effects (e.g., the Kirk effect) should be avoided. Finally, for faithful amplification of ac signals, the Early voltage must be high.

This section describes how these characteristics can be obtained through device design and processing technology.

7.3.1 Current Gain

As mentioned earlier, two of the figures of merit used to describe the performance of bipolar transistors are the current gains α and β . If the leakage current across the reverse-biased C-B junction is a negligibly small fraction of I_C , α can be expressed as the product of the emitter-injection efficiency, γ , and the base-transport factor, α_T :

$$\alpha = \gamma \alpha_T \quad (7-5)$$

Here γ is defined to be the ratio of the current in the base due to injection of electrons from the emitter, I_{En} , and the total emitter current (which is the sum of I_{En} and the current due to injection of holes into the emitter from the base, I_{Ep}):

$$\gamma = \frac{I_{En}}{I_{En} + I_{Ep}} \quad (7-6)$$

and α_T is defined as the fraction of I_{En} that reaches the C-B junction, as follows:

$$\alpha_T = \frac{\text{electron current reaching the collector}}{\text{electron current injected into the base}}$$

Accordingly, β will be given by

$$\beta = \frac{\gamma \alpha_T}{(1 - \gamma \alpha_T)} \quad (7-7)$$

For a transistor with a large β (i.e., $\alpha \rightarrow 1$), we can write

$$\frac{1}{\beta} \cong 1 - \gamma \alpha_T \cong \frac{1}{2} \left(\frac{W_B}{L_B} \right)^2 + \left(\frac{D_E G_B}{D_B G_E} \right) \quad (7-8)$$

where the first term is due to recombination within the base and the second to the injection into the emitter. In these terms, W_B is the base width (defined as the distance between the edges of the two depletion regions in the base); L_B is the minority-carrier diffusion length in the base; D_B and D_E are the diffusion coefficients of carriers in the base and emitter, respectively; and G_B and G_E are the base and emitter Gummel numbers (so named because they were first described by H. K. Gummel). These two parameters are defined as

$$G_B = \int_{\text{base}} N(x) dx \quad \text{and} \quad G_E = \int_{\text{emitter}} N(x) dx \quad (7-9)$$

and represent the total number of active impurity atoms in the base and emitter regions (per cm^2). The minimum value of G_B is $\sim 10^{12} \text{ cm}^{-2}$, because lower values would produce transistors with unacceptably small punchthrough-voltage values. (Typical values of N_B are in the $10^{12} - 10^{13} \text{ atoms/cm}^2$ range.) The typical range of G_E is 10^{15} cm^{-2} to 10^{16} cm^{-2} . Thus, with Eq. 7-8 used to find β , the maximum theoretical current gain in conventionally designed bipolar transistors is seen to be $\sim 10^4$.

For large current gain, Eq. 7-8 should be as small as possible. This in turn implies that the G_E/G_B ratio should be high, W_B should be small, and L_B should be large. A large G_E/G_B ratio also implies that the emitter must be much more heavily doped than the base. In modern, high-performance transistors, $W_B \ll L_B$ (i.e., L_B is greater than $10 \mu\text{m}$ for typical base doping concentrations, and W_B is less than $1 \mu\text{m}$ in low-voltage logic devices). Hence, the second term in Eq. 7-8 determines the current gain.

On the other hand, when the emitter is as heavily doped as it is in modern transistors, the silicon band gap in the emitter region is reduced and the recombination rate of minority carriers is increased. Both effects lower the emitter-injection efficiency. As a result, observed current gains in transistors with conventionally formed emitters (e.g., emitters formed by direct diffusion or ion implantation of the emitter dopant into the substrate) are not as high as those predicted by Eq. 7-8. If even higher doping concentrations were used in the emitter, the result would be decreased emitter-base junction breakdown voltages and increased emitter-base junction capacitances. Advanced emitter structures have been developed to overcome this problem and these will be discussed in section 7.7.

Note that in analog applications a high value of β (≥ 100) is generally desired, but in digital applications a lower value is often acceptable (in some cases as low as 20).⁹ On the other hand, in fast-switching transistors (which will be discussed next), small base-storage times and small intrinsic base resistances are needed. To reduce the base-storage times, the base width must be minimized, subject to base punchthrough. The minimum base doping to avoid punchthrough is proportional to $(1/W_B)^2$; therefore, the minimum G_B is proportional to $1/W_B$ and needs to be large if W_B is small. Since smaller

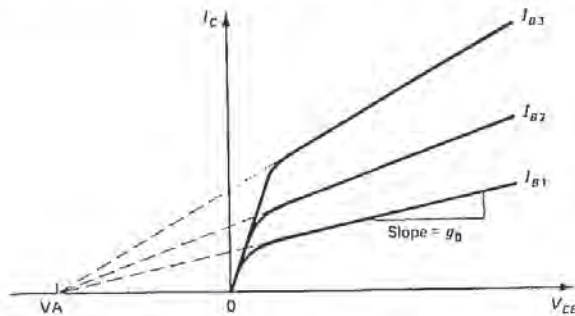


Fig. 7-9 Effect of base-width modulation on the C-E I-V output characteristics. The intercept on the voltage axis of the extrapolated curves is the Early voltage (V_A).

current gains can be tolerated in digital applications, some gain can be traded off by increasing the doping in the base to achieve a narrower W_B for a high switching speed.

7.3.2 The Early Voltage

The base width in bipolar transistors decreases as the C-B junction voltage increases because the depletion region increases in width. This effect is called *base-width modulation* (or the *Early effect*, after the scientist who first correctly explained its basis). The degree to which this effect impacts the operating characteristics of a bipolar transistor is represented by the Early voltage, V_A (volts). In the collector-emitter configuration, V_A is given by

$$V_A = I_C / (\partial I_C / \partial V_{CE}). \quad (7-10)$$

The dependence of $\partial I_C / \partial V_{CE}$ on I_C results in typical transistor output characteristics as shown in Fig. 7-9. Extrapolation of the characteristics of Fig. 7-9 back to the V_{CE} axis gives an intercept that represents V_A .

A large value of V_A (> 100 V) is desired in analog bipolar circuits for two reasons:

1. The *open-circuit voltage gain*, a_o (defined as the small-signal low-frequency voltage gain in the common-emitter configuration), is approximately found from¹⁰

$$a_o \equiv q V_A / kT. \quad (7-10a)$$

Since a_o is the maximum voltage gain that can be obtained from a bipolar transistor, it is a significant parameter in analog circuits. Because Eq. 7-10a indicates that a_o depends only on V_A and T , a large value of V_A will permit larger voltage gains to be achieved. (For a value of $V_A = 50$ V and room-temperature operation, $a_o \approx 2000$.)

2. Since $\Delta V_{CE} / \Delta I_C = V_A / I_C = r_o$ (where r_o is the small-signal output resistance of the transistor in the common-emitter configuration), small values of V_A imply a smaller output resistance, which is generally undesirable. Analog applications therefore require a minimum V_A of 30 volts, while a lower voltage is normally acceptable for digital applications (15-20 V).²²

The value of V_A is increased by raising the doping level in the base. Higher doping decreases the penetration of the depletion region as a function of increasing C-B reverse-bias voltage. In advanced bipolar transistors the doping levels are typically high enough (i.e., so as to prevent punchthrough, as will be described in section 7.3.4.2) that V_A is not a critical parameter in digital designs.

7.3.3 High-Level Injection Effects (Kirk Effect)

As the collector-current density (J_C) in an *npn* transistor increases, the density of electrons being transported across the C-B space-charge region also increases (i.e., this density is calculated from J_C / qv_1 , where v_1 is the electron scattering-limited velocity, $\sim 10^7$ cm/s). When the density of electrons crossing the C-B junction becomes comparable to the doping on the collector side of the space-charge region (a condition corresponding to *high-level injection*), the total charge in this region becomes significantly reduced from that under low-level injection conditions, leading to a lower electric-field gradient in the C-B junction. Since the collector voltage is constant, the integration of the electric field will remain the same, with or without current. However, with a lower maximum electric-field strength, the space-charge region edge in the base moves toward the collector, effectively increasing the base width. This phenomenon is known as the *base push-out effect** (or *Kirk effect*, after the author of the first paper analyzing it).²³

At first the base-width increase is small, and the current density at the onset of the Kirk effect, J_1 , is given by

$$J_1 = q v_1 N_{epi} \quad (7-11)$$

where N_{epi} is the doping concentration in the epi layer region of the collector. In practical device operation J_1 is easily exceeded in many applications. For example, in a transistor that has a 2×5 - μm emitter and an epi doping concentration of $5 \times 10^{15} \text{ cm}^{-3}$, Eq. 7-11 indicates that J_1 will be reached when $I_C = 0.8 \text{ mA}$.

As the current is increased beyond J_1 , the maximum electric-field strength continues to decrease. It eventually reaches zero, which represents a collapse of the original collector-base region. The base then spreads over the collector epitaxial layer and becomes very large. The current density at which this takes place is given by

$$J_{Ck} = q v_1 (N_{epi} + \{2K_s \epsilon_0 V_{CB} / q x_c\}) \quad (7-12)$$

*A more rigorous analysis of this effect is presented in chapter 7, reference 1.

where x_C is the thickness of the epi layer from the metallurgical C-B junction to the edge of the epi/buried-layer boundary.

The increase in base width leads to an increase in GN_B , and hence to a decreased current gain at high collector currents. In addition, the larger base width reduces f_T and the transistor speed. The transistor doping profiles must thus be optimized with this effect in mind. Once a transistor structure has been designed, operation at current levels above J_{Ck} for the device should be avoided.

From the equation used to calculate J_{Ck} , it is seen that this value can be raised by increasing the doping in the collector. Unfortunately, this also produces a decrease in the BV_{CBO} and BV_{CEO} of the device, as well as an increase in the collector-base junction capacitance, C_{CB} . Since transistors cannot be fabricated with smaller breakdown voltages than those required by the system power-supply voltage levels, the breakdown voltage constraints will set the maximum value of N_{epi} .

On the other hand a tradeoff between J_{Ck} and C_{CB} can be made, based on the operating point of the transistor. That is, if the transistor is to be operated at low current densities, the speed will be limited by C_{CB} . Thus, C_{CB} should be minimized (by reducing N_{epi}), since the Kirk effect will not impact the speed performance. However, for high-current applications, the speed will be limited by high-level injection effects, and C_{CB} can thus be allowed to rise when N_{epi} is increased in order to minimize speed degradation due to the Kirk effect.

Finally, it should be noted that in digital ICs, where BV_{CEO} is smaller than in analog ICs, a larger value of collector doping can be tolerated. This also implies that higher collector currents can be sustained in such devices without the onset of the Kirk effect.

7.3.4 Operating-Voltage Limits in Bipolar Transistors

The mechanism of avalanche breakdown in a pn junction limits the maximum reverse-bias voltages that can be applied to a junction and is also responsible for some of the maximum operating voltage values in bipolar transistors. The breakdown voltage of the C-B junction with $I_E = 0$ (BV_{CBO}) is due to this mechanism. The doping on the lightly doped side of the collector typically determines BV_{CBO} (except in cases of *reachthrough*, as will be discussed in section 7.2.4.1).

In C-B junctions in planar bipolar transistors, however, the pn junction is formed by diffusion into the silicon at openings in the surface SiO_2 layer. Thus, impurities will diffuse downward and sideways under the edges of the SiO_2 layer. Hence, the junction has a plane region with nearly cylindrical edges (as can be seen in Fig. 7-4c). Such junction curvature enhances the electric field in the curved part of the depletion region (see chap. 5, Fig. 5-10a), which reduces the breakdown voltage below that predicted by one-dimensional junction theory. Figure 5-10a also shows that as the junction depths get shallower (making the radius of curvature, r_j , of the spherical and cylindrical structures smaller), the breakdown voltage is significantly reduced, especially for small substrate-impurity concentrations. Thus, the measured value of BV_{CBO} is less than the value of the breakdown voltage that would be predicted if the junction were only one-

dimensional (as would be the case if the transistor had a structure like that shown in Fig. 7-1a).

The breakdown voltage of the E-B junction with $I_C = 0$ (BV_{EBO}) is also limited by avalanche breakdown. Since the emitter is so heavily doped, it is the base region that is the lightly doped side of this junction. Since the base is more heavily doped than the collector, BV_{EBO} is much smaller than BV_{CBO} and is typically 6 – 8 V. Since the base is formed by diffusion from the surface, the doping in the base region is highest there so the E-B junction will undergo breakdown first in this area. One of the reasons for performing a drive-in step following predeposition when a diffusion step is used to form the base is to reduce the surface concentration of boron, which thus increases BV_{EBO} to a sufficiently high value.

While high doping in the base region adjacent to the emitter sidewalls is advantageous for reducing the base series resistance, R_B (see section 7.5.2.6), if the doping is too high, BV_{EBO} becomes too small. Tradeoffs must therefore be made in the selection of the doping value in this region. The curvature of the junction enhances the electric field and reduces BV_{EBO} in the same manner as described for BV_{CBO} .

For a bipolar transistor with a voltage applied between the emitter and collector and in which $I_B = 0$, breakdown occurs at BV_{CEO} . In this situation, the breakdown may be due either to avalanche breakdown multiplied by the effect of transistor action, or to punchthrough.

In the first instance, the E-B junction will be very slightly forward-biased as a result of the base region's acquiring a potential that is intermediate between the potentials of the collector and emitter. Thus, the collector current will also be greater than I_{CBO} (the leakage current of the C-B junction when $I_E = 0$) because most of the carriers injected from the slightly forward-biased E-B junction will be collected by the collector. This increased current will precipitate avalanche breakdown at a smaller applied C-B voltage, which implies that BV_{CEO} must be smaller than BV_{CBO} . An empirical expression has been derived that relates BV_{CEO} and $BV_{CBO}(\text{plane})$:¹¹

$$BV_{CEO} = \frac{BV_{CBO}(\text{plane})}{(B)^{\frac{1}{n}}} \quad (7-13)$$

where $BV_{CBO}(\text{plane})$ is the collector-base breakdown voltage along the planar (flat) portion of the base region. The empirical parameter n has been found to have values of between 3 and 6.

7.3.4.1 Reachthrough Breakdown. Since most bipolar transistors used in ICs are fabricated with collector regions that consist of a lightly doped epitaxial layer over a heavily doped buried layer, BV_{CBO} may in some cases be limited by *reachthrough breakdown*. If the epitaxial-layer thickness, t_{epi} , is too small, under increasing reverse-bias voltage the depletion region of the C-B junction will entirely penetrate (or "reach through") the epitaxial layer before the predicted BV_{CBO} voltage has been reached. If the voltage is increased beyond this point, the depletion region must enter the heavily doped buried layer, resulting in BV_{CBO} values smaller than those

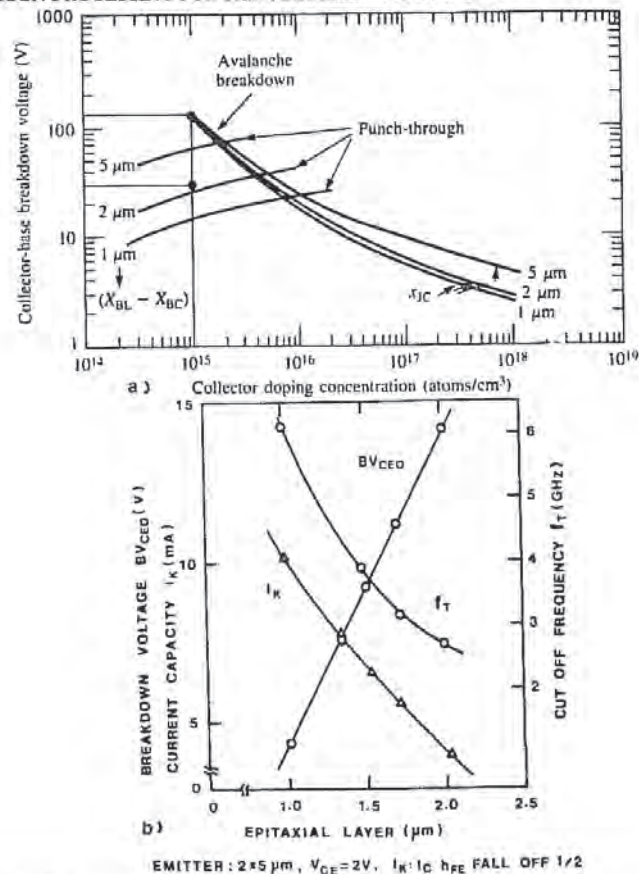


Fig. 7-10 (a) Collector-base-junction breakdown voltage as a function of collector-doping concentration with collector-base junction depth and punchthrough limits as parameters.¹²⁵ Reprinted by permission of the publisher, The Electrochemical Society, Inc. (b) Epitaxial layer thickness dependence of cutoff frequency, f_T , breakdown voltage BV_{CEO} , and knee current, I_K .¹² (© 1987 IEEE).

predicted merely by the doping in the epi layer. Figure 7-10a gives the collector-base junction-breakdown voltage as a function of the collector-doping concentration, with the collector-base junction depth and the thickness of the lightly doped region of the epi layer ($x_{BL} - x_{BC}$) as parameters.¹²⁵ Figure 7-10b shows how a decrease in t_{epi} causes BV_{CEO} to decrease in a bipolar transistor designed for use in 5-V digital ICs.¹² Section

7.12.1.2 gives another example of a BV_{CEO} calculation in which this effect is considered.

7.3.4.2 Punchthrough Breakdown. In some cases the value of BV_{CEO} can be limited by *punchthrough* rather than by avalanche breakdown. If the base width is too narrow and/or the doping level in the base is too small, the depletion region of the C-B junction will penetrate the base region and join the E-B junction depletion region before the avalanche-limited value of BV_{CEO} has been reached. (Section 7.2.2 describes such base-width modulation.) If the collector voltage is increased beyond this value, the potential-energy barrier at the E-B junction will be lowered by ΔV , as shown in Fig. 7-11. As a result, a large emitter current (and correspondingly, a large collector current) can flow in the device, even with no avalanche breakdown. (Note that a large current flow across the C-B junction when both the E-B and C-B junctions are reverse biased is a manifestation of breakdown.)

Only a slight increase in the collector voltage beyond that needed to merge the two depletion regions is required for a large increase in I_C . Thus, punchthrough appears as a form of device breakdown. To prevent punchthrough, appropriate combinations of base doping and base width for the circuit operating voltages must be used. Most bipolar transistors are designed so that avalanche breakdown of the C-B junction will occur before punchthrough.

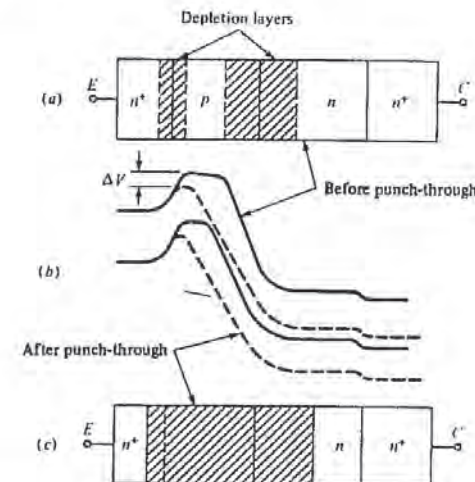


Fig. 7-11 Punchthrough in an npn transistor: (a) space-charge regions before punchthrough, (b) energy-band diagram (and space-charge regions) after punchthrough. Copyright, 1988, McGraw-Hill Book Company. Reprinted with permission from ref. [2].

7.3.4.3 Breakdown Voltage and High-Level Injection Limits in Advanced Bipolar Transistors. To maintain the sufficiently high current levels needed to achieve minimum gate delays (see section 7.3.8), the current density is increased as the emitter area is scaled down. Higher collector doping levels must be used to avoid the Kirk effect. Furthermore, higher doping concentrations in the base are needed to avoid punchthrough of the narrow base regions. Thus, as of 1990, the base and collector doping concentrations in advanced bipolar devices are typically on the order of $1 \times 10^{18} \text{ cm}^{-3}$ and $1 \times 10^{17} \text{ cm}^{-3}$, respectively. The resulting peak electric field in the C-B junction is high enough to produce intolerably large avalanche-multiplication currents at low C-B reverse bias voltages (e.g., less than 3 V). Lu and Chen describe this problem in detail in reference 13.

7.3.5 Parasitic Series Resistances in Bipolar Transistors

Due to the planar structure of IC transistors, the current that flows from the emitter to the collector must travel parallel to the surface of the wafer after it has been transported vertically through the base. It must then flow upward to a contact at the wafer surface. Because the current-flow paths within the transistor have significant resistivity, parasitic series resistances exist in each of the three bulk regions of the device: a collector series resistance, R_C ; a base series resistance, R_B ; and an emitter series resistance, R_E . Each must be as small as possible.

7.3.5.1 Collector Series Resistance, R_C . A low value of R_C is important in both high-frequency ac circuits and high-speed digital circuits. In the former, a large R_C value would decrease the cutoff frequency and introduce a voltage drop that would increase V_{CEsat} . In the latter, a large R_C value would reduce both f_T (see section 7.3.7) and the maximum current-drive capability. Figure 7-12 shows the components of R_C .

Several approaches are used to minimize R_C . First, the lightly doped epitaxial layer used as the collector is doped as much as can be tolerated with regard to breakdown-voltage and junction-capacitance considerations. This means that in low-voltage digital devices (in which a low value of V_{CEsat} is critical but lower operating voltages are

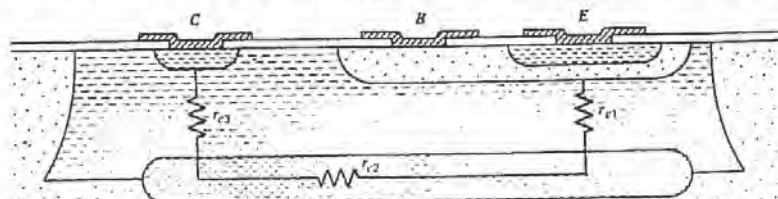


Fig. 7-12 Components of collector resistance, R_C . Copyright, 1984, John Wiley and Sons, Inc. Reprinted with permission from ref. [10].

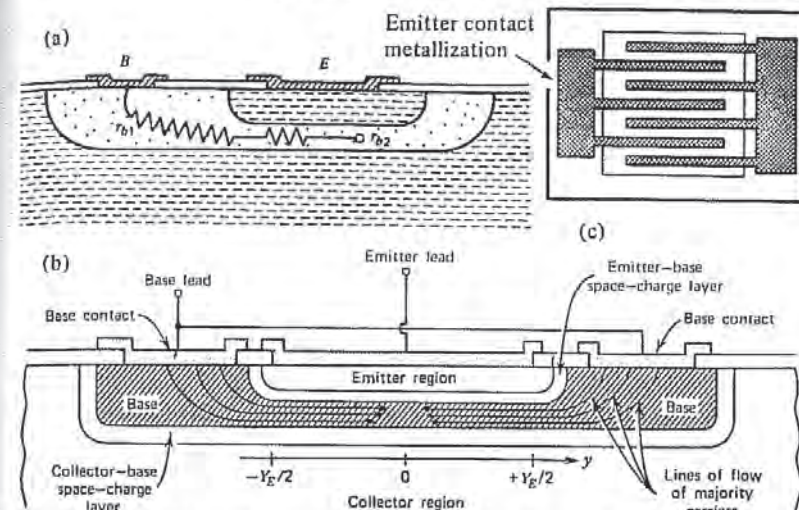


Fig. 7-13 Base-resistance components for the npn bipolar transistor. Copyright, 1984, John Wiley and Sons, Inc. Reprinted with permission from ref. [10]. (b) Cross section of a transistor under forward active bias. The base current is supplied from two side contacts and flows toward the center of the emitter causing the base-emitter voltage to vary with position. Copyright, 1986, John Wiley and Sons, Inc. Reprinted with permission from ref. [1].

used), the epi layer is doped somewhat more heavily than it is in analog transistors, which are operated at higher supply voltages. Second, heavily doped buried layers (subcollectors) are used beneath the more lightly doped collector regions to provide a parallel high-conductivity path while maintaining the high C-B breakdown-voltage capability provided by the lower-doped epi.

Inclusion of the buried layer reduces R_C from the kilohm range to a few hundred ohms in typical devices. For some applications, however, even this value of R_C is too high, so an additional, heavily doped, deep n^+ region is diffused beneath the collector contact until it reaches the buried layer. Such deep-collector contacts (also called *plugs* or *sinks*) can reduce R_C to the order of 10Ω . Details of the process steps used to form buried layers and deep-collector contacts will be discussed later in this chapter in the sections describing process sequences for bipolar and BiCMOS technologies. Chapter 2 of reference 10 gives a first-order-calculation technique for R_C when the sheet-resistance values of the buried and epi layers and their device-structure dimensions are known.

7.3.5.2 Base Series Resistance, R_B . Low values of R_B are also needed to achieve adequate gain at high frequencies and to obtain high speeds in digital bipolar ICs. As illustrated in Fig. 7-13a this resistance consists of two parts. The first part is

the resistance of the so-called *extrinsic* (or *inactive*) *base region*, R_{B1} (which makes up the path between the base contact and the edge of the emitter region). The second part, R_{B2} , is the resistance between the edge of the emitter and the site within the *intrinsic* (or *active*) *base region* at which the current is actually flowing. R_{B2} is also known as the *base-spreading resistance*.

One technique for reducing R_{B1} is to use a separate process step to selectively add dopants to the extrinsic base region in order to increase its conductivity. If the extrinsic base is doped too heavily, however, excessive emitter-base capacitance will result along the emitter sidewall, and a low emitter-base breakdown voltage may also be produced. Self-aligned techniques have been exploited to reduce R_{B1} by decreasing the distance between the base contact and the intrinsic base region. This approach is very effective and will be discussed in more detail later.

7.3.5.3 Base-Spreading Resistance, R_{B2} (and Emitter Current Crowding). R_{B2} is not well modeled by a single resistor, for two reasons. First, this effect is actually distributed (or spread) throughout the intrinsic base region, and two-dimensional effects are important.* Second, the effect of *current crowding* sets in, even at moderate current levels, causing most of the carriers to be injected into the base at the periphery of the emitter diffusion (since the base current in the narrow base flows from the edges of the emitter to the center [parallel to the wafer surface; see Fig. 7-13b]). An ohmic drop thus occurs between the emitter's edge and the center. Because the base is so narrow, the distributed resistance along it is quite high, and the value of the ohmic drop can cause the base-emitter voltage drop to vary with position along the E-B junction. The forward bias of the emitter junction will be largest at the edge of the emitter diffusion. Therefore, most of the emitter current will be injected into the base near the edge, and the remainder will decrease to a minimum at the center.

Although current crowding can have the beneficial effect of reducing R_B (in the extreme, the effective value of R_B approaches R_{B1}), it can also give rise to localized heating at current levels that might be tolerable only if the current were uniformly distributed. Furthermore, the Kirk effect can occur at lower currents because of the uneven current density across the active region of the device.

The most effective solution is to distribute the current along a relatively long emitter edge, thereby reducing the current density at any one point. In bipolar transistors designed for low-noise and/or high-frequency applications, the emitter-edge length is maximized by designing the emitter layout in the form of many long, narrow stripes with base contacts between them (this layout approach is sometimes called an *interdigitated geometry*, see Fig. 7-13c).

Since R_{B1} and R_{B2} are both inversely proportional to the *emitter* length (i.e., a longer emitter implies a wider horizontal base region) and R_{B2} is proportional to the emitter width, it is preferable for emitters to be long and narrow (i.e., with width-to-length ratios of one-third, or less).

* An analysis that rigorously formulates two expressions for the effective value of R_{B2} is given in reference 47.

7.3.5.4 Emitter Series Resistance, R_E . Until recently, bipolar device performance has not been significantly impacted by R_E . Since the emitter is a shallow, heavily doped region, the series resistance of its bulk region is quite small. In addition, the contact resistance of ohmic metal-semiconductor contacts has also been small enough that it has not increased R_E to values that would significantly impact the performance of analog or digital ICs. However, the introduction of polysilicon-emitter structures for advanced bipolar ICs has brought with it the penalty of an increased emitter resistance, and its value in such structures can significantly degrade their performance. This topic will be discussed further in section 7.7.

7.3.6 Parasitic Junction Capacitances in Bipolar Transistors

The parasitic capacitances associated with the E-B, C-B, and collector-substrate (C-S) junctions of the planar bipolar transistor can limit the high-frequency and switching characteristics of these devices. We will describe in section 7.3.9 how these capacitances impact the propagation delays of bipolar digital ICs.

The C-B and E-B capacitances in junction-isolated bipolar transistors are composed of the capacitances of the C-B and E-B junctions along the flat bottom portions and along the sidewalls. The C-S capacitance in junction-isolated bipolar transistors with buried-layer collectors consists of three parts: (1) that of the junction between the buried layer and the substrate; (2) that of the sidewall of the isolation diffusion; and (3) that between the epitaxial material and the substrate. In general, the value of the capacitance is proportional to the area of the junction and the doping concentration on the lightly doped side. (Techniques for first-order calculations of the values of these capacitances in junction-isolated transistors are given in chapter 2 of reference 10.) Since junction-isolated bipolar transistors for digital ICs are operated at smaller power-supply voltages than those used in analog ICs, thinner epi layers are used. Hence, the areas of the transistors are smaller, resulting in lower values of parasitic capacitance.

The benefit of reduced parasitic capacitance is extended even further if oxide or trench isolation is used instead of junction isolation. In such structures the sidewall-capacitance components are reduced (or eliminated), and the ability to use smaller feature sizes reduces the area of the devices (and hence the planar-capacitance components). Examples of parasitic capacitance values (as well as of the device layouts) for each of these four types of IC bipolar transistors are given in Fig. 7-14.

7.3.6.1 Storage Capacitances in Bipolar Transistors. The capacitances associated with the storage of minority carriers in the bulk regions under conditions of forward bias can also significantly impact the switching speed of digital bipolar circuits. This topic is discussed further in section 7.3.9.3.

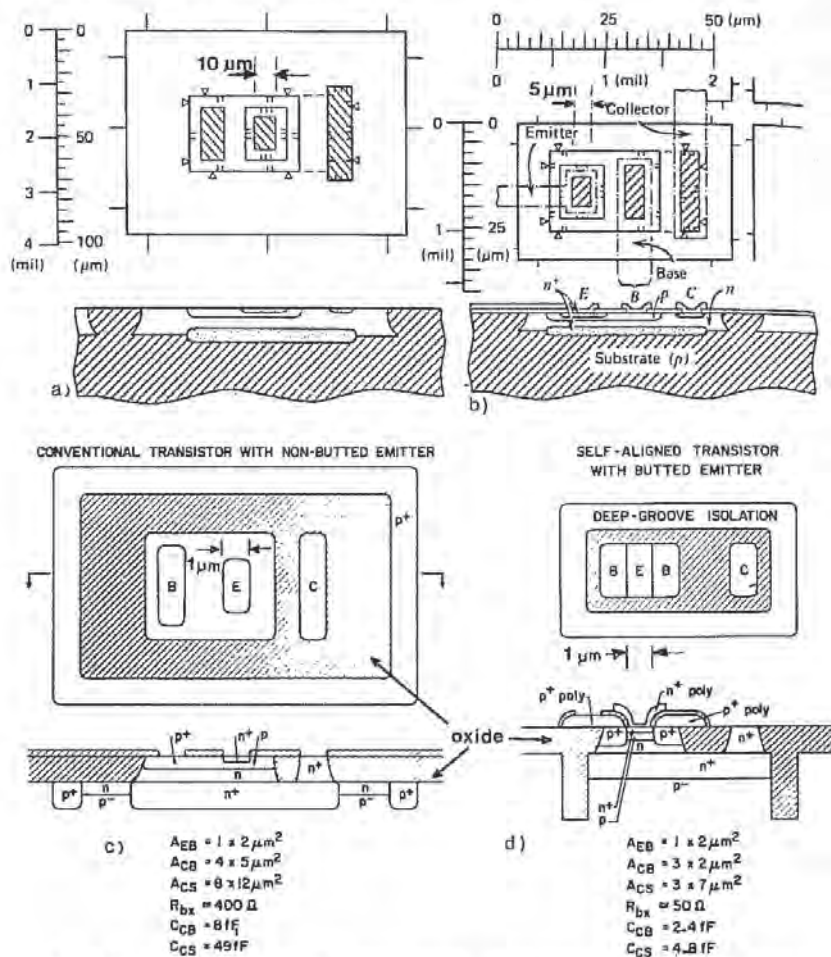


Fig. 7-14 Top views (showing dimensions), cross-sections and device structure characteristics of an: (a) junction-isolated amplifying (35 V) bipolar transistor, 10- μm minimum dimensions;¹²⁶ (b) junction-isolated digital (5 V) bipolar transistor, 5- μm minimum dimensions;¹²⁶ (c) oxide-isolated, digital (5 V) bipolar transistor, 1- μm minimum dimensions;²⁶ and (d) trench-isolated, digital (5 V) bipolar transistor, 1- μm minimum dimension.²⁶ [(a) and (b) Copyright, 1972 by Litton Educational Publishing, Inc., Reprinted with permission of Van Nostrand Reinhold Company from reference [126].] (c) and (d) (© 1986 IEEE).]

7.3.7 Bipolar Transistor Unity-Gain Frequency, f_T

Another important characteristic of bipolar transistors is their ability to provide current gain at high frequencies. The figure of merit that is typically used to describe the high-frequency behavior of a bipolar transistor is the *unity-gain frequency*, f_T , which is the frequency at which the current gain of the transistor decreases to unity. An approximate expression for f_T is given by

$$f_T = \frac{1}{2\pi\tau} \quad (7-14)$$

where τ is the *transit time* of the transistor. An approximate expression for τ is

$$\tau = \frac{W_B^2}{\eta D_B} + (C_{CB} + C_{CS}) R_C \quad (7-15)$$

The first term ($W_B^2/\eta D_B$), called the *base transit time* (τ_B), represents the average time per carrier spent in diffusing across the neutral base region of width W_B . D_B is the diffusion coefficient of minority carriers in the base, and the parameter η is associated with the doping profile in the base. For a uniformly doped base region, $\eta = 2$, whereas in a graded base, η can be larger by a factor of 10. The second term is the delay associated with the charging of the capacitances connected to the collector node (C_{CB} and C_{CS}) through R_C .

For a minimum value of τ to be obtained, several features of the bipolar transistor structure must be optimized. The base width is made as narrow as possible, and buried layers and deep-collector contacts are used to minimize R_C . In addition, light doping, oxide isolation, and reduced device area are all employed to decrease the values of C_{CB} and C_{CS} .

Note that τ in fact also depends on I_C . For medium to high values of I_C , the time required to charge C_{CB} and C_{CS} becomes small, and thus approaches a constant value (τ_B). At low values of I_C , the term involving C_{CB} and C_{CS} dominates. For very high I_C values, however, τ again increases as a result of high-level injection effects (Kirk effect).

7.3.8 First-Order npn Device Design

A first-order method for designing bipolar IC transistor structures, proposed by Solomon and Tang,¹⁴ is graphically depicted in Fig. 7-15 (which also summarizes the discussion of bipolar-device parameters presented in this section).¹⁵

The circuit applications specify several characteristics that set the device design boundaries, including collector-current density, current gain, BV_{CEO} , R_B and τ_B . The epi layer thickness and doping concentration are then selected based on these requirements and on Kirk effect considerations (with the emitter area also used as a parameter to establish the collector current density). The base profile is then optimized

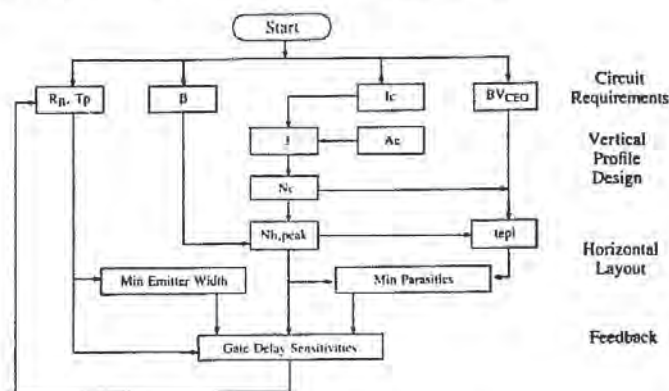


Fig. 7-15 First order npn device design. The vertical profile is set to optimize device delay, while optimal horizontal design minimizes the parasitics. From [14] © 1979 IEEE.

based on the design decisions and on the constraints of meeting the target β , R_B , and τ_B specifications.

7.3.9 Switching-Speed Behavior in Bipolar ICs

MOS and bipolar transistors have separate advantages with respect to their use in digital circuits. Circuits built with CMOS consume less power than bipolar circuits and can be fabricated with higher packing densities, while those built with bipolar transistors can operate at higher speeds with minimum-size transistors. In addition, small voltage swings are possible with bipolar gates, resulting in a relatively small speed degradation due to on-chip wiring capacitance. Finally, the power-supply voltage and the gate output-voltage swings of bipolar digital circuits can be minimized independent of transistor size. Bipolar circuits are thus preferred for applications in which ultimate speed performance is desired.

7.3.9.1 Propagation-Delay-Time Calculation in Bipolar Transistors.

Historically, the unity-gain cutoff frequency, f_T , has also been widely used as the figure of merit for describing the switching speed of bipolar transistors. Unfortunately, as shown in Fig. 7-16a, this parameter often does not correlate well with the gate delays measured in actual ECL circuits. An alternative delay expression has been derived by Stork¹⁶ that agrees very well with reported data of ECL ring-oscillator delays (Fig. 7-16b). The objective in deriving this figure of merit was to obtain a simple delay expression that is most accurate when the minimum delay is approached.

For the simplified case in which the emitter-follower resistance equals the load resistance, Stork's expression for the minimum delay, τ_s , of the switching action of a bipolar transistor, is given by

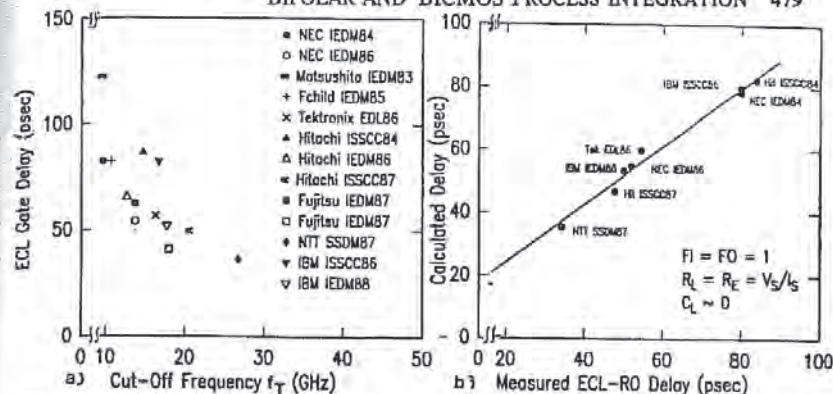


Fig. 7-16 (a) Scatter plot of ECL ring-oscillator speed: the correlation of published delay data with f_T is poor. (b) Calculated versus measured/published ECL ring-oscillator delay: the correlation is generally within 10%.¹⁶ © 1988 IEEE.

$$\tau_s = \sqrt{\left\{ 1 + \left(\frac{2\alpha R_B I_{EE}}{\Delta V} \right) \right\} \left(\frac{1}{2\tau_{fT}} \right) (3C_{CB} + C_{CS}) \left(\frac{\Delta V}{I_{EE}} \right)} \quad (7-16)$$

where ΔV is the total output-voltage swing of the gate, and I_{EE} is the current of the of the ECL gate's current source (see Fig. 7-6c).

7.3.9.2 Propagation Delay In Digital MOS versus Digital Bipolar

Circuits. Figure 7-17 shows two stages of a ring oscillator built with arbitrary devices. The logic swing is given by $\Delta V = V_H - V_L$. The load capacitance (C_L) is the sum of the junction capacitances (C_j), the on-chip wiring capacitance (C_w), and the input capacitance (C_{in}) of the device in the second stage being driven. In the case of a bipolar ring oscillator, C_{in} is a nonlinear capacitance with charge $Q = I_C \tau_B$ (where τ_B is the transit time of the bipolar transistor). With an MOS ring oscillator, the capacitance is the (linear) gate capacitance, C_g . As a first order approach, neglecting device parasitic series resistances, the propagation-delay times through the stages of these two circuits are given by

$$\tau_{dbipolar} \sim \{ (C_j + C_w) \Delta V / I_C \} + \tau_B \quad (7-17)$$

and

$$\tau_{dMOS} \sim \{ (C_j + C_w) \Delta V / I_{DS} \} + \{ C_g \Delta V / I_{DS} \} \quad (7-18)$$

Since the collector current of the bipolar transistor is exponentially dependent on V_{BE} (Fig. 7-17b), the current in the bipolar ring oscillator can be increased by several orders of magnitude by a small change in V_{BE} (e.g., ~60 mV at room temperature) without the need to increase the transistor size. According to Eq. 7-17, this means that the ideal

propagation-delay time can be decreased to τ_B , by increasing the value of I_C and/or reducing ΔV .

It is much more difficult to increase I_{DS} in the case of MOS circuits, due to the quadratic current/voltage characteristic (see chap. 5, Eq. 5-10, in section 5.1.4). While I_{DS} could be increased by making all of the MOS devices in the circuit larger, this would also increase C_g . In practice, the performance of digital circuits with minimum-size transistors is found to be strongly affected by the on-chip wiring capacitances. These capacitances will in fact prevent the minimum ideal propagation delay of the circuit from ever being reached.⁶

In summary, it can be said that the two main advantages of digital bipolar circuits over their digital MOS counterparts are that (1) optimum speed can be obtained with minimum size transistors, and (2) it is feasible to use small gate-output-voltage-swings as a result of the exponential dependence of output current on input voltage. This implies that digital bipolar circuits will continue to have an edge in the IC market whenever ultimate speed performance is required.

7.3.9.3 General Switching-Speed Behavior of Digital Bipolar Circuits.

In general, the propagation-delay time $\tau_{dbipolar}$ decreases when the current

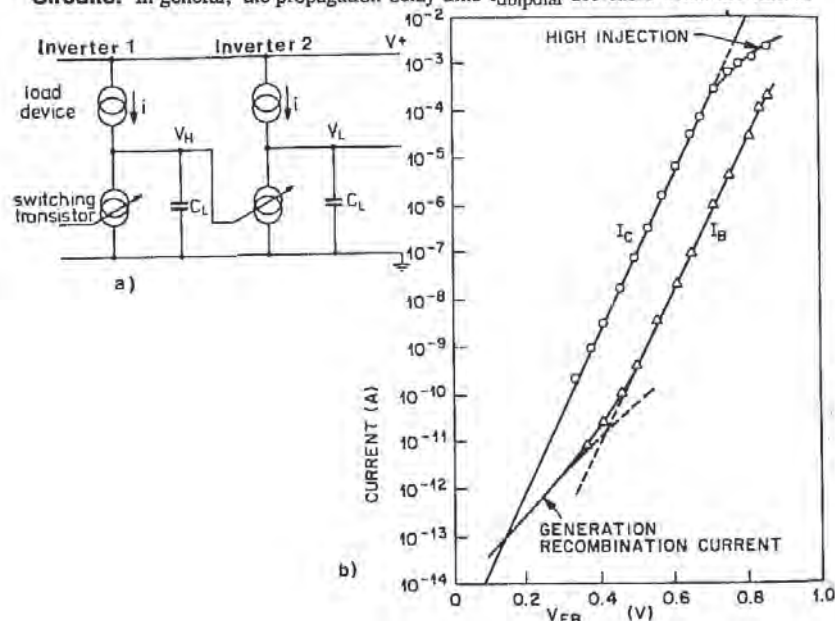


Fig. 7-17 (a) Two basic inverter stages of a ring oscillator built with arbitrary devices. C_L represents the total capacitance. (b) Collector and base currents as a function of V_{BE} .

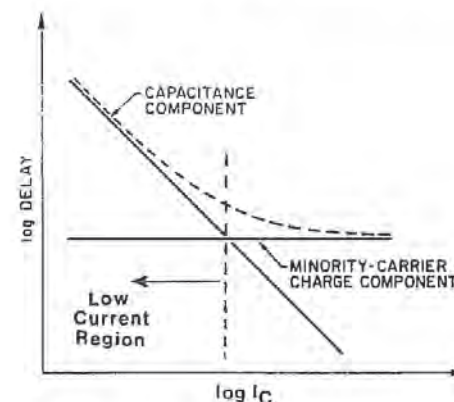


Fig. 7-18 Schematic of delay-versus- I_C characteristics of a typical bipolar logic circuit. It is assumed that the Kirk effect has been avoided by the proper tailoring of the collector doping profile.²⁶ (© 1986 IEEE).

per logic gate increases. At low current levels the storage of minority carriers can be ignored, and the switching speed is determined only by C_L (the sum of the junction and on-chip wiring capacitances). In such cases, $\tau_{dbipolar}$ is approximately equal to $C_L \Delta V / I_C$. If the series resistances are negligible, $\tau_{dbipolar}$ decreases linearly with increasing current, as shown in the low-current part of Fig. 7-18.

If the current is sufficiently increased, the storage of minority carriers begins to play a dominant role. Eventually, the minimum propagation-delay time, $\tau_{dbipolar}(\min) = \tau_B$, is reached. The Kirk effect can cause $\tau_{dbipolar}$ to increase with increasing current, but the effect can be minimized by adjusting the collector doping profile (see section 7.2.3).

In practice, the parasitic series resistances of the device, R_B and R_C , can also significantly increase $\tau_{dbipolar}(\min)$. At high currents, high values of R_B or R_C produce a large increase in $\tau_{dbipolar}(\min)$ because as the minority-carrier storage increases, the available current to discharge this charge is limited by the series resistances. If the wiring capacitance makes C_L relatively large, $\tau_{dbipolar}(\min)$ can increase even more as it begins to be limited by RC-time effects.

It can be concluded that the performance of digital bipolar circuits is optimized by doing the following:

- Minimizing C_L and ΔV .
- Minimizing the minority-carrier storage (by realizing a high f_T for the switching transistors and eliminating saturation effects - e.g., by using Schottky clamps or non-saturating logic circuits, such as ECL).

- Reducing the parasitic series-resistance values to acceptably low values.

The parasitic capacitances in junction-isolated (and even oxide-isolated) non-self-aligned bipolar transistors are so large that it is necessary to use rather large gate currents in order to reach propagation-delay times that are limited by the minority carrier storage component. For this reason, bipolar circuits are customarily associated with high power dissipation. The scaling of the device area with self-aligned structures, together with reductions in the parasitic-junction-capacitance values through the use of trench isolation, offers the promise of dramatically reduced power per gate in bipolar circuits. These two advances will be described in sections 7.8 and 7.9.

7.4 NON-OXIDE-ISOLATED BIPOLAR *n*pn TRANSISTOR STRUCTURES

Three types of planar transistor structures were developed in early bipolar integrated circuits, all of which used a reverse-biased *pn* junction to provide isolation between collectors on the same chip. These device structures, named after the processes used to fabricate them are as follows: (1) *standard buried collector* (SBC) transistors (Fig. 7-19a), (2) *collector-diffused isolation* (CDI) transistors (Fig. 7-19b), and (3) *triple-*

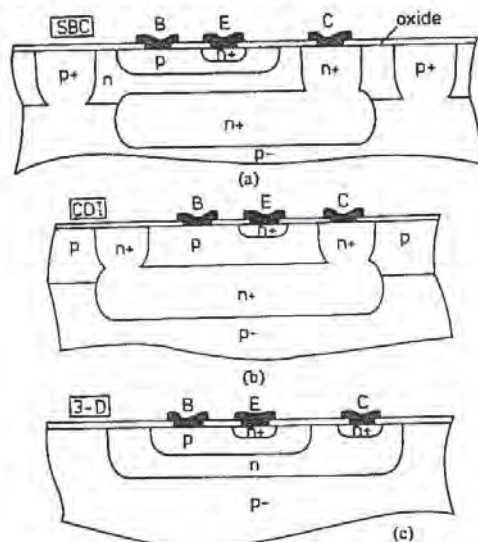


Fig. 7-19 Cross sections of three junction-isolated processes: (a) Standard buried-collector (SBC) process, (b) Collector-diffused isolation (CDI) process, (c) Triple-diffused (3D) process.¹² (© 1981 IEEE).

diffused (3D) transistors (Fig. 7-19c). The SBC transistor has been most widely used and a detailed description of the process flow used in its fabrication will be presented in section 7.5. The CDI process is described in chapter 2, section 2.1.1.2, and the 3D process is also discussed here.

7.4.1 Triple-Diffused (3D) Process

A cross-section of an *n*pn bipolar transistor fabricated with the triple-diffused process is shown in Fig. 7-19c.¹⁷ No buried layers or epitaxial layers are used. Instead, the collector is formed by means of ion implantation (and diffusion) directly into the substrate, as are the base and emitter regions. The chief advantages of this process are its simplicity (which allows for high yields and low manufacturing costs), and the fact that smaller-area transistors can be fabricated than is possible with the SBC process, when the same design rules are used (since no *p*-type isolation diffusions are needed to surround each collector). The benefit of process simplicity also makes the 3D transistor the bipolar structure of choice in low-cost digital BiCMOS technology (see section 7.12.1.1). An oxide-isolated, triple-diffused process has also been developed.¹⁸ The incorporation of high-pressure oxidation to improve the properties of devices fabricated using this process has also been described.¹⁹

Higher values of BV_{CEO} can also be obtained with smaller 3D devices than with either CDI or thin-epi-layer SBC transistors (the latter of which are used for 5-V digital ICs). Thus, 3D transistors have been used in such applications as analog-to-digital converters with greater than 8-bit resolutions (such circuits require the use of operating voltages greater than 5 V). The main disadvantage of 3D transistors is their high value of R_C , which results from the absence a buried-layer subcollector.

7.5 STANDARD-BURIED-COLLECTOR (SBC) PROCESS

7.5.1 Characteristics of *n*pn Transistors Fabricated with the SBC Process

A transistor fabricated with the SBC process has a collector region that consists of a heavily doped n^+ buried layer and a lightly *n*-doped epitaxial layer (Fig. 7-19a). The collectors are isolated from one another by *p* diffusions that extend all the way through the epi layer to contact the *p*-substrate. The base and emitter regions are formed by diffusions after the isolation structures have been created.

Until the mid-1970s the planar, standard-buried-collector process with junction isolation was the workhorse for the fabrication of both analog and digital bipolar ICs, since devices fabricated with this process had the best mix of characteristics for the largest number of applications.

Because it is easier to control the base width in the SBC than in the CDI process, higher current gains can be obtained (the graded doping profile in the base also helps to

Table 7-2 Typical Design Parameters for Junction-Isolated IC BJTs.

	Amplifying (Junction-Isolated)	Switching (Junction-Isolated)	Switching (Oxide-Isolated)
<i>Epitaxial Film</i>			
Thickness	10 μm	3.0 μm	1.2 μm
Resistivity	1 $\Omega\text{-cm}$	0.3 - 0.8 $\Omega\text{-cm}$	0.3 - 0.8 $\Omega\text{-cm}$
<i>Buried Layer</i>			
Sheet resistance		$\sim 20 \Omega/\square$	$\sim 30 \Omega/\square$
Up diffusion	2.5 μm	1.4 μm	0.3 μm
<i>Emitter</i>			
Diffusion depth in base	2.5 μm	0.8 μm	0.25 μm
Sheet resistance	5 $\Omega/\square$	12 $\Omega/\square$	30 $\Omega/\square$
<i>Base</i>			
Diffusion depth	3.25 μm	1.3 μm	0.5 μm
Sheet resistance	100 $\Omega/\square$	200 $\Omega/\square$	600 $\Omega/\square$
<i>Substrate</i>			
Resistivity	$\sim 10 \Omega\text{-cm}$		$\sim 5 \Omega\text{-cm}$
Orientation	(111)		(111)

produce such high gains). In addition, the use of an n -type epitaxial layer together with buried layers allows BV_{CBO} to be made higher than is possible in the CDI process, since the collector doping at the edge of the junction can be made much lower. Furthermore, R_C is much lower than it is in transistors fabricated with the 3D process (which does not use buried layers).

As will be explained more fully later, junction-isolated SBC transistors are unsuitable for digital VLSI circuits and have been replaced for this application by oxide-isolated SBC devices. Nevertheless, the junction-isolated SBC process continues to be the principal bipolar process for the fabrication of analog ICs operated with power-supply voltages in excess of 10 V.

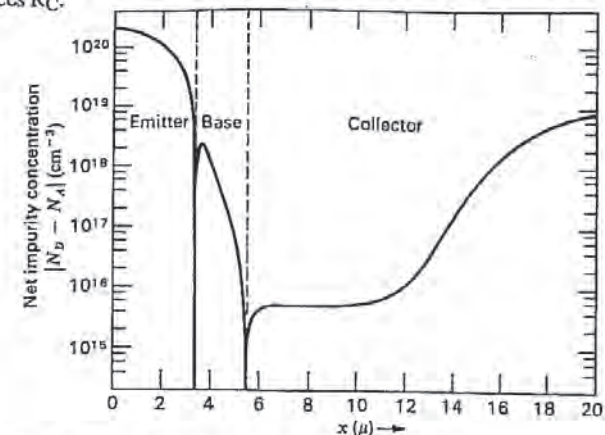
As shown in Table 7-2, SBC transistors used in analog ICs (primarily for amplification functions) are fabricated with different device structures than those used in digital ICs (in which the devices are used for switching). The differences result mainly from the higher operating-voltage levels of analog ICs (i.e., in digital ICs the typical power-supply voltage is only 5 V). As a result, analog-IC transistors must have thicker epitaxial layers and lighter collector-doping levels (at the edge of the base-collector junction).

Since digital ICs must often operate at higher current levels, the collector doping level must be increased in order to avoid device performance degradation caused by the Kirk effect. For example, epitaxial layers in analog junction-isolated bipolar-SBC ICs have thicknesses of between 7 and 15 μm and resistivities of 1 $\Omega\text{-cm}$, while their digital

counterparts have epi thicknesses of between 2.0 and 5.0 μm and resistivities of 0.3-0.8 $\Omega\text{-cm}$. The doping profile of a digital SBC transistor with an arsenic-doped emitter along a coordinate perpendicular to the surface and passing through the emitter, base, and collector is shown in Fig. 7-20. In this example, a single implant is used to form both the extrinsic and intrinsic base regions. Because the diffusion of the base region is significantly altered by the diffusion of the arsenic as it forms the heavily doped emitter, the intrinsic-base profile does not decrease smoothly into the silicon.

As an example of how the thickness and doping concentration in the epi layer impacts the breakdown voltage of the device, circuits operated at power-supply voltages of 36 V must use devices in which BV_{CEO} exceeds this value. Using Eq. 7-13 and assuming that $n = 4$, $\beta = 100$, and the base-junction depth is 3 μm , for a BV_{CEO} of 36 V, we calculate that $BV_{CBO}(\text{plane})$ must correspondingly be greater than 120 V. However, BV_{CBO} is usually determined by avalanche in the curved region of the collector, which means that a lower value is more correct for this equation. From Fig. 5-10b, for the case of a base junction depth of 3 μm (which also gives rise to a junction whose periphery has a radius of curvature of 3 μm), a $BV_{CBO}(\text{plane})$ of 120 V is reduced to ~ 90 V. An impurity concentration in the collector of approximately 3×10^{15} atoms/cm³ (equivalent to a resistivity of $\sim 1 \Omega\text{-cm}$) is needed to obtain a BV_{CBO} of 90 V.

The thickness of the depletion region must also be large enough to accommodate the depletion layer associated with the collector-base junction. At 36 V, it is calculated (using Eq. 5-17, chap. 5) that the depletion thickness is approximately 6 μm . Since Table 7-2 indicates that the buried layer diffuses upward approximately 2.5 μm during subsequent processing, and the base diffusion is approximately 3 μm deep, a total epitaxial-layer thickness of 12 μm is required for a 36-V circuit. For circuits with lower operating voltages, thinner and more heavily doped epitaxial layers can be used, because this reduces R_C .

**Fig. 7-20** The impurity distribution in an SBC bipolar transistor.

If thinner layers are used, BV_{CEO} will be limited by reachthrough of the base region instead of by avalanche multiplication. That is, the depletion layer in the lightly doped collector will hit the n^+ buried layer at voltages less than that of BV_{CEO} . At this point, further depletion-layer expansion will occur in the base, and any increase in the collector-base voltage will quickly cause breakdown to occur (see the second set of curves in Fig. 7-10b).²⁴ For example, if the epi layer is $8.5\ \mu\text{m}$ thick in the device just described, the value of $x_{BL} - x_{BC}$ will be $3\ \mu\text{m}$, and BV_{CBO} thus will be limited to a value of 30 V.

7.5.1.1 Limitations of Junction-Isolated SBC Transistors for VLSI Circuits. These transistors have several limitations. First, their packing density is relatively low since so much silicon area is taken up by inactive isolation regions. (Although the window defining the isolation diffusion may be a minimum size at the surface, the total width of the isolation region there is determined by lateral diffusion. For example, if the epitaxial layer is $12\ \mu\text{m}$ thick and the minimum feature size is $3\ \mu\text{m}$, the isolation region will approach $24\ \mu\text{m}$ in width (assuming that lateral diffusion is approximately 80% of the vertical diffusion). Thus, the width of the isolation region ends up about twice the thickness of the epitaxial layer in order for the acceptor dopants to penetrate through this layer. As a result, transistors fabricated with the SBC process are not small enough to meet the packing density requirements of VLSI circuits.

Second, these large isolation-region areas give rise to large parasitic capacitances, which degrade the speed performance.

Third, lateral *pnp* transistor action with minority-carrier collection in the isolation-diffusion region can occur, possibly leading to latchup. This parasitic *pnp* transistor is normally cut off, because the collector-substrate junction and the collector-base junctions are usually reverse-biased (i.e., the collector-substrate junction is kept at reverse bias by having the *p*-doped substrate connected to the most negative dc voltage in the electronic circuit). If the *nnp* transistor saturates, however, the collector-base junction becomes forward-biased, and the parasitic lateral *pnp* enters its active region. The resultant conducting transistor diverts current from the collector circuit to the substrate; under some circumstances latchup can even occur. All three of these limitations can be reduced by using oxide isolation instead of junction isolation, as is being done in high-performance, 5-V digital bipolar ICs.

Fourth, since the base region is formed by a single diffusion, the parasitic series base resistance will often be high enough to limit high-performance operation. A separate extrinsic base-formation process is used in advanced bipolar structures to overcome this problem.

7.5.2 Standard-Buried-Collector (SBC) Process Flow

7.5.2.1 Starting Material. The starting material is a *p*-type, lightly doped wafer ($\sim 10^{15}\ \text{cm}^{-3}$). The substrate doping is light enough to minimize the parasitic collector-to-substrate depletion-layer capacitance, but heavy enough to prevent it from being

changed to *n*-type during subsequent processing. In the early days of IC processing (111) wafers were used, but (100) is now the standard.

7.5.2.2 Buried-Layer Formation. Heavily doped n^+ regions called *buried layers* or *subcollectors* are the first features to be formed. These regions are used to reduce R_C by providing a low-resistance path from the collector contact to the active portion of the transistor.

The process is begun by growing an oxide ($0.5\text{--}1.0\ \mu\text{m}$ thick) on the wafer. *Mask #1* and an etch step are used to open windows in the oxide wherever the buried layer is to be formed (Fig. 7-21a). Since this is the first pattern on the wafer, no alignment is necessary. Instead of phosphorus, either arsenic or antimony is selectively introduced into the silicon (these impurities are used because each has high solid solubility and a small diffusion coefficient). The predeposition step can be performed using either chemical diffusion (in which case the surface oxide acts as a mask to prevent the *n*-dopant from entering the substrate, except where windows exist), or ion implantation ($\sim 30\ \text{keV}$, $\sim 10^{15}\ \text{atoms}/\text{cm}^2$; in this case the resist may remain on the wafer as a mask layer during the implant). Even if implantation is used, a thermal step is also employed to activate the implanted impurities and anneal out the implant damage.

In either case, the oxide that was initially grown on the wafer is stripped following the introduction of the dopants into the substrate and prior to the drive-in step. The anneal/drive-in step is then conducted in an oxidizing ambient, which simultaneously

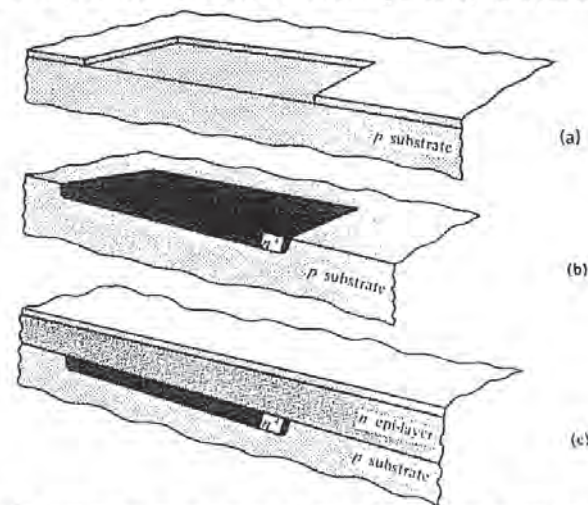


Fig. 7-21 (a) Oxide windows are opened for n^+ buried layer deposition. (b) Oxide layer is removed in preparation for epitaxial growth. (c) *n*-type epitaxial layer and oxide layer are grown. Copyright, 1982, Prentice-Hall. Reprinted with permission from reference [127].

causes a new oxide to be grown on the wafer surface. The oxide grows faster over the heavily doped n -regions than over the more lightly doped p -regions. Since the oxide consumes silicon to a depth of $0.44t_{ox}$ (where t_{ox} is the grown oxide thickness), when this new oxide is etched away, 100-200-nm-high steps exist on the bare silicon wafer at the edges of the n^+ regions (Fig. 7-21b). These silicon steps propagate through the epitaxial layer and thus become the alignment marks that allow subsequent mask levels to be aligned with the buried layer.

Although antimony has lower solid-solubility and requires higher anneal/drive temperatures (1250°C as compared with ~1000°C for arsenic), it has historically been more popular as a buried-layer dopant because it exhibits less autodoping during epitaxy and subsequent thermal cycles. On the other hand, since As atoms are closer to the size of Si atoms, As provides a better fit to the silicon lattice. This results in less strain and, hence, in fewer defects in the subsequently grown epi film. In addition, it has been found that arsenic autodoping can be decreased if the epi layer is deposited at reduced pressures.²⁵ The recent commercial introduction of reduced-pressure epitaxial reactors has renewed interest in the use of As as a buried layer dopant.

The sheet resistance of the buried layer should be as low as possible in order to reduce R_C . After the dopant is introduced into the substrate, it is driven in; this spreads out the doping profile so lowers the impurity concentration. One benefit of a lower doping concentration is higher carrier mobility, which results in lower sheet-resistance values. Furthermore, if the surface-doping concentration is too high, excessive outdiffusion into the more lightly doped collector will occur during epi growth. This can also cause defects in the epitaxial layer. Following the drive-in step, the wafer oxide is stripped.

To ensure a clean, defect-free surface, some silicon is removed from the surface during the HCl etch step in the subsequent epitaxial growth step. The loss of this highly doped layer increases the buried-layer sheet resistance and so must be taken into account. (Typical buried-layer sheet-resistance values are between 15 and 50 $\Omega/\text{sq.}$)

7.5.2.3 Epitaxial Growth. Following the stripping of the oxide, a lightly doped n -type epitaxial layer (10^{15} - 10^{16} atoms/ cm^3) is deposited. This layer serves to "bury" the heavily doped subcollector regions (Fig. 7-21c). Because of its small diffusivity, arsenic is most often used as the epi-layer dopant. The minimum thickness and maximum doping concentration of this layer are determined by the avalanche breakdown and reachthrough limitations on the value of BV_{CEO} (as was discussed in section 7.2.4).

The lighter the epitaxy doping, the smaller the collector-base capacitance of the device (which is an important parameter limiting the high-frequency performance). The minimum doping concentration, however, is generally limited by one of two factors:

1. If the transistor is to be operated at high collector currents, the conductivity of lightly doped n -epitaxy layer becomes modulated by the collector current. This causes the base-collector junction to be pushed out into the lightly doped epitaxial layer (Kirk effect; see section 7.2.3), resulting in gain degradation as well as in a reduction in the high-frequency performance. To avoid device performance degradation due to the Kirk effect, the doping in the collector should be greater

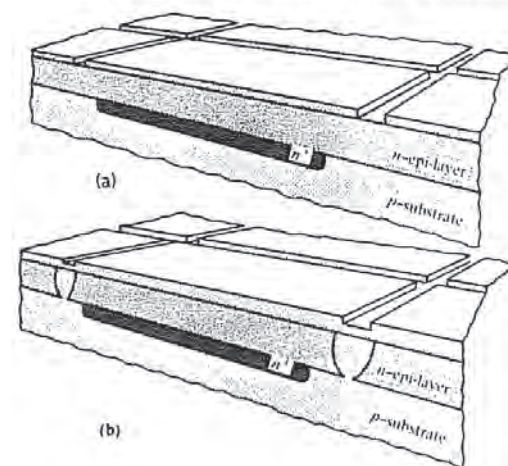


Fig. 7-22 (a) Oxide windows are opened for p -type isolation diffusion. (b) Oxide layer is regrown over p -type isolation regions. Copyright, 1982, Prentice-Hall. Reprinted with permission from reference [127].

than J/qv_1 , where J is the maximum expected collector-current density in the device and v_1 is the saturation velocity ($\sim 10^7$ cm/s).

2. Even if the Kirk effect is not important, too light a doping concentration can be difficult to control in the epitaxy growth process, due to autodoping by the buried layer.

The epi layer in early SBC transistors was deposited by means of the hydrogen reduction of SiCl_4 . However, because this process must be carried out at $\sim 1200^\circ\text{C}$ in order for sufficiently rapid growth rates to be achieved, it has been replaced in most production processes by the thermal decomposition of dichlorosilane, SiH_2Cl_2 (which yields faster growth rates at $\sim 1050^\circ\text{C}$; see Vol. 1, chap. 5). This process has the added benefit of having chlorine present during the epitaxial film growth.

The thermal decomposition of silane (900 - 1000°C) is often used in high-speed bipolar devices. In such devices, the epi layer thickness is very small (1.0 - 1.5 μm) and the processing temperature is kept low so that the epi/substrate junction will be as sharp as possible. However, there are disadvantages to using silane. First, the presence of impurities in the gas phase (even at relatively small concentrations) can cause epi defects. Second, the low stability of silane makes it susceptible to gas-phase nucleation of particles (which can then "rain down" and become incorporated in the growing film).

7.5.2.4 Formation of Isolation Regions. Following epitaxial growth, a new masking oxide for subsequent diffusions is grown on the top surface of the

epitaxial layer. *Mask #2* is used to form the patterns of the isolation regions (Fig. 7-22a). After windows have been etched in the oxide, a boron predeposition and drive-in step is performed. (The boron concentration at the surface is the solid solubility of boron in silicon at the predeposition temperature; e.g., $4 \times 10^{20} \text{ cm}^{-3}$ at 1200°C .) The purpose of this step is to isolate the collectors of the transistors from one other with reverse-biased pn junctions. Thus, the acceptor concentration in the newly diffused regions must be higher than the donor concentration in the epi layer, and the junction depth must at least equal the epi thickness in order for complete isolation to be achieved. In fact, it is customary to overdrive the isolation diffusion to prevent the depletion region from extending underneath the isolation diffusion.

The result is the formation of n -type islands that are completely surrounded by p -type material. Because thick epi layers (e.g., $>10 \mu\text{m}$) are used in analog SBC transistors, the diffusion requires several hours at temperatures of 1200°C . The isolation-diffused layer has a sheet resistance of from 20 to $40 \Omega/\text{sq}$. (Additional details regarding junction-isolation-structure fabrication are described in chap. 2, section 2.1.1.1.)

Following diffusion, a low-temperature-oxidation and etch step is used to remove the boron-rich layer from the surface of the isolation regions, and another oxidation step is used to grow a new oxide layer over the isolation-diffusion openings (Fig. 7-22b).

As noted in chapter 2, the n^+ buried layer regions are usually not allowed to intersect the p -isolation diffusions, as this would decrease the breakdown voltage of the collector-substrate junction and increase its junction capacitance. Thus, a layout design rule specifies the minimum spacing between the n^+ region and the isolation diffusion. (In fact, the minimum buried-layer-to-buried-layer separation is determined primarily by the buried-layer-to-isolation spacing.)

In addition, successful fabrication of the isolation structures can be confirmed by stripping the oxide over the isolation diffusion and probing two adjacent structures. If complete isolation has been achieved, no current will flow unless the applied voltage exceeds the breakdown voltage of the isolation junction. If even a small gap exists at the bottom of the isolation, leakage current will occur. Poor isolation results in catastrophic device failure, since all of the transistor collector regions are shorted together.

As a final note, the windows opened by the isolation mask are generally reopened during the base-diffusion step in order to increase the p -impurity concentration at the surface. This is done because the surface can become depleted of boron as a result of impurity redistribution and dopant segregation during the intervening oxidations.

7.5.2.5 Deep-Collector Contact Formation (Optional). Even with the use of buried layers, in some applications the value of R_C is still too high. To reduce R_C even further, a deep n^+ contact diffusion (also referred to as a *plug*, or *sinker*) is used. The diffusion once again penetrates the epitaxial layer to contact the underlying buried layers (Fig. 7-23). Such deep-collector-contact diffusions are particularly useful for reducing the ON -state output voltage of saturating-type digital logic circuits.

In junction-isolated SBC devices, this contact is formed by means of a diffusion process, whereas in oxide-isolated bipolar devices – which use thinner epitaxial layers –

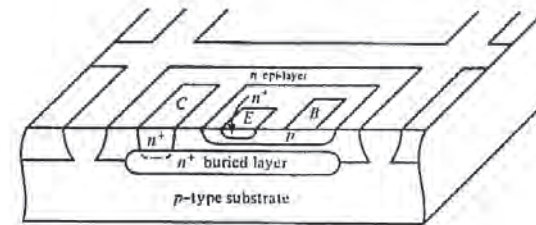


Fig. 7-23 Deep n^+ collector diffusion reduces R_C . Copyright, 1982, Prentice-Hall. Reprinted with permission from reference [127].

ion implantation (followed by diffusion) is used. Phosphorus is chosen for this application since it is a faster-diffusing impurity than arsenic. In any case, an additional mask must be employed to allow the dopant to be selectively introduced only into the collector regions.

Besides the additional masking step, a second penalty for using this approach is that a relatively large collector-to-base spacing must be maintained to avoid degradation of BVC_{BO} , since significant lateral diffusion of the deep n^+ will occur. In addition, in analog bipolar SBC transistors this process is usually performed before the base-formation step, since the diffusion must penetrate a thick epi layer. In digital bipolar SBC devices, however, the step may be carried out after the base-predeposition step. The deep-collector-contact and base-drive-in steps may then be carried out simultaneously.

7.5.2.6 Base-Region Formation. Formation of the base region, a critical process in bipolar transistor fabrication, begins with the use of the base mask (*Mask #3*) to pattern windows in the oxide where such regions are to be created (Fig. 7-24a). The base must be aligned so that the collector-base and collector-substrate depletion regions do not merge, following diffusion at the surface. The minimum allowable spacing between the isolation regions and the base region can be determined based on a knowledge of the applied voltages and the epitaxial-layer doping concentration. Additional space must be allowed to accommodate the alignment tolerances. In some cases, the base diffusion is also used to create diffused resistors for the circuit; the values of these resistors depend not only on the diffusion conditions, but also on the dimension of the opening made during etching.

The β of the transistor is heavily dependent on the base width and the doping profile. Thus, the base-junction depth and base-doping profile must be not only appropriately selected but also tightly controlled during production. As described earlier, the base region is made thinner and less heavily doped than the emitter. Since β (which is typically set to around 100), is inversely proportional to GN_b , the specified value of β sets a constraint on the *maximum* base doping allowed.

The *minimum* doping level in the base is set by a separate set of constraints. If the level is too low, the transistor will not be able to withstand the reverse-bias voltage

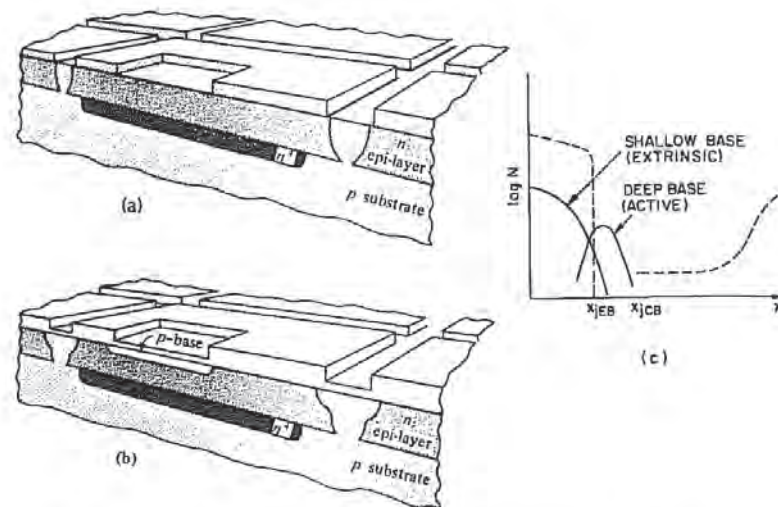


Fig. 7-24 (a) Oxide windows are opened for p-type base formation. (b) Base diffusion is completed and oxide is regrown. Copyright, 1982, Prentice-Hall. Reprinted with permission from ref. [127]. (c) Doping profile showing results of double-base implant.

applied across the C-B junction, and punchthrough will ensue. Also, as the base charge is reduced the transistor exhibits smaller values of V_A , which may be undesirable in some circuit applications. Furthermore, the parasitic series resistance in the base region becomes a problem at low base-doping concentrations. Finally, if the concentration at the surface is less than $\sim 5 \times 10^{16} \text{ cm}^{-3}$, poor metal-semiconductor contacts to the base region can occur. Typical values used for GN_B are between 10^{12} and $10^{13} \text{ atoms/cm}^2$.

The choice of base width, W_B , is made based on similar considerations. That is, small W_B values also produce shorter minority-carrier diffusion times across the base, improving performance. Given the constraints on GN_B , the value of W_B is minimized to improve the base-transit time. To allow for such narrower base widths, however, the doping level must be increased to prevent punchthrough. In high-performance bipolar transistors, the base-doping profile is graded so that a higher doping concentration will exist at the emitter-base junction, resulting in a built-in electric field that aids the transit of minority carriers.

The base may be formed through diffusion (i.e., a boron predeposition and drive-in step) or ion implantation ($\sim 10^{12}$ boron atoms/ cm^2) and diffusion. At the end of the process, the depth of the base is $\sim 3 \mu\text{m}$ in analog SBC transistors and $\sim 1.0 \mu\text{m}$ in digital SBC devices. The corresponding sheet-resistance values are 100 and 200 Ω/sq .

The drive-in step in either case is done in an oxidizing ambient so that an oxide is grown over the base region for subsequent masking (Fig. 7-24b).

One reason for performing the drive-in step is to lower the impurity concentration at the wafer surface. Avalanche breakdown of the emitter-base junction will tend to occur at the device surface, because the impurity concentration is highest there. Since such breakdown is dependent on the doping concentration of the more lightly doped side of a pn junction, the drive-in step reduces the value of the surface-doping-concentration that exists after the predeposition step. Higher emitter-base breakdown voltages can thus be achieved.

In most advanced SBC processes, a separate procedure (involving an additional masking step and a selective-doping step) is used to create an *extrinsic* base region with higher doping than that in the intrinsic base region. This provides a lower value of R_{B1} (with typical sheet resistances reduced to 100-200 Ω/sq). However, if the extrinsic base is too heavily doped, a low value of BV_{EBO} will result, and an excessively high C_{EB} value will be produced.

A double base-implant procedure is typically used to create the intrinsic and extrinsic base regions separately.²⁰ A higher-energy, lower-dose (deep-base) implant is used to establish the intrinsic base properties (Fig. 7-24c),²¹ while a lower-energy, higher-dose (shallow-base) implant is used to produce the extrinsic-base characteristics. An additional mask is used to selectively implant the additional dopants in the extrinsic base regions.

7.5.2.7 Emitter-Region Formation. The formation of the emitter region is the last step in the fabrication of the transistor. The emitter is heavily doped to obtain high current gain and to minimize R_E (e.g., for an emitter with a Gummel number of $\sim 10^{16}$ atoms/ cm^2 and an E-B junction depth of $0.5 \mu\text{m}$, the concentration is $\sim 2 \times 10^{20}/\text{cm}^3$). In addition, as the width of the base is decreased to improve performance, the junction depth of the emitter must be decreased to maintain base-width control and reproducibility. For example, in analog-SBC transistors in which the base width is on the order of $0.5 \mu\text{m}$, this depth is around $2.5 \mu\text{m}$, while in junction-isolated digital SBC transistors (which have a base width of $\sim 0.3 \mu\text{m}$) it is less than $1 \mu\text{m}$. (In advanced digital oxide-isolated, self-aligned SBC transistors, in which the base width is less than $0.1 \mu\text{m}$, the emitter depths are smaller than $0.1 \mu\text{m}$.) Control of the E-B junction depth is obviously a key step in determining bipolar-transistor behavior. Minor process variations will produce significant changes in β .

The emitter mask (*Mask #4*) is used to open the emitter-region windows in the wafer-surface oxide. The opening must lie wholly within the base region (7-25a). A shallow, high-concentration diffusion (or shallow, high-dose implant) of phosphorus or arsenic is used form the emitter structure. The drive-in that follows the predeposition (or implant) step is done in an oxidizing ambient so that oxide will cover the entire wafer. Since the thermal-oxidation rates are enhanced over heavily doped silicon regions, this oxidation cycle can be relatively short. (Note that the collector contact regions are also opened and simultaneously doped during the emitter formation step, allowing low-resistance, ohmic contacts to be made to the collector.)

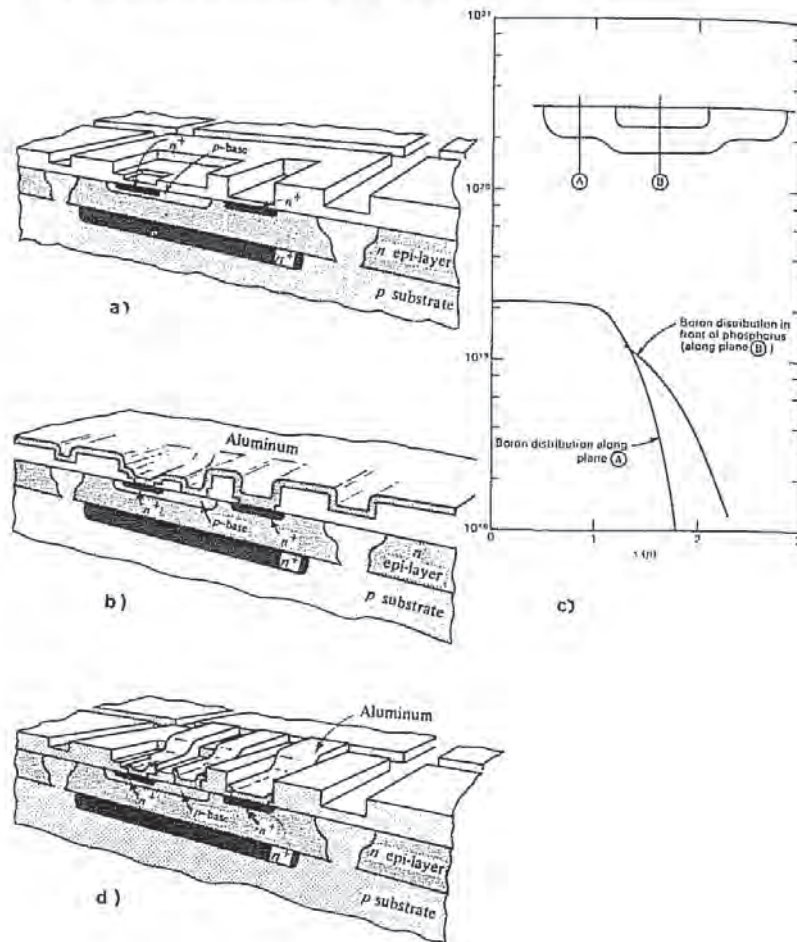


Fig. 7-25 (a) Oxide windows are opened for n^+ emitter formation. (b) n^+ emitter diffusion is completed and oxide layer is regrown. (c) The emitter push effect; anomalous diffusion due to a second high concentration phosphorus diffusion.¹²⁸ (d) Completed npn transistor; aluminum is removed as needed to complete the circuit. Copyright, 1982, Prentice-Hall. Reprinted with permission from ref. [127].

Phosphorus was used as the dopant in early bipolar transistor emitters, since the diffusion coefficients of boron and phosphorus have about the same value. Therefore, as the phosphorus is diffused, the base-collector junction should move downward at about the same rate. However, control of β is complicated by the *emitter push effect* which occurs when phosphorus is diffused at high concentrations. Because this effect enhances the diffusion of boron in the intrinsic region beneath the emitter, the C-B junction is deeper in the intrinsic base region than in the extrinsic base regions (Fig. 7-25b; see also Vol. 1, chap. 8). As a result, the base width becomes wider than it would be if the emitter push effect played no role.

Since they allow shallow E-B junctions to be fabricated more easily, arsenic emitters are therefore now more commonly used. In addition, the concentration-dependent behavior of arsenic diffusion in silicon allows for more-abrupt doping profiles than can be obtained with phosphorus. If ion implantation is used, a low-energy, high arsenic dose ($\sim 10^{16} \text{ cm}^{-2}$) is utilized.

The active area of the transistor (i.e., the region directly beneath the emitter) is a very small fraction of the total device area in SBC transistors (roughly 5%, as can be estimated from Fig. 7-26a). The rest of the area is needed to make contacts, support depletion layers, and provide isolation. In fact, the isolation region takes up the majority of the device area. Thus, its minimization represents an important issue in developing high-performance devices, not only for density improvement but also for junction-capacitance reduction as well. This limitation of junction-isolated transistors has led to the development of oxide-isolation and trench-isolation structures, as will be described in later sections.

7.5.2.8 Contact and Interconnect-Layer Formation. Part of the emitter drive-in step is done in an oxidizing ambient, so that an oxide will be grown over the exposed emitter and collector contact regions. *Mask #5* (the *contact mask*) is then used to pattern the areas of the wafer surface where holes will be cut in the oxide to allow for electrical connection of the device. (Contact openings are also made to the base, the collector, and the isolation diffusion with the contact mask.)

Since the emitter contact window must lie wholly within the emitter region (Fig. 7-25c), the emitter area must extend at least one alignment tolerance beyond the contact edge in all directions. The size of the contact thus controls the minimum emitter size. For this reason, contacts are usually fabricated at the smallest dimension that the masking process can reproducibly create.

The actual contact structure in SBC transistors has most commonly been an Al-Ti:W-PtSi-Si contact. (The details of the fabrication such contact structures are presented in chap. 3, section 3.6.1.)

A bilayer film about $1 \mu\text{m}$ thick, consisting of Ti:W and Al (or Al alloy), has been the most commonly used interconnect layer. This film is generally sputter deposited, then patterned using *Mask #6*. The metal mask is designed with sufficient overlap that the metal covers the entire contact window (7-25d). Following etching of the interconnect film, a passivation layer is deposited (see chap. 4, section 4.8), and an

anneal step is performed (see chap. 3, section 3.4.2.7). Opening of the pads is accomplished using Mask #7 and an etch step.

The basic SBC process is thus seen to be more lengthy and complex than the basic five-mask NMOS process described in chapter 5 (which also implies higher production costs per wafer). Furthermore, the use of a deep collector contact and a separate process for forming a more heavily doped extrinsic region would involve the use of two additional masks; if a double-level-metal process were implemented, yet another two masks would be required.

7.5.2.9 Washed Emitters. The washed-emitter contact-formation process allows the device area to be scaled down without a scaling of the design rules. While conventional contact masking and etching steps are used to open the base and collector contacts in this approach, the emitter areas remain covered by resist during the contact-opening step.

The resist is then stripped and the wafer is dipped in a dilute HF solution for a short time in a process known as *washing*. This washing step removes the oxide that has formed over the emitter during the emitter drive-in step, without significantly thinning the other oxides on the wafer surface. (Since the oxide grown over the heavily doped emitter regions contains a high concentration of phosphorus, it etches much more rapidly than normal in an HF-containing etchant.)

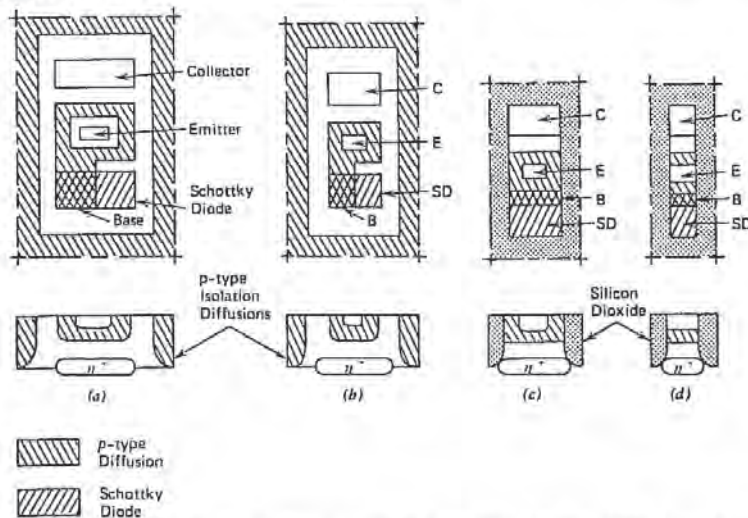


Fig. 7-26 Composite layout and cross-sections of Schottky-clamped bipolar transistors. (a) Conventional junction-isolated transistor. (b) Washed-emitter junction-isolated transistor. (c) Oxide-isolated washed-emitter transistor. (d) Walled-emitter oxide-isolated transistor.

The entire emitter-diffusion area is thus available as a contact window for the subsequently deposited metal (compare Figs. 7-26a and 26b). Since no extra area is needed for possible misalignment of the contact hole to the emitter-diffusion pattern, the emitter dimensions can be made as small as the minimum lithographic line width.

The risk of using washed emitters is that the metallization may short out the emitter-base junction at the periphery of the window. Since the emitter dopant diffuses laterally under the thicker oxide that exists over the base during emitter formation, the distance of this diffusion must be great enough that the E-B junction at the surface is not exposed by the washing step.

7.5.2.10 Schottky Contacts. A metallization technique involving the use of Schottky-barrier-diode clamps to the collector region can be used to keep bipolar transistors out of saturation. If the output voltage of a bipolar transistor without a Schottky clamp becomes small enough, the C-B junction becomes strongly forward-biased (i.e., the transistor enters the deep-saturation region of operation). As a result, a large number of minority carriers are injected into the base from the collector. Removal of these carriers when the device must be brought out of deep saturation is time consuming, and thus slows down the switching speed of the transistor.

To reduce the degree of collector-base-junction forward-biasing, a Schottky-barrier diode to the collector can be formed (as shown in Fig. 7-27a). Schottky diodes turn on at a smaller forward-bias voltage (e.g., -0.3 V for a PtSi Schottky diode to n -type Si) than do pn junction diodes (-0.7 V). Since the Schottky and C-B pn diodes are in

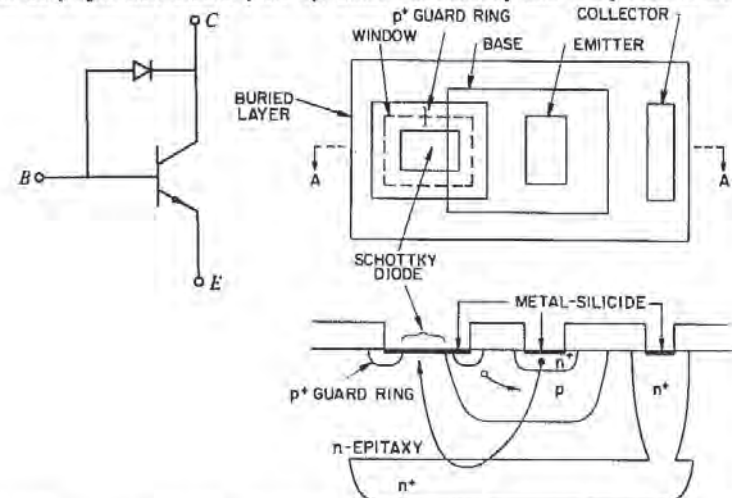


Fig. 7-27 Schematic top and side view of Schottky-clamped transistor (with p^+ guard ring) to prevent saturation. Copyright, 1983, Bell Telephone Laboratories, Inc. Reprinted with permission.

parallel, the voltage across the C-B junction is held (*clamped*) at the turn-on voltage of the Schottky diode, and hence cannot become forward-biased enough to conduct significant current.

Figure 7-27b shows the layout of a junction-isolated SBC bipolar transistor that includes a Schottky diode. The PtSi layer simultaneously contacts the *n*-epitaxy and the *p*-base regions. Hence it forms an ohmic contact to the base and a Schottky contact to the collector. A p^+ guard ring can also be used around the periphery of the Schottky diode to raise its reverse-breakdown voltage, as described in chapter 3. MoSi₂ Schottky diodes for bipolar LSIs have also been described.^{46,48}

7.6 OXIDE-ISOLATED BIPOLAR TRANSISTORS

As discussed in section 7.5.1.1, the large size of junction-isolated SBC transistors makes them unable to meet the functional density requirements of VLSI applications, and also causes them to possess high parasitic-capacitance values. While digital transistors exhibit these limitations to a lesser degree than do their analog counterparts, the drawbacks of junction isolation still prevent digital SBC transistors from being suitable for VLSI ICs. By replacing junction isolation with oxide isolation, however, it becomes possible to shrink the size and propagation-delay times of bipolar transistors so that they can be used to implement VLSI ICs.

The evolution of oxide-isolated bipolar transistors has occurred in several stages, each of which resulted in smaller device sizes and lower parasitic capacitances. In addition, the use of oxide isolation has reduced the current gain of the parasitic *pnp* transistor, decreasing the likelihood that the parasitic *pnpn* structure will latch up (a potential problem inherent in bipolar ICs).

Details of the fabrication of LOCOS-based, fully recessed field-oxide-isolation structures are presented in chapter 2, sections 2.3 and 2.5. A variety of similar processes for bipolar ICs, each with its own acronym, have been reported (e.g., Isoplanar, OXIS, Planox, OXIM, and ISAC). In early oxide-isolated bipolar structures, the field oxides exhibited relatively large bird's beak and bird's head features. More advanced processes have substantially reduced these undesired effects (see chap. 2). Figure 7-28 illustrates a process sequence for fabrication of high-performance, oxide-isolated bipolar transistors, and Fig. 7-29 shows a three-dimensional view of a completed device. With the exception of the oxide-isolation structure fabrication, most of the details of this process are identical to those of the junction-isolated SBC process described in section 7.4.2.

To ensure complete isolation between adjacent collectors, it is necessary that the field oxide be thick enough to penetrate the entire depth of the epitaxial layer. The oxide must also make contact to the n^+ buried layer. Because the collector is surrounded by an oxide collar, the sidewall contribution to the collector-to-substrate capacitance is essentially eliminated. (Note that while the growth of field oxides greater than $\sim 1.5 \mu\text{m}$ is impractical, if thicker epitaxial layers need to be employed, a combination of oxide and junction isolation may be a viable option. One name given to such a combination process is *recessed oxide isolation* [ROI].) The isolation oxide is also formed over a

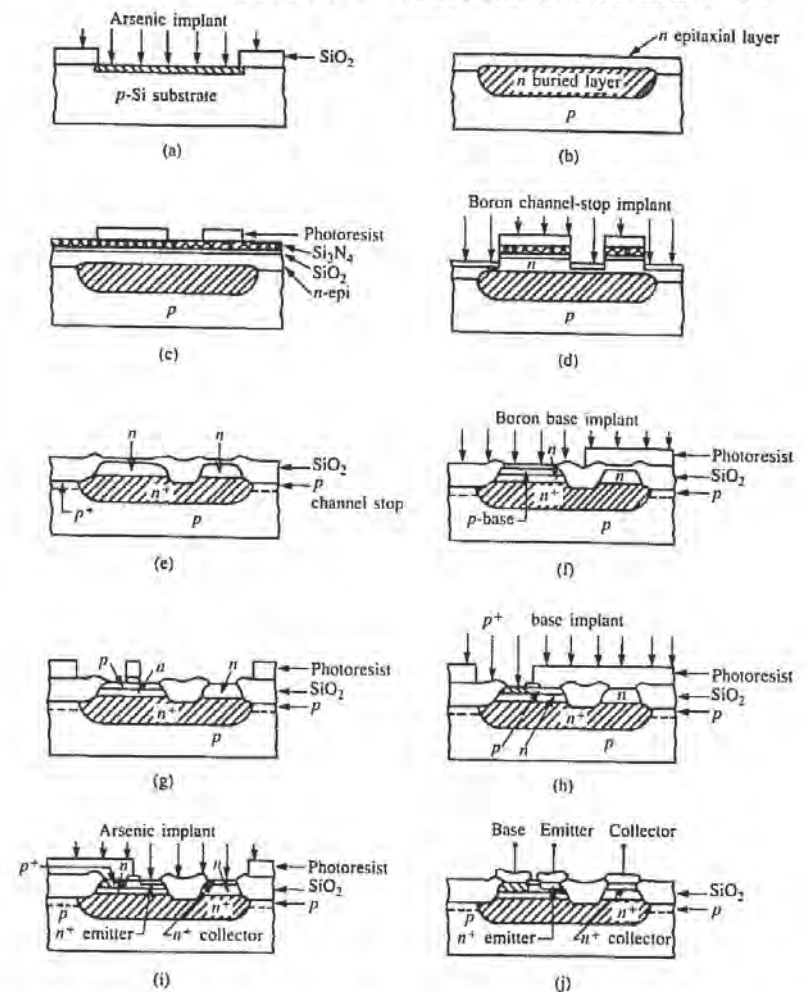


Fig. 7-28 Process sequence for a high-performance oxide-isolated bipolar transistor. (a) Buried-layer formation; (b) Epitaxial layer growth; (c) Mask for selective oxidation; (d) Boron implant prior to recessed oxide growth; (e) Selective oxidation; (f) Base mask and boron implantation; (g) Emitter, base contact, and collector contact mask; (h) p^+ base contact implantation; (i) Arsenic implantation for emitter and collector contact; (j) Structure completed with multilayer metallization. Copyright, 1985, John Wiley and Sons, Inc. Reprinted with permission from ref. [130].

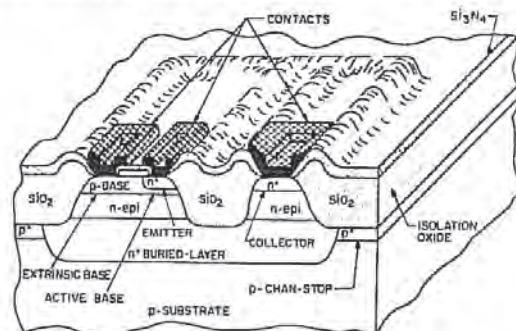


Fig. 7-29 Three-dimensional view of oxide-isolated bipolar transistor. Copyright, 1980, John Wiley and Sons, Inc. Reprinted with permission from ref. [131].

fraction of the active device area (i.e., between the collector and base contacts, to eliminate unnecessary area and capacitance between them; see Fig. 7-29).

A boron channel-stop implant is used beneath the field oxide to prevent the lightly doped substrate p -regions beneath the field oxide from inverting, as this would destroy the isolation between adjacent collectors.

In early oxide-isolated bipolar device structures, oxide surrounded the periphery of the collector as noted above. In addition, the base region butted up against the oxide on all sides, and C_{CB} was thus reduced. The emitter, however, was still wholly enclosed within the base region (a so-called *nonwalled emitter*). Nevertheless, since the collector-metal and the base-metal stripes could overlap the sidewall oxide, and the collector n^+ and the p base contacts could extend over the isolation region, it was possible to reduce the total device area to somewhat less than 50% of that needed by a junction isolated device (see Fig. 7-26c).

In the second generation (Fig. 7-26d), the size was further reduced by allowing two sides of the emitter (and in some device structures, even three sides) to abut the oxide-isolation region. Devices fabricated with this approach are referred to as *walled-emitter transistors*. The value of also C_{EB} in walled-emitter transistors is also reduced. Nonwalled emitter transistors will have smaller R_{B2} values than walled-emitter transistors using the same minimum dimension, but they will also be larger.

7.7 ADVANCED BIPOLAR-TRANSISTOR STRUCTURES FOR VLSI AND ULSI

Although CMOS has been the major driving force in the rapid progress in silicon technology in recent years, it is less well known that there has also been rapid progress in bipolar technology.²⁶ The advances here have been due to a number of process

breakthroughs, including polysilicon emitters, self-aligned bipolar structures, deep-trench isolation, and heterojunction-bipolar transistors.

7.8 ADVANCED EMITTER STRUCTURES

As pointed out in section 7.3.1, the maximum theoretical current gain in conventionally designed bipolar transistors is $\sim 10^4$. However, such high gains are not obtained in actual devices, since the heavy doping in the emitter reduces the band gap and increases the recombination rate of minority carriers. Both effects lower the emitter-injection efficiency, which also implies that attempts to increase the current gain by raising the emitter-doping levels to even higher levels would be to no avail. In addition, higher doping concentrations would lead to smaller values of BV_{EBO} and larger E-B junction capacitances. Finally, when the emitter junction depths are scaled below 200 nm (to allow control to be maintained over the base width as it is reduced), the minority-carrier diffusion length becomes larger than the emitter depth, resulting in further reduced current gain.²⁷

Two approaches to circumventing these problems have been pursued, each based on the use of alternative emitter structures: polysilicon emitters, and heterojunction bipolar transistors (HBTs).

7.8.1 Polysilicon Emitters

Besides being formed through ion implantation of dopants into the substrate, emitters can be formed through outdiffusion of dopant from a polysilicon layer in direct contact with the monocrystalline substrate. In 1979, it was reported that the current gain of transistors fabricated with such polysilicon emitters is three to seven times greater than that of transistors with conventional emitters.²⁸ Figure 7-30 gives an example of such enhanced device performance, comparing the current gains of polysilicon-emitter npn

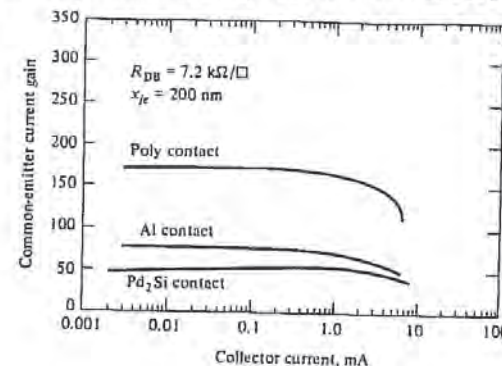
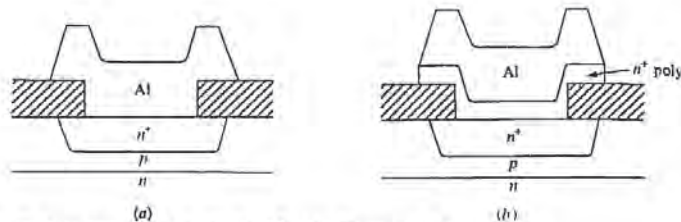


Fig. 7-30 Current gain as a function of collector current for an emitter 200-nm deep contacted either by Pd_2Si , Al, or 100 nm polysilicon.²⁸ (© 1979 IEEE).

Fig. 7-31 Emitter contacts: (a) Al, (b) n^+ polysilicon + Al.

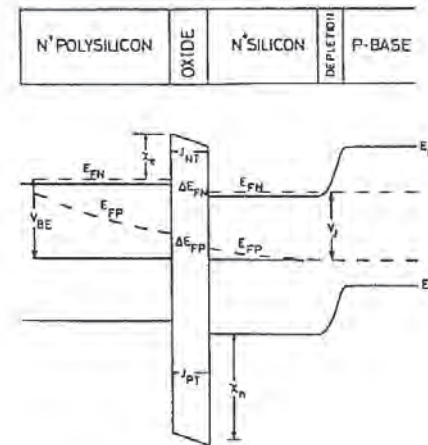
transistors with those of two types of transistors fabricated with conventional emitters. Each transistor is fabricated with a shallow (200-nm-deep) emitter and a base width of 100 nm.²⁹ (The polysilicon layer in such structures typically has a thickness of between 150 and 250 nm and is covered with an Al layer; see Fig. 7-31.) The transistors with polysilicon emitters exhibit higher gains than those made with direct metal contacts to the emitter.

In applications where higher transistor current gains do not significantly increase performance (e.g., digital circuits), the enhanced current gain can be traded off for an increase in the base doping level. This allows the base width to be decreased without the occurrence of premature punchthrough (a significant benefit, because as noted earlier, narrower base widths yield transistors with higher maximum switching speeds.) Polysilicon-emitter n pn transistors with f_T 's of 17-30 GHz and sub-50-ps ECL gate delays have been produced.³⁰

The development of polysilicon emitters has also been pursued for other important reasons:

- The lateral and vertical dimensions of the emitter can be scaled in a coordinated manner, allowing the peripheral component of the E-B junction capacitance to be maintained at a reasonable value.
- Defects in the vicinity of the emitter-base junctions of conventional ion-implanted emitters give rise to severe yield problems.³¹ These defects are associated with the residual damage caused by the emitter-formation implant (i.e., damage that remains even after post-implant annealing; see Vol. 1, chap. 9). In polysilicon emitters, this problem is avoided, since the dopants are implanted into the poly-Si and the emitter-base junction is formed by diffusion into undamaged silicon.
- The incidence of Al spiking of the emitter-base junctions is significantly decreased.

7.8.1.1 Models That Describe Polysilicon-Emitter Behavior. Three models have been proposed to explain why polysilicon-emitter contacts increase the current gain:

Fig. 7-32 Band diagram of a polysilicon emitter transistor with an interfacial oxide.³³ (© 1989 IEEE).

1. Reduced-Mobility Model. This model postulates that the current gain increases because the polysilicon film increases the length of the emitter, while the low minority-carrier mobility in the polysilicon retards the transport of the injected minority carriers. However, this early model is no longer considered correct, as it fails to explain many of the other characteristics observed in poly-Si emitters.³²

2. Interfacial-Oxide Tunneling Model. A layer of oxide can be formed on the single-crystal-silicon substrate during the processing that precedes polysilicon deposition. The tunneling model assumes that this native oxide acts as a tunneling barrier to hole injection into the polysilicon (Fig. 7-32).³³

This model predicts current gain improvements of ten or more, depending on the thickness and height of the tunneling barrier. It also predicts an increase in emitter resistance due to the transport of majority carriers through the interfacial oxide. This is consistent with the high values of emitter resistance observed in polysilicon emitters, especially in devices with deliberately grown interfacial oxides. It appears that tunneling through an interfacial oxide plays a significant role in poly-Si emitters if the oxide is more than about 10 Å thick.

3. Interface Dopant-Segregation Model. This model describes why increased current gains are observed for poly-Si emitters with oxide-free interfaces (and is also apparently applicable to poly-Si-emitters with interfacial oxides thinner than ~10 Å – e.g., the native oxides that unavoidably form even at "clean" poly-Si interfaces). The model is based on the premise that dopants in polycrystalline silicon segregate to the grain boundaries (see Vol. 1, chap. 6). Since the poly/mono interface is essentially a large continuous grain boundary, dopant from

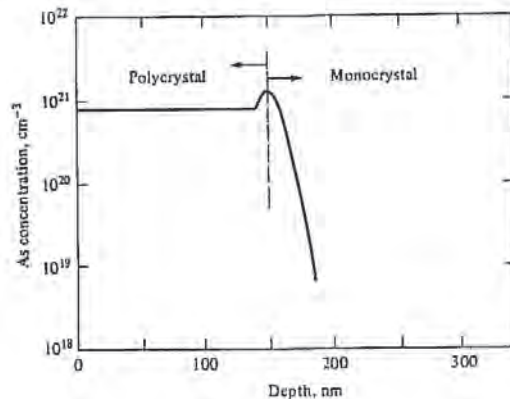


Fig. 7-33 Measured doping of arsenic concentration in a polysilicon emitter contact. (© 1979 IEEE).

the polysilicon should segregate to it. SIMS-analysis results confirm that such dopant pileup does occur (Fig. 7-33).³⁴ This gives rise to a potential-energy barrier that hinders the injection of holes into the emitter during forward bias. The blocking action of the barrier also decreases the slope of the hole concentration in the emitter, leading to a smaller hole current and an increased current gain.^{35, 36}

7.8.1.2 Process Technology for Polysilicon-Emitter Fabrication.

Either phosphorus or arsenic can be used to dope the polysilicon. Arsenic is selected for transistors in which an f_T of more than 10 GHz is required. The smaller diffusion coefficient of As in Si permits the formation of shallower emitter regions and narrower bases, while the emitter push effect that occurs when phosphorus is diffused makes control of narrow base widths more difficult in phosphorus-doped poly emitters (see Vol. 1, chap. 8). However, the latter are less sensitive to the presence of interfacial oxide layers than are those doped with arsenic. Thus, phosphorus-doped polysilicon emitters yield transistors with lower emitter resistances but wider base widths, making them suitable for bipolar processes that can tolerate *npn* transistors with less-than-optimum performance characteristics. (Polysilicon emitters doped with phosphorus typically exhibit minimum emitter interfacial contact resistances of $20 \Omega\text{-}\mu\text{m}^2$ compared to $\sim 60 \Omega\text{-}\mu\text{m}^2$ for arsenic-doped emitters.)³⁷

The presence of an intentionally grown oxide layer at the interface (typically $\sim 14 \text{ \AA}$ thick)³⁸ leads to higher current gain than if oxide-free interfaces are contacted (actually, "oxide-free" implies that only native oxides thinner than $10 \text{ \AA}$ are present). Since the intentional oxide layers are normally formed through a chemical oxidation of the silicon

in a solution of ammonium hydroxide (which is one of steps of the RCA-prefurnace wafer cleaning procedure; see Vol. 1, chap. 15), they are sometimes referred to as *RCA-clean interfacial layers*.

Unfortunately, as noted above, interface oxides also represent a significant electrical barrier for *majority carriers*; if the layers are continuous they increase the minimum emitter resistance by more than an order of magnitude.³⁹ The negative impact on device performance caused by such high polysilicon-emitter resistance is significant.⁴⁰ For small-area emitters this resistance severely limits the speed and transconductance of the device; consequently, emitter contact resistances of less than $50 \Omega\text{-}\mu\text{m}^2$ are required for high-performance bipolar transistors.

Annealing the emitter at temperatures of 1000°C can reduce the emitter resistance by breaking up the interfacial oxide (i.e., this causes the oxide layer to become discontinuous) and diffusing the emitter deeper into the substrate. However, these improvements are obtained at the costs of lower current gains and larger base width. Thus, attempts to integrate such chemically formed oxide layers into polysilicon-emitter fabrication processes have been abandoned.⁴¹

Nevertheless, it has been reported that high current gains can still be obtained if no intentional oxide is grown, but if instead the silicon substrate surface is cleaned using an HF dip and DI rinse just prior to polysilicon deposition.³⁹ A native-oxide layer still forms at the interface (with an average thickness of slightly less than $8 \text{ \AA}$) (see Fig. 7-34a).⁴² This layer is also effective in blocking minority-carrier injection into the emitter (which further reduces the base current). Annealing at temperatures of 900°C or more causes some of this oxide layer to be broken up (Fig. 7-34b), and also causes some areas of the polysilicon layer to epitaxially realign with the substrate Si.

Nevertheless, the nonuniform and variable nature of even this non-intentionally grown interfacial region has represented a source of poor reproducibility,* especially as device regions have been scaled down.⁴³ However, several processing techniques, have been studied for reliable production of poly-Si emitters that exhibit low interfacial contact resistances and that also effectively block minority-carrier transport. These techniques include the following:

- * The emitter drive-in step is carried out by means of RTP at 1050°C ³⁶ (following a careful cleaning procedure and HF dip just prior to poly deposition). This allows a shallow ($\sim 50\text{-nm}$ deep) As-doped emitter region to be formed without disruption of the continuous grain boundary at the poly-mono interface (i.e., during *furnace* drive-in steps, some of the poly is epitaxially realigned with the mono-Si, reducing the effectiveness of the barrier). The emitter resistance following this procedure is reduced to $30 \Omega\text{-}\mu\text{m}^2$. In a second approach, a $1050\text{-}1150^\circ\text{C}$, 10-sec RTP step was carried out following a 900°C furnace anneal. This also produced enhanced current gain and reduced emitter resistance.⁴⁴

- * In analog bipolar ICs, where low and reproducible values of R_E are required, the formation of emitters by means of direct implantation and metal contact to the substrate is still more widely practiced.

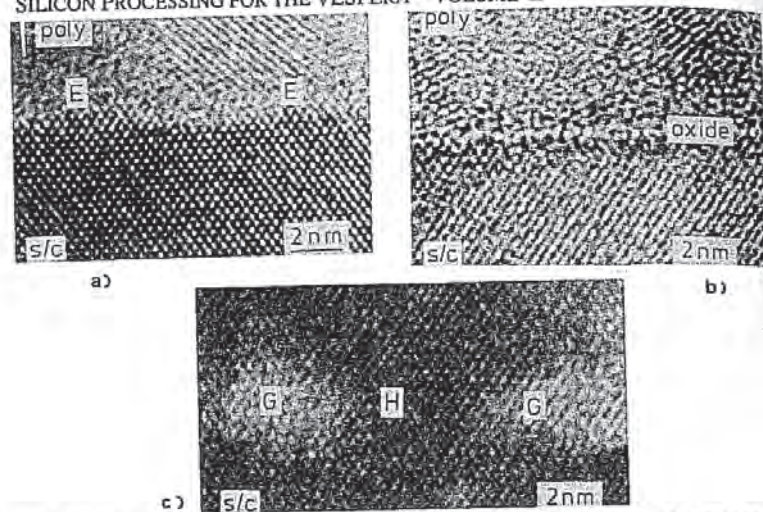


Fig. 7-34 (a) High-resolution-TEM micrograph of a Si surface given an HF etch prior to polysilicon deposition, an As implant and a drive-in of 30 mins. Interfacial oxide is not continuous. (b) Si surface given HF etch prior to polysilicon deposition, but with no heat treatment yet. A thin (~4 Å thick), continuous oxide is present. (c) Same surface, but after an interface anneal at 1000°C and full emitter processing. Interface oxide is broken up, and consists of balls to 50 Å in diameter. (after Wolstenholme et al.,⁴² copyright 1987 AIOP).

- The polysilicon deposition is performed at 275°C using a PECVD process.⁴⁵ This approach produces highly reproducible enhanced current gains at relatively low post-deposition anneal temperatures (900°C).

7.8.2 Heterojunction Bipolar Transistors (HBTs)

As the vertical dimensions of the bipolar transistor are scaled even further, it is predicted that some serious device-operational limits will be encountered, even if polysilicon emitters are used. The following scenario will illustrate this point: When the base thickness of the bipolar transistor is decreased, the doping level must be increased in order to control punchthrough and to maintain a sufficiently low value of R_{B2} . Since the doping level in the emitter must remain high in order for a sufficient emitter-injection efficiency to be maintained, CEB will increase. The current density will thus need to be increased so that the propagation delay associated with this extra capacitance can be overcome. To handle the increased current density without suffering the Kirk effect, it will be necessary to raise the doping in the collector. Consequently, even though such a device will have a higher f_T , it will have a narrower operating-current range. Furthermore, the increased doping levels will reduce the values of BV_{EBO} or BV_{CEO} to such a degree that it will not be possible to use the device in many practical applications.

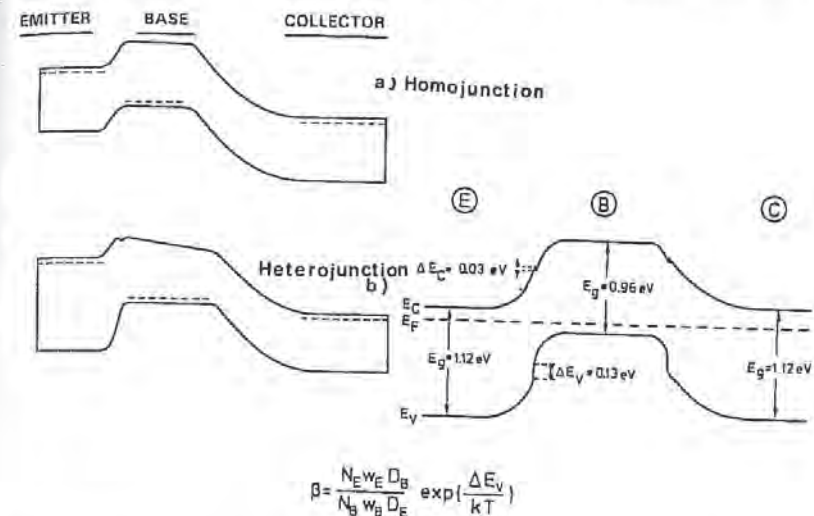


Fig. 7-35 Band diagrams of representative npn (a) homojunction and (b) heterojunction bipolar transistors.⁴⁹ (© 1989 IEEE).

One actively studied approach to overcoming the above problems is to build transistors with emitter materials whose band gaps are larger than that of the material used in the base (i.e., *heterojunction transistors*). This technique appears promising because the emitter-injection efficiency, γ , depends on the band-gap difference according to the equation⁴⁹

$$\gamma = (n_e v_e / p_b v_h) \exp(\Delta E_g / kT) \quad (7-19)$$

where n_e is the emitter electron concentration, v_e is the effective velocity of electrons injected into the base, p_b is the base hole concentration, v_h is the corresponding hole velocity injected into the emitter, and ΔE_g is the difference in the band gaps of the emitter and base. (From Fig. 7-35, it can be seen that the larger band gap in the emitter gives rise to a larger potential-energy barrier for holes going from the base to the emitter than for electrons going from the emitter to the base.) Thus, if the band gap of the emitter in silicon devices is larger than that of the base, when ΔE_g is 0.2 eV or more, $\exp(\Delta E_g / kT)$ will exceed 2000. As a result, a sufficiently high emitter-injection efficiency will be assured, no matter what doping levels exist in the base and emitter.

For example, doping levels in the base can be increased to above 10^{19} cm^{-3} , and in the emitter they can be dropped to $5 \times 10^{17} \text{ cm}^{-3}$ (Fig. 7-36a). Under such conditions, many other device characteristics are also improved (among the improvements are a lower base resistance, a higher punchthrough voltage, a

higher Early voltage, and a decreased E-B junction capacitance). However, for such devices to be successful in silicon technology, the technology must be compatible with advanced silicon bipolar processing.

Most approaches to heterojunction fabrication are based on growing an epitaxial layer on the Si substrate, such that the epi material has a different band gap than the Si, but also has a lattice constant that is at least approximately matched to the lattice constants of the Si substrate. If such deposited layers are sufficiently thin, the requirement for exact lattice-match is relaxed. That is, thin layers with an almost-matched lattice constant will deform so that they can accommodate to the lattice constant of the substrate. In this deformed condition they are either "coherently strained" or "pseudomorphic" layers. The latter are thermodynamically stable for thicknesses below a strain-dependent "critical" value. However, it is important that subsequent processing temperatures be restricted to prevent the stress of the layers from being relaxed through the formation of an array of misfit dislocations.

From the point of view of process integration, the most attractive approach to building devices with hetero-E-B junctions is to increase the band gap of the emitter (since this represents the only change from a conventional bipolar process). Several methods for implementing this approach have been attempted, including the use of the following materials as the wider band gap emitters: GaP,⁵⁰ semi-insulating polycrystalline-silicon (SIPOS, $E_g \approx 1.7$ eV);⁵¹ oxygen-doped Si-epitaxial films (OXSEF, $E_g \approx 1.7$ eV);⁵² β -SiC ($E_g \approx 2.2$ eV);⁵³ and phosphorus-doped hydrogenated microcrystalline silicon (μ c-Si, $E_g \approx 1.7$ eV).⁵⁴

Despite an intense effort to develop such processes, no entirely successful fabrication procedure has yet been demonstrated. Although numerous high-gain transistors have been reported, these devices exhibited other problems, including high-temperature requirements, high bulk and contact resistivities, high emitter resistances, or poor stability during the multilevel-interconnect fabrication steps. The most promising candidate of this group is the μ c-Si emitter. Fujioka et al. reported achieving reduced emitter resistances with such structures and built ECL gate arrays with them using a non-self-aligned emitter structure.⁵⁴ Basic gate delays of 295 ps at 6.5 mW/gate were obtained.

An alternative approach is to use a material in the base of the transistor that has a smaller band gap than that of Si. Layers of Si to which Ge has been added exhibit such reduced band gaps (Fig. 7-36b). For example, pseudomorphic epitaxial $\text{Si}_x\text{Ge}_{1-x}$ alloy layers of up to 100 nm in thickness can be formed with a Ge content of 20%, and the ΔE_g between the emitter and base in Si transistors fabricated with such base layers (i.e., $\text{Si}_x\text{Ge}_{1-x}$ -base HBTs) can be as large as 0.1 eV. An excellent, highly detailed review of the science and technology of such devices is presented in reference 55.

Pseudomorphic epitaxial $\text{Si}_x\text{Ge}_{1-x}$ layers can be deposited by means of molecular beam epitaxy (MBE) or chemical vapor deposition (CVD). The most successful CVD methods have involved *ultrahigh-vacuum CVD*⁵⁶ or *limited-reaction-processing (LRP) CVD*.⁵⁷ Note that appropriate levels of boron doping must also be incorporated into these films during the epi growth process of the base region.

It has been calculated that HBTs with narrow band-gap $\text{Si}_x\text{Ge}_{1-x}$ bases will be able to achieve very high performances (e.g., f_T 's as high as 75 GHz have been predicted for

a 20% Ge concentration in a 100-nm base, a 0.2- μm -wide Si emitter, and a collector current density of 1×10^5 A/cm²).⁵⁸ Conversely, $\text{Si}_x\text{Ge}_{1-x}$ -base HBTs offer the potential benefits of the same performance characteristics as those of conventional homojunction transistors, but with a relaxed profile (i.e., the base can be doped more heavily for a lower base-sheet resistance, or it can be made thicker to achieve lower doping levels with higher breakdown voltages).

The chief drawback of the reduced-band-gap/base approach is that it is more difficult to integrate into existing advanced bipolar processes. The metastable nature of the $\text{Si}_x\text{Ge}_{1-x}$ films places constraints on the Ge content, layer thickness, subsequent annealing cycles, and thermal oxidation. On the other hand, this method does not suffer from the other problems exhibited when wide band-gap emitter materials are used. In addition, to provide the ultrathin base profiles for future bipolar devices, the conventional ion-implantation approaches are becoming increasingly difficult to implement successfully, while the low-temperature epitaxial processes show great promise.

While $\text{Si}_x\text{Ge}_{1-x}$ -base HBT devices have demonstrated excellent performance potential

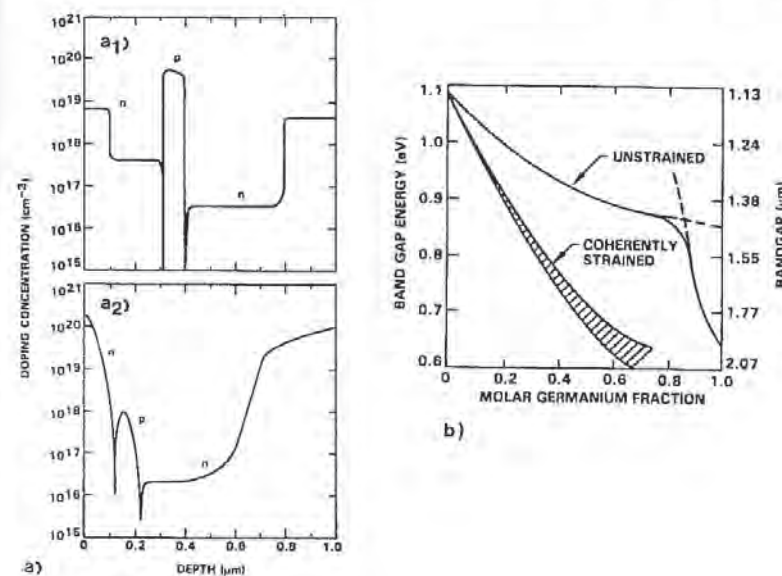


Fig. 7-36 (a) Representative doping profiles for: (a₁) npn heterojunction and; (a₂) npn homojunction bipolar transistors.⁴⁹ (© 1989 IEEE). (b) Band gap energy versus alloy composition for Si-Ge alloys.¹³² (after People and Bean,⁴² copyright 1986 AIOF).

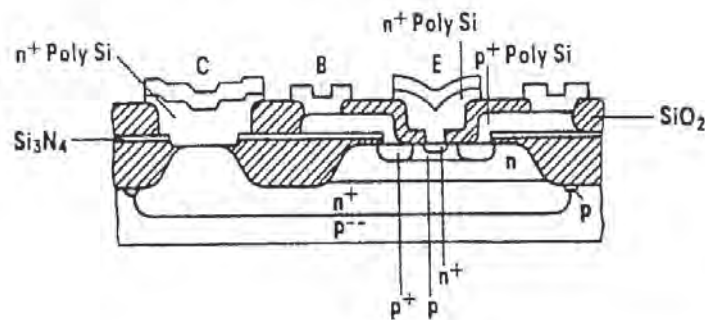


Fig. 7-37 Cross-section of a double-polysilicon self-aligned bipolar transistor.¹³³ (© 1986 IEEE).

to date, they must still pass several tests with respect to process integration, manufacturability, and long-term reliability before they can be incorporated into mainstream IC production.⁵⁵

7.9 SELF-ALIGNED BIPOLAR STRUCTURES

Although oxide isolation reduces the bipolar device size by requiring less isolation area, it does not reduce the distances between the contacts within the device. That is, relatively large inactive device areas are needed within the isolation boundaries to accommodate mask misalignment errors. (For example, the optical-stepper alignment accuracy is about $\pm 0.4 \mu\text{m}$). However, these same inactive regions not only waste valuable silicon area but also cause parasitic capacitances and resistances that degrade transistor performance. The most effective techniques developed to reduce the areas of these regions involve the use of *self-aligned* (SA) structures. The chief disadvantage of SA structures is that they increase process complexity; in addition, their smaller tolerances make them difficult to manufacture with high yield. Process-development efforts in this area will thus focus on these two issues.

7.9.1 Double-Polysilicon Self-Aligned Structures

The most widely used SA approach employs a double-polysilicon process to form an emitter that is self-aligned to the base, as shown in Fig. 7-37.^{59,60} The first layer of polysilicon (*poly 1*) is doped p^+ and serves as the base electrode, while the second (*poly 2*) is doped n^+ and makes contact to the transistor's emitter region. The emitter and the

extrinsic base regions are formed by means of dopant outdiffusion from poly 1 and poly 2, respectively.

Figure 7-38 shows the sequence of steps that form this self-aligned structure.⁵⁹ Following the growth of the field oxide and the removal of the oxide from the surface of the active regions, poly 1 is deposited and heavily p -doped with boron. This layer is then covered with a CVD oxide (Fig. 7-38a). The emitter mask is used to pattern the emitter-area regions, and a dry-etch process is used to produce openings in the CVD oxide and poly 1 (Fig. 7-38b). A thermal oxide is then grown over the etched structure, and a relatively thick oxide (on the order of $0.1\text{--}0.4 \mu\text{m}$) is grown on the vertical sidewalls of the heavily doped poly. The thickness of this oxide determines the spacing between the edges of the base and emitter contacts. The extrinsic base regions are also formed during the thermal oxide growth step, as a result of the outdiffusion of boron from the poly 1 into the substrate. Because the boron diffuses laterally as well as vertically, the extrinsic base region will be able to make contact with the intrinsic base region that is formed next (and that will be located under the emitter contact).

The parasitic series resistance of the extrinsic base region, R_{B1} , will now consist of the sum of the resistances of the poly 1 electrode, $R_{B1\text{poly}}$, and of the path between the edges of the base and emitter contact, $R_{B1\text{sub}}$. Since the latter distance is so small, $R_{B1\text{sub}}$ is negligibly small. The doping of poly 1 can be much heavier than that of the extrinsic base region in the substrate, and the value of $R_{B1\text{poly}}$ can therefore be made much smaller than R_{B1} in a non-SA transistor.

Following the oxide-growth step, the intrinsic base region is formed through the ion implantation of boron (Fig. 7-38c). This serves to self-align the intrinsic and extrinsic base regions. Following a cleaning of the contact to remove any oxide layer present, poly 2 is then deposited and implanted with As or P. A shallow emitter region is then formed through dopant outdiffusion from poly 2 (this critical step will be discussed in greater detail in section 7.8.1.2). The use of RTP for the base and emitter out-diffusion steps has also been reported to facilitate the formation of shallow E-B and C-B junctions.⁶⁰

This SA structure allows the fabrication of emitter regions smaller than the minimum lithographic dimension. When the sidewall-spacer oxide is grown, it fills the contact hole to some degree (since the thermal oxide occupies a larger volume than the original volume of polysilicon; see Vol. 1, chap. 7). Thus, a $0.8\text{-}\mu\text{m}$ -wide opening will shrink to about a $0.4\text{-}\mu\text{m}$ width if a $0.4\text{-}\mu\text{m}$ -thick sidewall oxide is grown on each side. This is another reason why SA structures can be made smaller than conventionally fabricated devices.

7.9.1.1 Limitations of Double-Polysilicon SA Structures. A major obstacle to producing small W_B values in advanced bipolar devices is the channeling tail of the intrinsic-base boron implant (see Vol. 1, chap. 9). Furthermore, the use of the normal 7° off-axis implant to reduce such channeling tails results in shadowing effects that adversely impact the device's perimeter-punchthrough and tunneling-leakage-current characteristics (Fig. 7-39).⁶² Several techniques have been reported for overcoming this drawback.

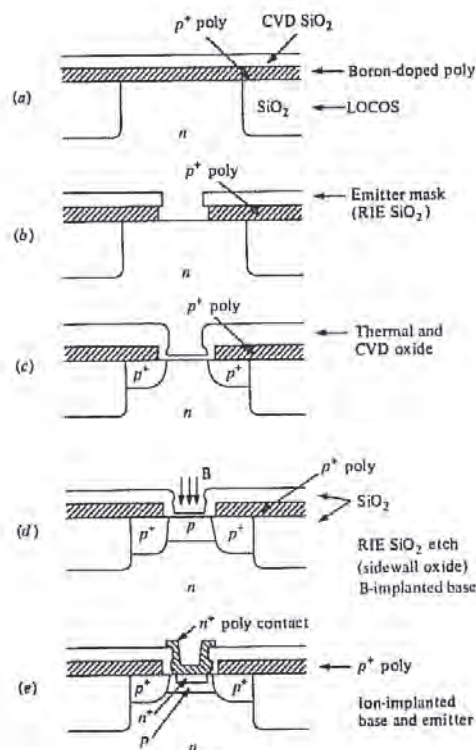


Fig. 7-38 Process sequence for fabricating double-polysilicon self-aligned *npn* transistors.⁵⁹ (© 1980 IEEE).

In one approach, the direct implantation of boron into the substrate is eliminated.⁶³ Instead, after the sidewall spacer oxide has been grown, poly 2 is deposited and implanted with boron. Following an outdiffusion step to form the intrinsic base region, an As implant into the poly 2 is performed. A second diffusion then forms the emitter region. This approach has produced *npn* transistors with base widths of 100 nm, emitter junction depths of 50 nm, and f_T 's of 16 GHz.

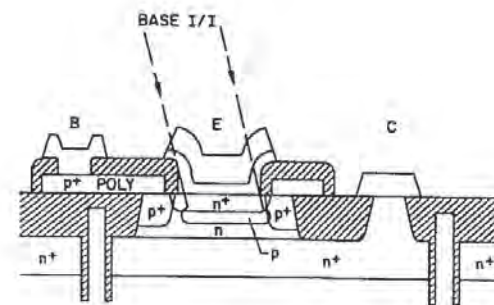


Fig. 7-39 Device cross-section of an advanced polysilicon self-aligned bipolar transistor. The base profile is offset with respect to the emitter profile due to the sidewall shadowing effect.⁶² (© 1987 IEEE).

It should be noted that both diffusions cannot be done simultaneously, since boron is nearly immobile in the presence of high arsenic concentrations.⁶⁴ Successive diffusion of both dopants, as shown in Fig. 7-40a, solves this problem and yields very narrow base widths. On the other hand, since the poly is partially recrystallized during the first boron drive-in step, the implanted As no longer diffuses as rapidly, resulting in a shallower As profile with a steeper slope.

In another approach, the thickness of the poly 1 layer is reduced so that the intrinsic base implant can penetrate it.⁶⁵ Another method also eliminates the base-implant step, forming the emitter by means of a selective epitaxial process (Fig. 7-40b).⁶⁶ This approach was extended by using a low-temperature-epi (LTE) process for forming the base, and an epitaxial-lateral-overgrowth (ELO) process for forming the extrinsic base (i.e., instead of using *p*-doped polysilicon for the latter).¹³⁶ *NPN* transistors with 60-nm-wide epitaxial-base regions and 7.5 k Ω /sq intrinsic bases sheet resistances were successfully fabricated with this method. Yet another method involves the deposition of borophosphosilicate glass (BSG) into the etched emitter window and the outdiffusing of the base dopant from the BSG to form the intrinsic base (Figs. 7-41a).⁶⁷ The BSG film is then stripped before the polysilicon-emitter film is deposited. An advanced BSG-based process was subsequently described in which the epitaxial collector region under the intrinsic base was implanted with phosphorus to suppress the Kirk effect and W was selectively deposited over the emitter polysilicon to reduce the emitter resistance (Fig. 7-41b).¹³⁷ Transistors with a 40 GHz f_T were fabricated with this approach. Other approaches are also described in section 7.8.2.

As noted earlier, in the SA structures that use polysilicon-base electrodes, the extrinsic-base-region series resistance is the sum of R_{B1sub} and R_{B1poly} . The latter is a function of the polysilicon sheet resistance. With 300-nm-thick polysilicon used, an acceptably low value of 50 Ω for R_{B1poly} has been measured.⁶⁸ However, as the polysilicon is reduced in thickness (to avoid topography problems in submicron processes), R_{B1poly} will increase. A potential solution is to replace the polysilicon with a polycide. One example of such an approach has been reported by Tashiro et al., who formed a PtSi on the polysilicon toward the end of the process.⁶⁹ Silicided base regions in single-poly SA structures will be described in section 7.8.2.

There are three other problems associated with the double-poly SA process. First, the silicon surface can be damaged during the poly 1 etch.⁷⁰ Second, the topography of the double-poly structure has severe steps, which give rise to increased complexity in the metallization and lithography processes. Third, perimeter effects play an important role in the device characteristics (analyses of such effects are presented in references 71 and 72). Increasingly complex structures have been devised to solve these problems.

7.9.1.2 Current-Gain Degradation Due to Sidewall Injection In SA Bipolar Structures.

Injection of minority carriers into the base occurs at the

Bipolar Structures. Injection of minority carriers into the base occurs at the sidewalls (or perimeter) of the emitter, as well as at the emitter bottom. If the injections from these two regions are considered separately, it can be seen that the current gain due to transistor action from the sidewall injection is much smaller than that due to the bottom injection. This is because both the base width and the doping

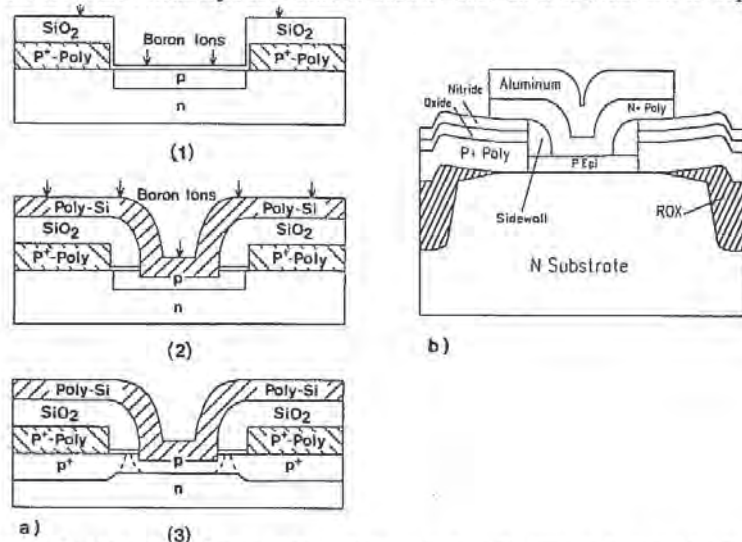


Fig. 7-40 (a) Double diffusion of the polysilicon to form the intrinsic base and the emitter regions without ion implantation into the Si substrate.⁶³ (b) Cross-section of selective-epitaxy base transistor.⁶⁶ (© 1988 IEEE).

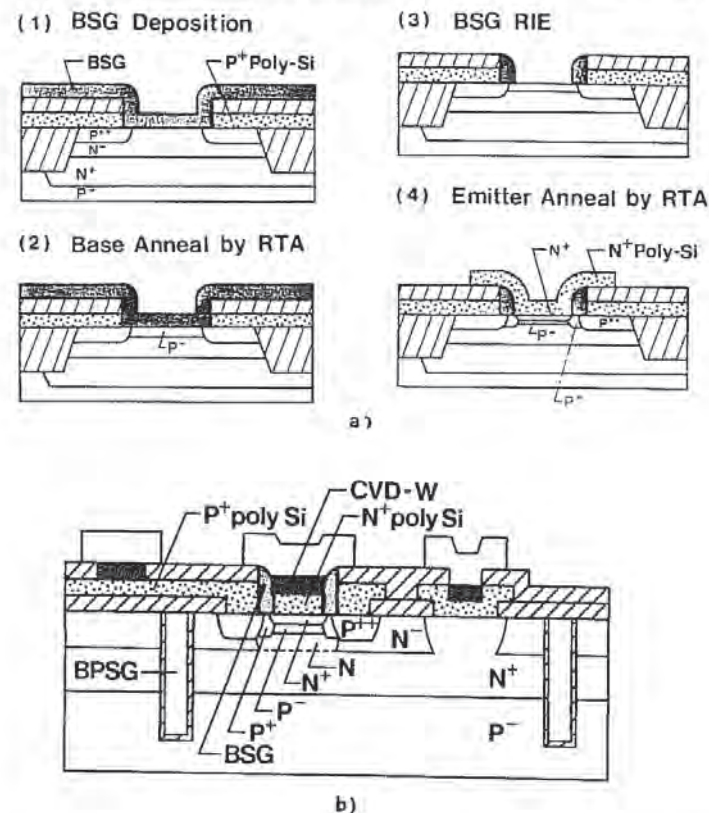


Fig. 7-41 (a) Key processing steps for borosilicate-glass, self-aligned transistor formation.⁶⁷ (© 1987 IEEE). (b) Cross section of the advanced BSG transistor.¹³⁷ (© 1989 IEEE).

concentration are smaller in the base region beneath the emitter than they are along its sidewalls. In bipolar transistors in which the minimum emitter dimension is greater than about $1.5 \mu\text{m}$, the fraction of the current injected at the emitter sidewalls is so small that the overall current gain, β , is not significantly reduced by the perimeter injection. However, in bipolar devices in which the emitter widths are $1 \mu\text{m}$ or less, the perimeter-to-area ratio of the emitter increases to such a degree that β is significantly reduced by the perimeter-injection effects.

For example, a 65% reduction in current gain of has been reported in devices with an emitter width of 0.5 μm , compared to those having 4.25- μm -wide emitters (and identical

emitter lengths).⁷³ This effect can become even more severe in SA structures in which the spacing is very small between the heavily doped extrinsic base and the emitter edge. An obvious remedy is to continue to scale the depth of the emitter so that the perimeter remains a small fraction of the total emitter area. Careful design of the base profile at the edges of the emitter is also mandatory.

7.9.1.3 Linkup-Region Formation. As noted, the connection between the extrinsic and intrinsic base regions in the double-polysilicon SA bipolar structure (i.e., the *linkup*, or *graft-base*, region) is critical. The space between the edge of the base and the emitter contacts is determined by the thickness of the sidewall oxide spacer. The amount of overlap between the two base regions depends on this thickness and on the amount of lateral diffusion from the p^+ poly. Control of these two parameters is vital in order to fabricate these SA structures with high yield.

If a heavy overlap is produced between these regions in the linkup region (Fig. 7-42a), low values of BV_{EBO} , β , and f_T are exhibited. On the other hand, if the space between the edge of the intrinsic base implant and the extrinsic base diffused regions is too great (Fig. 7-42b), the result will be anomalously high values of R_{B1} and low values of *perimeter punchthrough voltage* (i.e., between the emitter and collector through the lightly doped link region of the base). If the structure is formed according to the process indicated in Fig. 7-38, a proper choice of sidewall-spacer thickness (Fig. 7-42c) thus involves a trade-off between these device characteristics.⁷⁴ A study of such trade-offs by Sawada⁷⁵ and Chuang et al.⁶⁵ measured the effects of several different poly 1 outdiffusion cycles on the device characteristics.

An alternative approach is to use a separate procedure to form the linkup region.⁷⁶ In one such process, the contact hole is etched, and a thin oxide is grown by RTP on the poly sidewalls and the epi surface. A shallow BF_2^+ implant is used to form the lightly doped linkup region, and only then are oxide sidewall spacers produced (Fig. 7-40). Following the deposition of poly 2, a boron implant and diffusion are used to form the intrinsic base, and an arsenic implant and an anneal step are employed to create the emitter region. (A "local" collector implant to reduce the channeling tail of the boron linkup implant, as described in section 7.9.2, may also be performed in the same process step in which the boron-linkup implant is carried out.) Another approach for separately forming the link-up region is given in reference 77.

7.9.2 Single-Polysilicon Self-Aligned Bipolar Structures

Several other SA techniques that use only a single polysilicon layer (and thus offer the advantage of somewhat reduced process complexity) have also been developed. The first takes advantage of the fact that regions of highly doped silicon oxidize much more rapidly than lightly doped ones. A heavily doped, patterned polysilicon-emitter structure (beneath which an intrinsic base region has already been formed, shown in Fig. 7-43a) is oxidized at the same time as the bare, lightly doped substrate.⁷⁸ A much thicker oxide is grown on the poly, and a low-energy boron implant is used to form the extrinsic base region. The boron is prevented from penetrating into the polysilicon by the thick oxide,

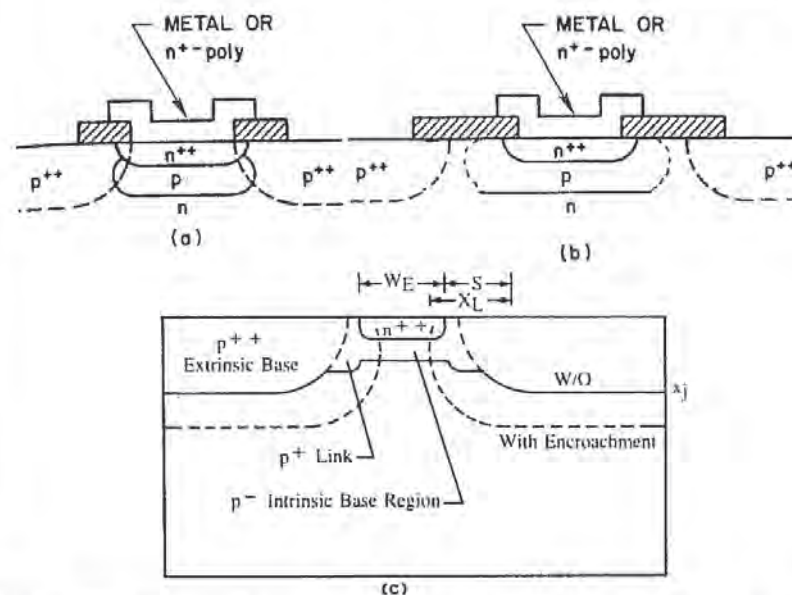


Fig. 7-42 (a) Heavy overlap between the intrinsic and extrinsic base regions. (b) Wide space between intrinsic and extrinsic base regions.⁷¹ (c) Selecting the appropriate overlap, based on various tradeoffs – s is the spacing from extrinsic base implant mask to emitter; x_L is the lateral diffusion from the edge of the implant mask; x_j is the junction depth of the extrinsic base.¹³⁴ (© 1987 IEEE).

but it can enter the silicon substrate through the thinner oxide grown there. The arsenic outdiffuses from the poly during the oxide-growth step to form the emitter, allowing the base to be self-aligned to the intrinsic base and the emitter diffusion.

In a slightly more complex version of this approach, a linkup boron implant is performed following etching of the 250-nm thick poly layer (Fig. 7-44).⁷⁹ An oxide spacer is then formed around the poly emitter, and the p^+ extrinsic-base implant is performed. The emitter is diffused during the anneal step following the extrinsic base implant. The base and emitter resistances are further reduced through the use of a silicide technique by which a silicide is formed on all of the exposed silicon.

An enhancement of the above process involves the deposition of an amorphous silicon layer in place of poly.⁸⁰ Following the formation of the linkup and extrinsic base regions, the amorphous silicon is implanted with boron. (Since the silicon is amorphous, the formation of a channeling tail is avoided.) The intrinsic base is formed through out-diffusion from the amorphous silicon. During this step, the silicon crystallizes with a larger grain size and different grain orientations than those of

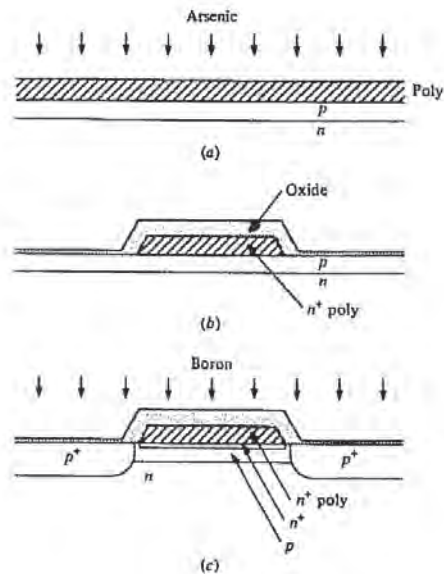


Fig. 7-43 Forming a single-polysilicon self-aligned transistor.⁷⁸ (© 1985 IEEE).

polycrystalline silicon. The altered poly grain structure results in a shallower E-B junction and a narrower base width than were the case when polysilicon was used as the diffusion source. This approach produced devices with good high-frequency characteristics.

Another single-poly process that uses removable sidewall spacers to self-align the linkup region to the intrinsic and extrinsic base regions has been described by Chen et al.^{77,87} As shown in Fig. 7-45,⁸⁷ the resulting topography is much less severe than that in double-poly SA structures. This approach was used to build sub-50-ps ECL circuits.

Another single-poly SA process, called SPOT (Single Poly Transistor) has been used to fabricate structures with minimum feature sizes of 1.25 μm .⁸¹ The primary goal in designing this structure was to create a simple, high-yielding process without seriously degrading device performance. A cross-section of a SPOT device is shown in Fig. 7-46. Following the conventional formation of buried collectors, deep collector contacts, and intrinsic base regions, a 300-nm-thick layer of polysilicon is deposited. Poly resistors and p^+ electrode regions are implanted with boron, and a nitride layer is

deposited. The poly electrode patterns are defined through PR patterning and RIE of the nitride and poly. Following stripping of the resist, the exposed substrate silicon and poly sidewalls are thermally oxidized to a thickness of 150 nm. A self-aligned extrinsic base region is formed by means of boron implantation through this oxide. The nitride is then removed, and arsenic is implanted into the emitter and collector regions of the poly. Following a final emitter drive-in step, PtSi is formed over the exposed regions of the poly electrodes. A two-level-metal process completes the process sequence.

In this simpler approach, however, the etching and oxidation may form a nonvertical boundary between the polysilicon and the adjacent oxide (Fig. 7-47). For example, since selective oxidation between the emitter and base is used to provide electrical isolation, the shape of the polysilicon over the emitter may be trapezoidal rather than rectangular following oxidation.

Because dopant is usually implanted near the surface of the poly and diffused through it into the underlying silicon substrate to form the emitter region, in trapezoidal poly structures dopant atoms must diffuse laterally as well as vertically to act as an emitter-forming dopant source. The resulting junction may be shallower near the emitter's corner than near the center, which can lead to degradation of the transistor gain, especially in devices with narrow emitters.⁸²

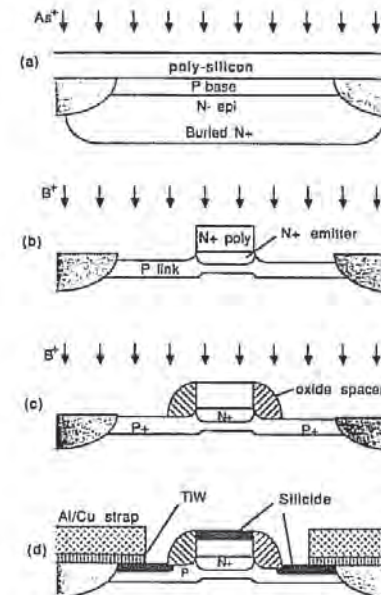


Fig. 7-44 Single-polysilicon self-aligned transistor with extra link-up implant.⁷⁹ (© 1988 IEEE).

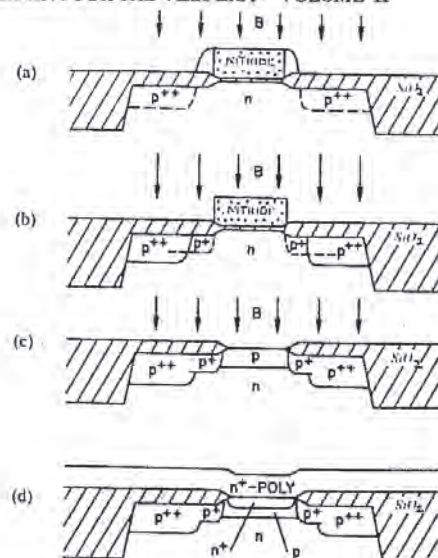


Fig. 7-45 Alternative approach for fabricating single-polysilicon self-aligned transistor.⁸⁷ (© 1988 IEEE).

7.9.3 Sidewall-Base-Contact Structures (SICOS)

Another self-aligned structure that can potentially result in even smaller nonactive device regions than the double- or single-poly SA structures is the sidewall-base-contact structure (SICOS; Fig. 7-48).^{83,84} In the structure shown, only the active transistor region is formed in the single-crystal-silicon epitaxial layer. The linkup base regions are formed through the diffusion of boron from heavily doped polysilicon layers. A buried oxide beneath these polysilicon base electrodes significantly reduces C_{CB} . Figure 7-49 shows the steps used to fabricate the SICOS structure shown in Fig. 7-48a. ECL gates with a minimum propagation-delay time of 63 ps have been fabricated using this structure.

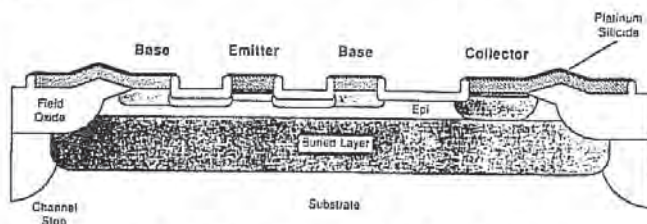


Fig. 7-46 Cross-section of the SPOT transistor.⁸¹ (© 1988 IEEE).

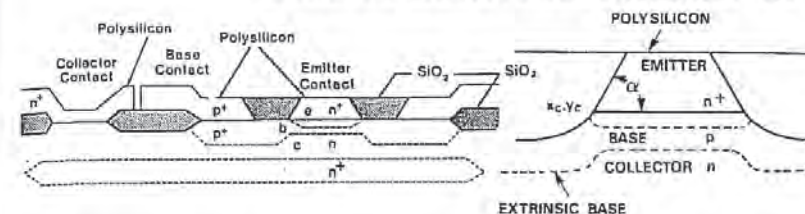


Fig. 7-47 Cross-section of the emitter region in SPOT transistor showing trapezoidal shape, which leads to nonuniform distribution of dopant emitter area.⁸² (© 1989 IEEE).

A SICOS structure in which a submicron linkup region is provided (to permit control of such problems as E-B forward tunneling and low E-B-junction breakdown voltages) has recently been described.¹³⁸ A cross section of this structure, called the *Best Alignment with Sidewall Contact (BASIC)* device is shown in Fig. 7-48b.

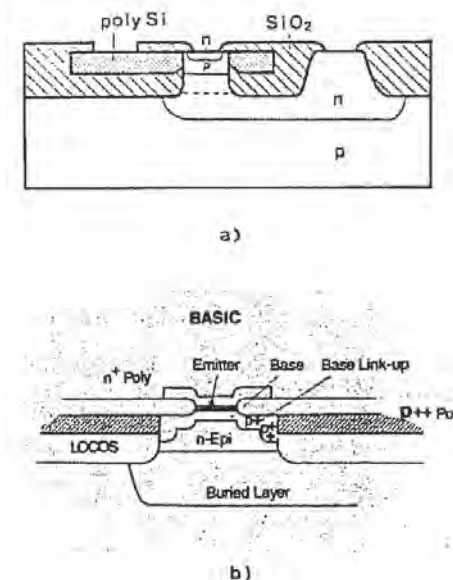


Fig. 7-48 (a) An advanced SICOS transistor.⁸⁴ (© 1979 IEEE). (b) Cross section of the BASIC SICOS device.¹³⁸ (© 1989 IEEE).

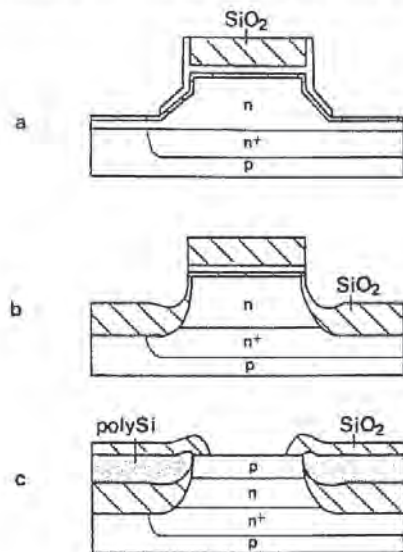


Fig. 7-49 Key process steps for fabricating a SiCOS transistor.⁸⁴ (© 1979 IEEE).

7.10 TRENCH-ISOLATED BIPOLAR TRANSISTORS

Because trench-isolated bipolar transistors greatly reduce the device area and the associated parasitic collector-substrate capacitance, they allow a significant increase in the density of bipolar circuits and contribute to a reduction in the power-delay product. In addition, the deep trenches produce larger collector-to-collector breakdown voltages. Such trenches also decrease the latchup susceptibility of the device structures. Figure 7-50 shows a cross-section of a trench-isolated self-aligned bipolar structure.⁸⁷ Deep trenches are used to isolate the collector regions, and LOCOS is used to form the field-oxide regions that abut the edges of the base and emitter regions (as is done in oxide-isolated bipolar transistors with walled base and emitter regions).

Several reports of such structures have been published.^{76,85,86} Still another report describes a bipolar isolation approach that uses 5- μm -deep trenches to isolate the collector regions and shallower BOX isolation structures as the field oxides.⁷⁷ This technology exhibits a 33-V breakdown between isolated collector regions. (The details of fabricating deep trench structures for isolation purposes are presented in chapter 2, section 2.6.3.)

Another paper dealing with the scaling property of trench isolation capacitance and its impact on high-performance ECL bipolar circuits has been published.¹³⁹ It shows that the trench isolation capacitance has a distinct dependence on the trench width, and that

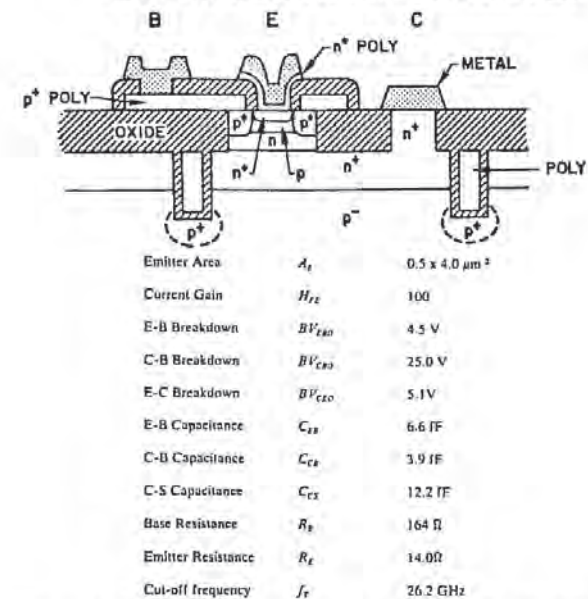


Fig. 7-50 Cross-section of a self-aligned, trench-isolated transistor.⁸⁷ (© 1987 IEEE).

the type of trench used is also significant. Four types of trenches (in order of decreasing capacitance: (1) Open-bottom, p^+ -poly fill; (2) Close-bottom, p^+ -poly fill; (3) Close-bottom, undoped poly fill; and (4) Close-bottom, oxide fill), were studied.

7.11 BiCMOS TECHNOLOGY

BiCMOS is a technology that integrates both CMOS and bipolar device structures on the same chip. This capability can be exploited in a number of ways to produce circuits with capabilities that exceed those that are possible when only one of the device types is used.

For example, while CMOS can implement low-power, high-density digital ICs with TTL outputs, such circuits are slower than ECL-based ICs implemented with bipolar devices (especially when driving large capacitive loads). Such loads are encountered on-chip when long interconnects are driven or high fan out occurs, as well as when signals must be sent off-chip.

Bipolar transistors, on the other hand, can deliver large drive currents, allowing heavy loads to be rapidly charged. Bipolar digital circuits implemented with ECL gates can also operate with small logic swings and high noise immunity. Conversely, ECL gates

exhibit high power consumption, poor density, and limited circuit options, and they do provide TTL outputs (i.e., an ECL output is produced).

BiCMOS offers the benefits of both bipolar and CMOS circuits. By appropriately trading off the characteristics of each technology, the speed-versus-power middle ground (not achievable with either bipolar or CMOS ICs alone) can be reached. For example, high-density, low-power logic arrays can be integrated with high-speed bipolar drivers to provide gate arrays that are faster than comparable CMOS versions, but that consume far less power than ECL arrays having approximately the same density. BiCMOS can also offer analog and digital functions on the same chip. Furthermore, either a TTL or ECL interface can be produced.

These benefits, however, are attained at the expense of longer chip-fabrication times and higher costs (due to the more complex technology-development and chip-manufacturing tasks). Thus, BiCMOS is not being pursued as a replacement for CMOS in digital systems with low and medium performance. However, the benefits of BiCMOS are often perceived as being significant enough to justify the extra cost for many high-performance digital applications, as well as in mixed analog/digital systems (although it should be noted that this opinion is not universally held).

7.11.1 Device and Circuit Advantages of BiCMOS

Integrating CMOS with bipolar devices can provide superior performance for a variety of circuit applications (particularly those that require circuits capable of operating between the very-high-speed but power-hungry ECL, and the very-high-density, medium-speed CMOS regimes). Advantages of BiCMOS include the following:

- Digital BiCMOS logic gates can provide more drive current than can comparably sized pure CMOS gates, and as a result, they can operate at higher speeds when large capacitive loads must be driven.
- The dc power consumed by BiCMOS gates can be as small as that consumed by pure CMOS gates, and the ac power can be even smaller.
- Digital BiCMOS circuits can be fabricated with either a TTL or ECL output.
- In the move from a pure CMOS to a BiCMOS process, the performance of the purely CMOS circuits can be increased without the need to reduce the minimum feature sizes of current-generation technology. Thus, the existing fabrication tools can be used for at least one more generation. For example, while 1-Mbit CMOS SRAMs have been built with access times of 9 ns, these have required the use of 0.5- μm design rules. 1-Mbit BiCMOS SRAMs with the same access times have been built using 0.8 μm design rules. As of this writing, the task of fabricating parts with 0.5- μm dimensions is quite risky, in part because it entails the use of new, expensive production equipment that is also difficult to work with.

7.11.1.1 Comparison of BiCMOS and CMOS Propagation-Delay Times. BiCMOS gates exhibit smaller propagation-delay times when driving large

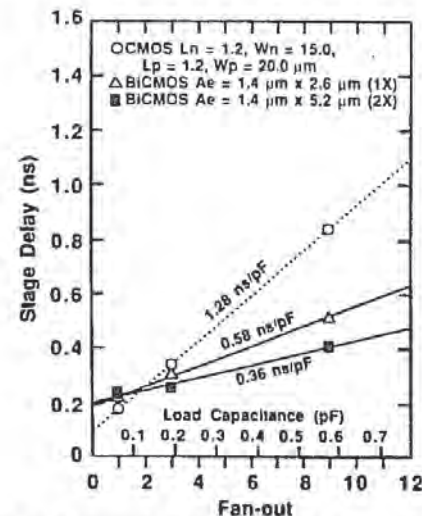


Fig. 7-51 CMOS and BiCMOS propagation delay times at 5 V versus load capacitance and fan-out.⁸⁹ (© 1987 IEEE).

capacitance loads than do pure CMOS gates of equal area. This is demonstrated in Fig. 7-51, where it is seen that the slope of the τ_{pd} -versus- C_L curve is smaller for BiCMOS gates than for CMOS gates.⁸⁹ It should be noted that under no-load (and very light loading) conditions, τ_{pd} is indeed smaller if CMOS gates are used. However, as the value of C_L grows larger, the reverse becomes true. In the example shown, when the loads exceed ~ 0.3 pF, BiCMOS is faster.

While heavy loads are normally encountered when gates must drive large wiring-capacitance or off-chip loads, they also exist in logic circuits when fan-out exceeds 2. Figure 7-51 also indicates that the delay times of BiCMOS gates do not increase as rapidly as those of CMOS gates with increasing C_L . This means that approximately constant values of τ_{pd} can be maintained, independent of loading conditions.

The basis of this advantage of BiCMOS can be explained with the aid of Figs. 7-52a and 7-52b, which show CMOS and BiCMOS inverters driving the same load capacitance, C_L . To a first order, τ_{pd} of an inverter can be expressed as:

$$\tau_{pd} = \tau_o + \Delta V_o C_L / \Delta I \quad (7-20)$$

where τ_o is an intrinsic delay time of the gate itself. For small values of C_L , CMOS gates are faster than BiCMOS gates because the bipolar stage of the BiCMOS gate represents a load to the CMOS portion of the gate that drives it. Roughly speaking, if the parasitic junction capacitances of the bipolar transistor are larger than C_L , more time

may be required for the CMOS portion of the gate to charge these capacitances than if only the C_L of the following gate had to be charged.* The value of C_L at which the BiCMOS-gate delay becomes equal to that of a comparably sized CMOS gate is important. This point changes as a function of many device parameters (e.g., L_{eff} , W_B , f_T , g_m , and C_j), as well as of external operating conditions (e.g., temperature). The degree to which the C_L -crossover value can be reduced without the need for complex processing measures is one of the factors determining the cost-effectiveness of BiCMOS in digital systems. (Note that for small values of C_L , this fact also implies that the use of CMOS inverters may yield higher performance.)

At sufficiently large values of C_L , however, the second term of Eq. 7-20 dominates. In such cases, τ_{pd} becomes inversely proportional to the transconductance ($g_m = \Delta I/\Delta V_{in}$) of the driver device in the gate, as shown by

$$\tau_{pd} \sim \Delta V_o C_L / g_m \Delta V_{in} \quad (7-21)$$

and when $\Delta V_o = \Delta V_{in}$,

$$\tau_{pd} \sim C_L / g_m. \quad (7-22)$$

Since it is well known that for equal-area devices bipolar transistors exhibit g_m values that are much larger than those of MOS transistors,¹⁰ the values of τ_{pd} in BiCMOS gates should be smaller than those in comparably sized CMOS gates.

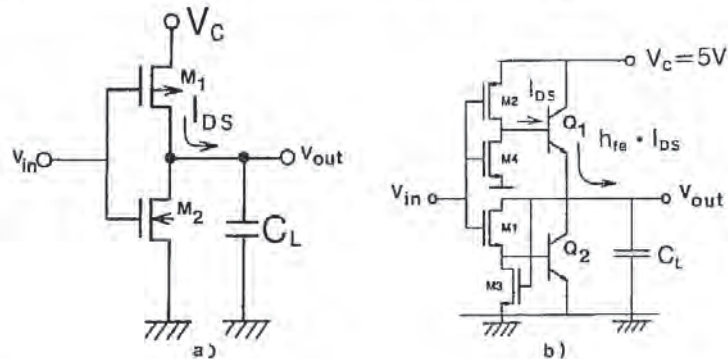


Fig. 7-52 (a) CMOS logic gate. (b) BiCMOS logic gate.

* When the basic digital inverter shown in Fig. 7-17a switches by having the input go from high to low, the current supplied by the load element charges C_L . In this case, $\Delta V_o/\Delta t = (\Delta Q_o/\Delta t)/C_L = I/C_L$. In the case of the CMOS inverter, the load element is an MOS device, and $I = I_{DS}$. In the case of a BiCMOS gate with a totem-pole output, as shown in Fig. 7-52c, the load element is a bipolar transistor, and $I = I_C$. More rigorous analyses of BiCMOS-gate propagation delays can be found in references 61 and 90.

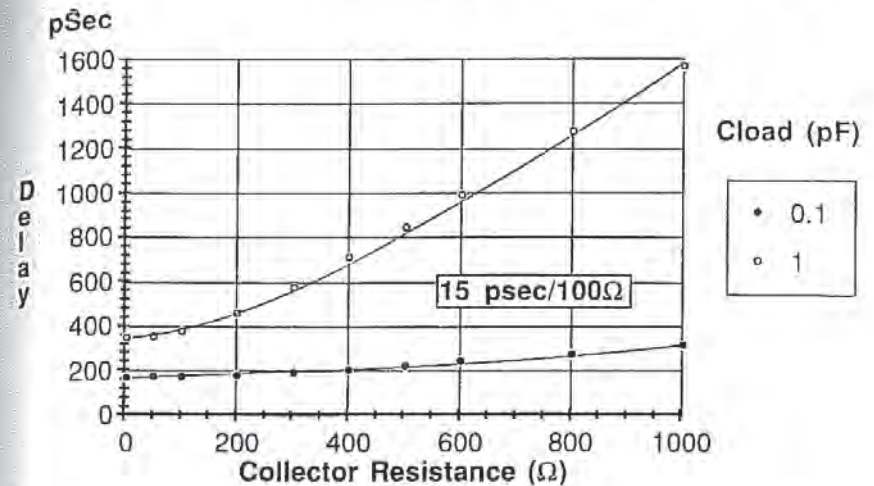


Fig. 7-53 BiCMOS gate delay versus R_C .⁹¹ (© 1987 IEEE).

Another way to explain this result is to note that the CMOS inverter can only provide a drive current I_{DS} for charging C_L , whereas the BiCMOS inverter can supply a current βI_{DS} where β is the current gain of the bipolar device (I_{DS} represents the base current to the bipolar load element in the BiCMOS inverter.) Note that the effective β can be much less than the small-signal β exhibited by bipolar transistors (e.g., the effective β can be as small as 5-7, compared to a typical small-signal β of 100 or more). The lower β value applicable to this case results because during switching the collector resistance, R_C , of the bipolar load element causes the transistor to go into saturation. Less drive current is then available to charge C_L . Figure 7-53 shows how an increase in R_C results in greater τ_{pd} values.⁹¹ Reference 90 gives more details about the changes in the BiCMOS gate-drive current during different stages of the switching cycle.

The difference between the g_m of bipolar and MOS transistors also gives BiCMOS some other advantages. The g_m of bipolar transistors is given by

$$g_{mbipolar} = qI_C/kT \quad (7-23)$$

whereas the g_m of MOS transistors is given by

$$g_{mMOS} = (2\mu C_{ox} I_{DS} Z/L)^{1/2} \quad (7-24)$$

Since $g_{mbipolar}$ depends on I_C , and I_C is directly proportional to the area of the emitter (A_E), an increase in A_E will lead to a linear increase in g_m . On the other hand, g_{mMOS}

products did not occur until the mid-1980s, when 5-V digital BiCMOS technologies were developed for logic (e.g., gate arrays) and memory (SRAM) applications as fabrication of device features smaller than $2\ \mu\text{m}$ became possible. Such small feature sizes permitted the implementation of functional densities that would have consumed too much power if they had been fabricated using bipolar technologies alone. (At larger feature sizes, pure bipolar digital ICs dissipated high – although still tolerable – power levels, and BiCMOS thus did not yet represent a significant advantage.) However, when it became feasible to fabricate devices with feature sizes smaller than $\leq 1.5\ \mu\text{m}$, BiCMOS made it possible to implement circuits with these densities which consumed less power than bipolar ICs, but offered speeds faster than those possible with CMOS alone.

Three categories of BiCMOS ICs have emerged: (1) low-cost, medium-speed 5-V digital; (2) high-performance, higher-cost 5-V digital; and (3) analog/digital. The distinctions among these are based on differences in the process flows used to produce them.

Low-cost BiCMOS parts are fabricated with slightly modified, single-well CMOS processes (i.e., only one or two mask levels are added to a baseline CMOS process). *High-performance digital BiCMOS ICs* are fabricated with twin-well, CMOS-based process flows that are significantly modified from the baseline CMOS process (e.g., three or four mask levels are typically added to the baseline flow). *Analog/digital BiCMOS ICs* are fabricated with processes designed to accommodate the larger voltage levels of analog applications (e.g., 10-30 V for low-voltage analog circuits, and more than 30 V for power applications). Analog/digital processes must also permit the production of the resistors, capacitors, and lateral *pnp* transistors needed in analog circuits, in addition to allowing the fabrication of *nnp* bipolar and CMOS structures. Analog-circuit requirements are such that they can still be met by devices built with less aggressive design rules than those needed in digital circuits (i.e., $1.5\text{-}2.0\ \mu\text{m}$ are sufficient for most applications).

Table 1-3 compares these three categories in terms of process complexity, with respect both to one another and to a CMOS and bipolar technology of the same minimum-feature-size generation.⁹⁶ (It is assumed that all the processes compared use double-level metal, and that the MOS devices are fabricated with LDDs.)

Table 7-3 Comparison of BiCMOS Technology Process Complexity.

Step	BiCMOS				
	CMOS	H. Perf.	Low cost	Analog	Bipolar
Masks	12	15	13	16	13
Etches (RIE)	11	12	11	12	11
Epi	Optional	Required	Optional	Required	Required
Furnace	16	19	16	19	16
Implant	8	12	9	13	7
Metal	2	2	2	2	2
Total	49	61	51	63	50

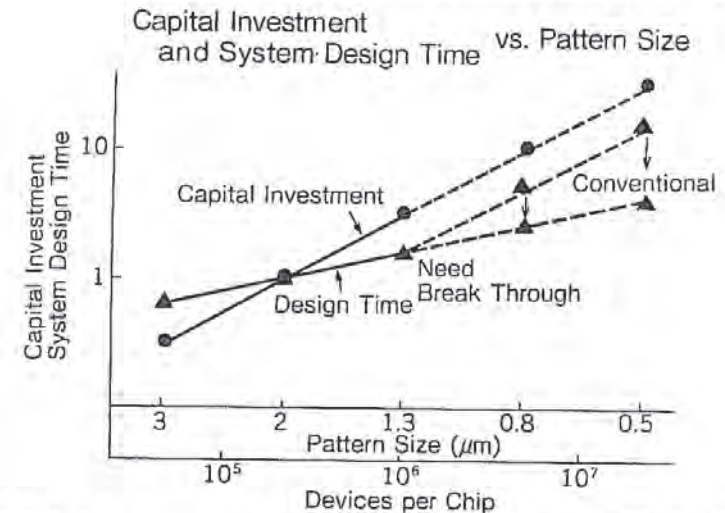


Fig. 7-55 Capital investment and system design time versus pattern size.⁹³ (© 1987 IEEE).

Since this book is essentially restricted to mainstream process-integration issues, the bulk of the BiCMOS process-integration discussion will be devoted to 5-V digital technology. However, issues involved in the fabrication of 10 to 30-V analog-digital BiCMOS technology will also be covered (see section 7.12.2).

7.12.1 Digital BiCMOS Technology

Digital ICs require high device density, low power dissipation, small propagation-delay times, and high noise immunity. Since the power-supply voltages of digital systems are normally 5 V or less, processes that produce devices with relatively small breakdown voltages can be used.

7.12.1.1 Low-Cost Digital BiCMOS Technology. As noted earlier, the first digital BiCMOS chips were designed with arrays of CMOS logic gates and bipolar I/O circuits. Since the I/O drivers could be designed to provide adequate drive current by increasing the size of the *nnp* transistors, it was acceptable to use bipolar transistors whose performance characteristics were inferior to those used in high-speed ECL circuits (in fact, in applications where the on-chip bipolar transistors are restricted to totem-pole-buffer applications, no significant speed performance is gained by using the bipolar transistors fabricated with a high-speed digital process).⁹⁷ Such early digital BiCMOS ICs still represented an incremental performance improvement over their CMOS counterparts. However, as long as the bipolar devices were restricted to the I/O

circuitry, the maximum benefits of a high-performance BiCMOS technology were not achievable.

Since even these early circuits contained many more CMOS than bipolar devices (due to the power-dissipation limitations of the latter), BiCMOS technologies have tended to evolve from CMOS processes. The initial approach used was to graft added steps to well-established CMOS processes to produce bipolar devices without degrading the characteristics of the CMOS transistors. Since less-than-optimum bipolar-transistor characteristics were acceptable in early BiCMOS designs, it was realized that such transistors could be produced by means of a baseline CMOS process with the addition of only one or two mask levels. As a result, only a small increase in fabrication costs was incurred.

The simplest low-cost BiCMOS technologies add only one mask to an existing n -well CMOS process (Fig. 7-56).⁹³ The added mask level is used to selectively produce lightly doped p regions, which serve as the p -bases of the bipolar transistors ($\sim 1 \mu\text{m}$ deep, with a doping level of about 1×10^{17} atoms/cm³). The n -well fabrication step also forms the collector regions of the $n\text{pn}$ transistors. The heavy implant used to form the NMOS source/drain regions is employed to produce the emitter regions and collector contacts of the bipolar transistors. The p^+ source/drain implant is also used to create the base contact, since this serves to lower the series base resistance, R_B . Semirecessed LOCOS structures isolate both the CMOS and bipolar devices. High-performance CMOS gates are achieved with this approach. The starting substrate is either a p -wafer or a p -epi layer on a p^+ wafer.

Such simple processes are also called *triple-diffused (3D) BiCMOS processes*, because they produce triple-diffused bipolar transistor structures. As noted earlier, the main drawback of 3-D bipolar transistors is the large collector resistance. (When the n -well is used as the collector, the collector sheet resistance is typically around $2\text{-k}\Omega/\text{sq.}$) The large R_C values that result cause internal debiasing of the base-collector junction at high currents (*quasi-saturation*), and this reduces the drive current (which in turn increases the gate delay). R_C in such transistors can be reduced by increasing the transistor size. Doing so, however, increases the collector and emitter junction capacit-

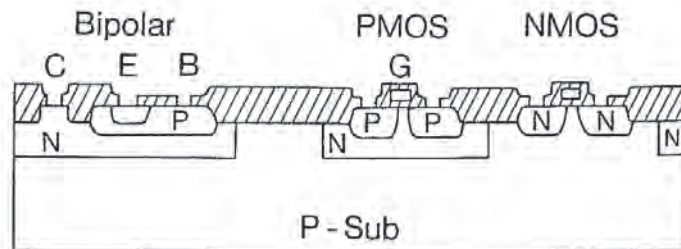


Fig. 7-56 3-D BiCMOS structure cross section.⁹³ (© 1987 IEEE).

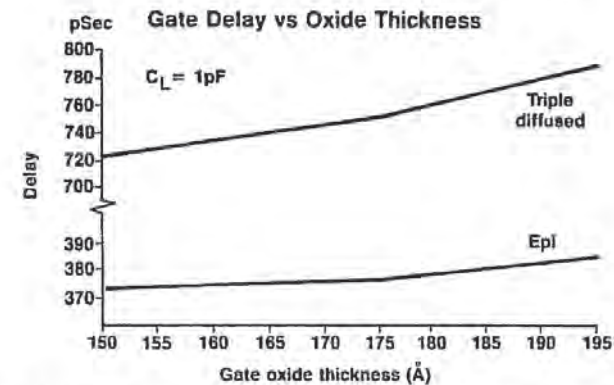


Fig. 7-57 BiCMOS inverter gate delay as a function of gate oxide thickness for 3-D and twin-well BiCMOS gates, $V_{CC} = 5 \text{ V}$, and $L = 1 \mu\text{m}$.¹³⁵ Reprinted with permission of Semiconductor International.

ances, so that the bipolar drivers must then be driven with larger values of I_{DS} from the CMOS gates. To produce sufficiently large I_{DS} values, the performance of the CMOS gates must be optimized; this implies the use of thin gate oxides. Figure 7-57 shows the gate delay of a 3D BiCMOS inverter as a function of gate-oxide thickness.

A second, somewhat more complex BiCMOS fabrication process adds two masking layers to a baseline n -well CMOS process. We refer to this method as the *collector-diffusion isolation (CDI) BiCMOS process*, since the bipolar transistors have the same structures as those resulting from the CDI process described in chap. 3, section 2.1.1.2.

In this process, n^+ buried layers are formed in a p -substrate beneath the n -wells. (The use of heavily doped n^+ layers beneath the n -well not only reduces collector resistance but also reduces the susceptibility to latchup; see chap. 6.) A thin p -epi layer, which will serve as the base region, is then grown. The n -wells are formed next and are driven to a sufficient depth so that they merge with the buried n^+ layers. The merged n -regions also form the low-sheet-resistance collector regions and the isolation structures of the $n\text{pn}$ transistors. To provide collector regions with even smaller R_C values, a second additional mask is used to create a deep n^+ connection between the surface and the buried n^+ layers. Note that the bipolar collector regions are isolated from one another by the p -substrate and p -epi regions.

The CDI BiCMOS approach improves the performance of the $n\text{pn}$ transistors by decreasing R_C . However, the performance is still not as high as it is in ECL gates for several reasons:

- The base is formed from the relatively lightly doped epi layer, resulting in a large R_B value.

* The emitter region (created during the ion-implantation step that forms the NMOS source/drain regions) does not yield transistors with as high a performance as is the case when polysilicon emitters are used.

* The minimum spacing of n^+ to n^+ regions is limited by the lightly doped p -regions that separate them. Thus, the maximum packing density is smaller with the CDI process than with submicron-CMOS processes (which do not employ local n^+ -buried regions).

7.12.1.2 High-Performance Digital BiCMOS Technology. This technology strives to produce CMOS and bipolar devices with performances equal to those of devices found in pure high-performance CMOS and bipolar ICs (so-called *uncompromised* device structures). Two basic approaches have been taken: (1) modification of a p -well CMOS process through the addition of three masking steps, and (2) modification of a twin-well CMOS process through the addition of three or four masking steps (Fig. 7-58).

In the first process, a lightly doped n -epi layer is deposited on a p -substrate in which n^+ regions have been selectively produced (Fig. 7-59a). This forms the collector regions of the bipolar transistors (as is done in the *standard-buried collector [SBC]* bipolar process described in section 7.4). Thus, such processes are sometimes referred to as *SBC BiCMOS processes*.

The p -well step is also used to form the junction-isolation structures of the bipolar transistor collectors (also akin to the procedure used in conventional junction-isolated

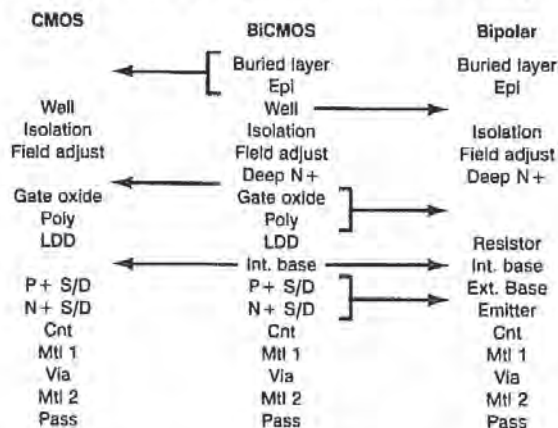


Fig. 7-58 Process flow for high-performance BiCMOS technology.

*Alternatively, trench isolation could be used, and such a process has in fact been reported.⁸⁸ Of course, process complexity is increased (the process just mentioned requires 18 masking steps).

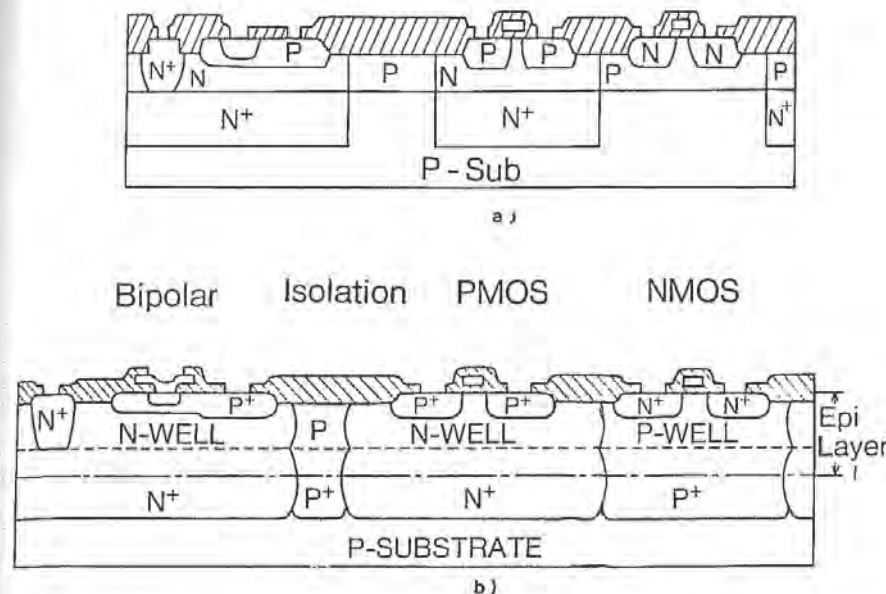


Fig. 7-59 (a) Cross-section of SBC BiCMOS structure. (b) Cross-section of twin-well BiCMOS structure.⁹³ (© 1987 IEEE).

SBC processes). The three masking steps added to the p -well CMOS process are the n^+ buried layer mask, the collector deep- n^+ mask, and the base p -mask.^{89,99}

While the SBC process is simpler than the modified twin-well process, the device performance is lower, for two reasons. First, the packing density is limited by the low p -substrate doping level (that is, the collector-to-collector spacing must be wide enough to prevent punchthrough from one bipolar collector to another). While the substrate doping could be raised to allow closer device spacing, this would increase the collector-to-substrate capacitance.* Second, the n -epi layer must be counterdoped to isolate the n -well regions and to form p -wells for the NMOS devices. However, adding sufficient p -dopant to the 0.3–1.0 $\Omega\text{-cm}$, 1.5–2.0 μm -thick n -epi layer can result in not only processing problems but also in a reduction of the NMOS performance through mobility degradation.

Therefore, modified-twin-well BiCMOS processes (Fig. 7-59b) have been the most widely used for achieving high-performance, high-density digital BiCMOS ICs. Many such processes have been reported.^{100,101,102}

7.12.1.3 Device-Design Issues of the Modified Twin-Well BiCMOS Process. The design of the bipolar and CMOS device structures in BiCMOS circuits follows many of the guidelines presented earlier (both in this chapter and in chapters 5 and 6) describing the relationships between device characteristics and processing in NMOS and CMOS ICs. This section focuses on the impact of BiCMOS process integration on bipolar and CMOS device design.

The three circuit-performance parameters that must be optimized in a digital BiCMOS technology are speed, power dissipation, and noise margins. These parameters are dependent not only on device characteristics but also on the circuit architecture (which also plays a role in determining the ratio of FETs to *npn* transistors on the chip). Thus, device optimization is normally carried out after the circuit-architecture and application operating conditions have been specified. The optimum set of device characteristics is then evolved by making the appropriate trade-offs in the device-structure design. A fabrication process sequence can then be developed.

Many tradeoffs must be made to obtain the smallest possible values of τ_{pd} . For example, if the device dimensions and doping concentrations were selected merely to yield minimum values of τ_{pd} , many other device parameters would likely be degraded to an intolerable degree. Some speed-performance benefits must therefore be sacrificed. As noted, such trade-offs will also be a function of the circuit architecture.

For example, if it is decided to restrict the bipolar *npn* transistors to the I/O circuits and voltage regulators, relatively low-performance *npn* transistors may be acceptable (i.e., transistors with large parasitic junction capacitances and R_C values could be used). However, the MOS devices would then have to be able to supply a large I_{DS} so that the high capacitive loads of such *npn* transistors could be rapidly charged. This implies that short-channel FETs with thin gate oxides would be desirable.

On the other hand, if advanced, self-aligned *npn* transistors were used, the circuit speed could be limited by on-chip wiring instead of by the parasitic capacitances of the bipolar transistors. In such cases, ease of manufacturing considerations would make it advantageous to use a process that produced FETs with thicker gate oxides (and consequently, small I_{DS} values).

To make this kind of design decision it is necessary to analyze the gate delay as a function of length and oxide thickness (using device and circuit simulators). As was shown in Fig. 7-57, the use of thinner gate oxides in a BiCMOS process that uses low-performance bipolar transistors (3D) will reduce the gate delay to some degree, whereas the gate delay is almost independent of gate-oxide thickness in a process that uses higher-performance *npn* transistors (designated as *epi* transistors in Fig. 7-57).

The use of high-speed bipolar transistors in digital BiCMOS ICs will certainly improve the propagation delay of BiCMOS logic gates. However, the incremental improvement in τ_{pd} versus f_T of the *npn* transistors in the gate must also be studied. For example, Fig. 7-60 shows that as f_T is increased, τ_{pd} decreases.¹⁰³ However, for f_T 's beyond about 8 GHz, the incremental improvement in τ_{pd} is small. Since significantly greater process complexity is required to fabricate *npn* transistors with f_T 's of 14 GHz than 8 GHz, such an extra processing effort would most likely not yield

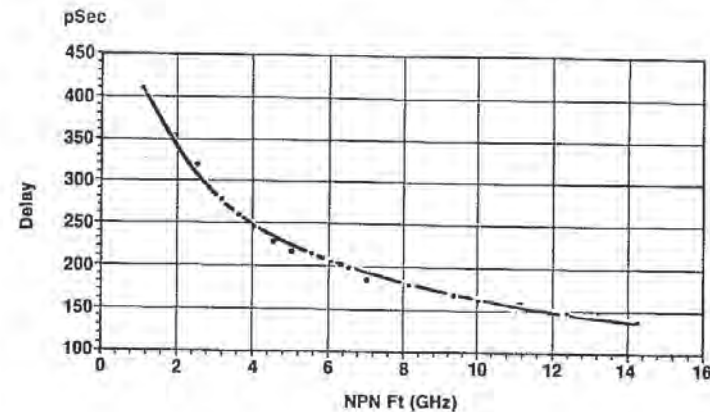


Fig. 7-60 Example of BiCMOS gate delay time versus the f_T of the *npn* transistors.¹³⁵ Reprinted with permission of Semiconductor International.

much improvement in speed performance.

The major design trade-offs in high-performance digital BiCMOS involve the characteristics of the epitaxial layer and the well profiles. From the perspective of bipolar device needs, the minimum epi thickness, $t_{epi(min)}$, is determined by the breakdown voltage of the bipolar transistors (BV_{CEO}), the collector capacitance, and the manufacturing controllability. As shown in Fig. 7-10b, for a BV_{CEO} of 7 V, a minimum epi thickness of $\sim 1.25 \mu m$ is needed.¹⁰⁴ On the other hand, the epi layers should not be too thick. That is, thicker epi layers will reduce f_T and increase R_C , leading to decreases in drive-current capability. In addition, thicker epi layers require longer *n*-collector-diffusion heat cycles, and *p*-isolation diffusion times. The latter considerations generally set an upper bound on the epi-layer thickness.

From the perspective of CMOS device requirements, the use of *p*-buried layers beneath the NMOS devices for adjacent *n*-well isolation dictates that $t_{epi(min)}$ be large enough to prevent the NMOS body-effect coefficient and junction capacitance from being excessively increased (due to the updiffusion of the *p*-buried layers). However, by keeping the *p*-buried layer dose low and the epi layer sufficiently thick, it is possible to decouple these NMOS parameters from the *p*-buried-layer parameters.

The *n*-well in twin-well BiCMOS not only impacts the PMOS devices but also serves as part of the collector structure of the *npn* bipolar devices. Thus, its profile is derived through the following tradeoffs: From the bipolar-device perspective, the well should be doped heavily enough to prevent the Kirk effect, yet lightly enough to provide an adequately high value of BV_{CBO} . (A more heavily doped well will also cut off the ion-implantation-channeling tail of the base, thus producing a smaller base width and consequent higher f_T .) The specifications of minimum drive-current capability, f_T , and

BVCBO dictated by the circuit application will drive the optimization of the well-profile design.

Both the n -well and p -well profiles must be selected so that the following device characteristics are optimized: (1) threshold voltage, (2) punchthrough voltage, (3) source/drain junction capacitance, and (4) body effect. A blanket-boron implant is typically used to provide the desired V_T for both the NMOS and PMOS devices. While the doping profile in the well beneath the surface must be high enough to prevent punchthrough, such higher concentrations will lead to larger source/drain capacitance and body effect values (see chap. 5). Compromises are therefore necessary.

An example of the details involved in the design of an n -well profile for a $0.8\text{-}\mu\text{m}$ digital BiCMOS technology is given in references 105 and 106.

7.12.1.4 A Process Sequence for Fabricating High-Performance 5-V Digital BiCMOS ICs. The following process sequence was developed at Texas Instruments to implement a $0.8\text{-}\mu\text{m}$ high-performance digital BiCMOS technology and has been described in several reports.^{105,107} Figure 7-58 shows the twin-well CMOS baseline flow and the additional process steps that have been added to integrate bipolar devices. Four masking steps are added to an advanced $0.8\text{-}\mu\text{m}$, 12-mask CMOS process¹⁰⁸ to integrate high-performance bipolar transistors. Since most of the steps are essentially identical to those used in a twin-well CMOS process (described in detail in chap. 6), we will focus here on the steps that have been added to incorporate the bipolar devices. Note that of the 16 masking steps used in this BiCMOS process, the extra ones are #1, #4, #5, and #6.

The starting material is a lightly p -doped ($\sim 10\ \Omega\text{-cm}$) silicon wafer with a (100) orientation. Buried n^+ layers (with doping levels exceeding $1 \times 10^{19}\ \text{atoms/cm}^3$) are selectively formed as described in the discussion of the SBC process in section 7.5.2 (Fig. 7-61a). *Mask #1* is used to pattern the buried-layer regions. Note that the lateral diffusion that occurs during the buried-layer anneal step (following the As or Sb implant) limits the spacing between adjacent subcollector regions and thereby also effectively sets the minimum n -well-to- n -well spacing.

Next, a selective boron implant ($1\text{--}2 \times 10^{13}\ \text{cm}^{-2}$) is performed (Fig. 7-61b) which places additional p -dopant between the n^+ buried layers.* A self-aligned approach requiring only one mask to create both the n^+ - and p -buried layers (as also described in ref. 106) is widely used, but two-mask processes can also be employed.

Following formation of the two types of buried layers, all oxide is removed from the wafer. A thin ($1.0\text{--}1.5\text{-}\mu\text{m}$), near-intrinsic epi layer (i.e., containing fewer than 1×10^{15} dopant atoms/cm³) is then deposited. The use of such low background-doping levels in the epi layer makes it unnecessary to perform excessive counterdoping during n - or p -

* These p -buried-layer regions serve to maintain adequate punchthrough breakdown-voltage values between adjacent n^+ buried layers. The p -concentration must thus be high enough to provide isolation between the buried n^+ layers, but not so great as to produce high sidewall capacitance for the n^+ buried layer.¹⁰⁶

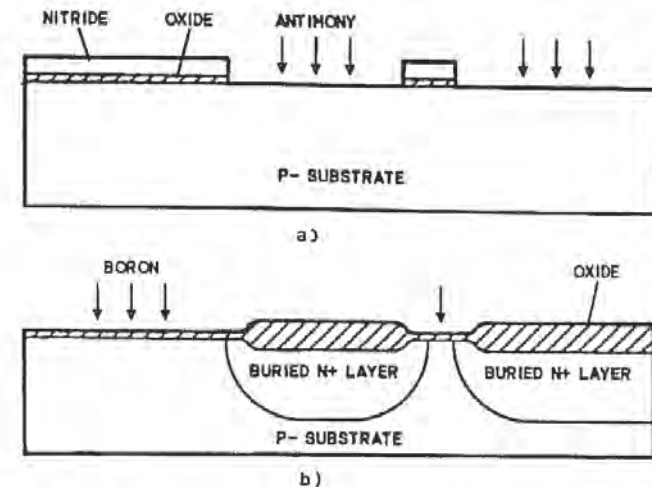


Fig. 7-61 (a) Formation of the n^+ -buried layers. (b) p -implant self-aligned to the n^+ buried layers to form p -buried layers.¹⁰⁷ Reprinted by permission of Semiconductor International.

well formation (which would degrade mobility). As described in the previous section, the choice of epi thickness is dependent on the particular circuit-performance requirements.

Nevertheless, the growth of a lightly doped, thin epi layer over two types of buried layers presents some difficult challenges for the epitaxial deposition process. Autodoping for both types of dopant must be minimized to avoid the need for excessive counterdoping in the wells. It has been shown that arsenic autodoping can be suppressed through the use of reduced-pressure epitaxy, but that, conversely, boron autodoping increases at lower pressures (Fig. 7-62).^{109,110,123} Antimony autodoping is only slightly reduced by a reduction in pressure. Compromises must thus be made in the selection of optimal epitaxial-growth conditions.

A novel two-step epi process has been reported¹¹¹ in which a high-temperature (1150°C), low-pressure (10 torr) SiH_2Cl_2 -epi cap is deposited first, followed by a low-temperature (900°C), low-pressure (10 torr) SiH_4 -epi-layer deposition. This process produces an abrupt epi substrate transition and minimal arsenic and boron autodoping.

The steps for forming the twin wells are conducted next (Fig. 7-63a). The doping profiles in the wells are tailored to optimize both the CMOS and bipolar device characteristics, as was described in the previous section. A single-mask process can be used to form both well types, as described in chapter 6 (*Mask #2*). In the TI process, a dual-phosphorus implant (deep and shallow) is used to form the n -well. Following the implant steps that introduce the well dopants, a well drive-in step is performed.

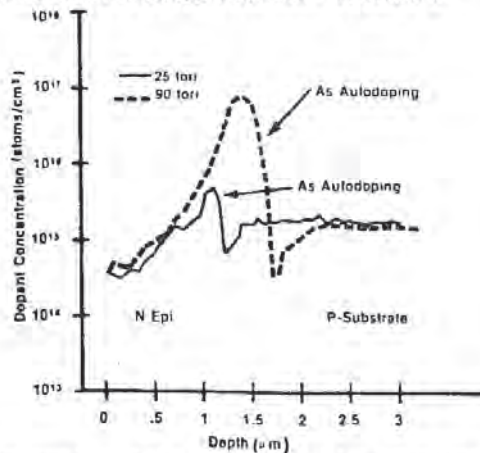


Fig. 7-62 Arsenic lateral autodoping peak off buried layer comparing 90 torr to 26 torr, 1080 °C (high-temperature/low-pressure) epi process.¹¹⁰ Reprinted with permission of Solid State Technology.

The active regions are then defined using *Mask #3*. A semirecessed LOCOS-isolation process is used to form the field oxide between the active regions (as was also described in chap. 6, section. 6.6). A boron channel stop is implanted prior to the field-oxide growth (Fig. 7-63b). In some processes, advanced semirecessed isolation structures are used (e.g., poly-buffered LOCOS or SILO), with a high-pressure field-oxidation growth step included to minimize dopant redistribution.

Following the stripping of the nitride, *Mask #4* is used to pattern the deep n^+ collector contact. A deep phosphorus implant and subsequent drive-in/anneal step are used to form the deep-collector-contact structure. A long thermal drive would have to be used to diffuse the implanted dopants all the way through the epi layer. However, because the lateral diffusion of the dopants from such a thermal cycle would be excessive for submicron BiCMOS processes, only a reduced thermal cycle can be tolerated. Unfortunately, this results in increased collector resistance.

To overcome this problem, a polysilicon-plug deep collector contact has been developed. A trench is etched in the silicon down to the n^+ buried layer. This trench is refilled with heavily in-situ-doped polysilicon to form a low-resistance plug to the buried collector. The sidewalls are covered with a dielectric to prevent lateral diffusion (Fig. 7-64). The use of this structure also reduces the area of the bipolar transistor by ~15% compared to devices that use conventional deep n^+ collector contacts. Following completion of this step, *Mask #5* is used for patterning the p -base and diffused resistor regions. A boron implant and drive-in is used to form these regions.

In some processes, an oxide is selectively grown over the base and resistor regions prior to the base implant. The boron is then implanted through the oxide (Fig. 7-65a).

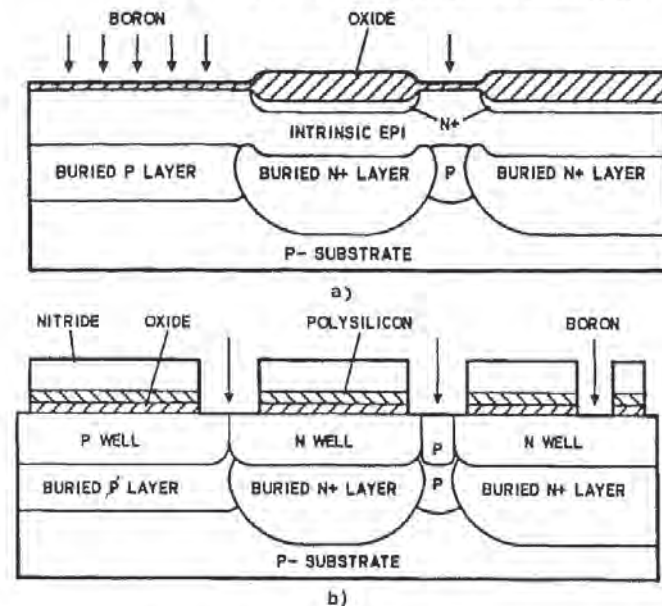


Fig. 7-63 (a) p -well implant self-aligned to oxide-masked n -well. (b) Channel-stop implant after etch of field isolation regions.¹⁰⁷ Reprinted with permission of Semiconductor International.

In this approach, the oxide-growth thermal cycle does not impact the base-doping profile. In other processes the oxide is deposited by means of CVD and is then selectively removed from the non-base/resistor regions. The purpose of the oxide (regardless of how it is formed) is to screen the base regions from the V_T -adjust boron implant that follows the base implant. That is, no mask is used during the V_T -adjust implant, but the oxide over the base regions prevents it from affecting the base. Following the base implant, the most recently formed oxide is removed.

In processes that use a polysilicon emitter, the emitter may be formed next. In the example TI process, this is accomplished through a *split-poly* process. A gate oxide is grown and is immediately capped by an initial poly layer. *Mask #6* and an etch step are then used to create openings down to the silicon substrate (i.e., by etching through the polysilicon and gate oxide). A second layer of poly is then deposited to make contact with the substrate. A blanket phosphorus implant is used to dope the poly so that a later thermal step will cause the phosphorus to diffuse and form the emitter region (see section 7.8.1). With the emitter formed in this fashion, the gate oxide of the MOS devices is protected by the first layer of poly from damage and/or contamination during the dry-etch, resist-strip step following the opening of the emitter contact. Other high-

performance BiCMOS processes that have used directly implanted emitters have also been reported.¹¹²

The polysilicon film is next patterned and etched using *Mask #7* and an anisotropic dry-etch process. *Mask #8* is used to selectively implant the phosphorus dose for the *n*-type LDD region. A CVD-oxide layer is then deposited and anisotropically etched back to form sidewall spacers on the patterned poly lines. *Mask #9* is used to selectively implant the NMOS source and drain regions, the collector and *n*-well contacts, and the NMOS gate-polysilicon lines. *Mask #10* is employed to allow selective implantation of the PMOS source and drain regions, the substrate and *p*-well contacts, the extrinsic base region of the bipolar transistor, and the PMOS gate-polysilicon lines (Fig.7-65b).^{*} Following these implants, a 900°C drive-in/anneal step is used to simultaneously form the diffused emitter and extrinsic base and to activate and drive-in the dopants of the NMOS and PMOS source/drain regions.

A silicidation of the polysilicon layers and the source and drain regions with a TiSi_2 is then performed, and a TiN local-interconnect layer is formed at the same time. As described in chapter 3, section 3.11.2.4, patterned TiN straps ($\sim 15 \Omega/\text{sq}$) are used as extensions of the source/drain regions and as local interconnects. *Mask #11* is used to pattern the TiN.

A double-level metal interconnect is then implemented. *Mask #12* is used to pattern contact holes in the CVD oxide deposited following the TiN processing. After Metal 1 is deposited, it is patterned with *Mask #13* and dry etched. *Masks #14* and *#15* are used to pattern the vias in the intermetal dielectric and the Metal 2 layer, respectively (Fig. 7-65c). *Mask #16* is the pad mask.

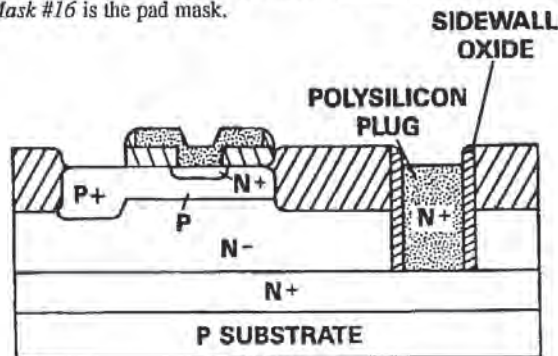


Fig. 7-64 Cross section of a bipolar transistor with a n^+ polysilicon plug used to minimize the collector series resistance. From A. R. Alvarez, Ed., *BiCMOS Technology and Applications*. Copyright, 1989, Kluwer Academic Publishers. Reprinted with permission.

^{*} Note that in this example process dual-doped gate-polysilicon lines are used. A dual-doped poly process can be successfully implemented because a TiN local interconnect layer is used to strap *p*-doped and *n*-doped poly lines together. The diffusion-barrier properties of TiN prevent counterdoping of one poly type by the other; see chap. 3, section 3.11.2.3.

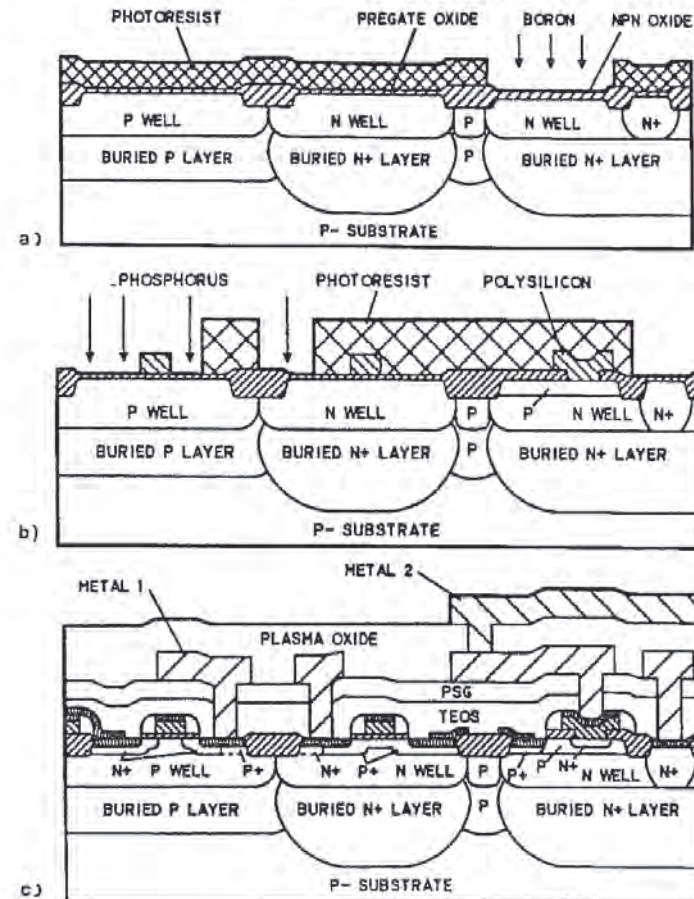


Fig. 7-65 (a) Patterned p^+ -type implant for forming base and resistor. (b) Phosphorus is implanted to form a shallow, *n*-type LDD region that is self-aligned to the edge of the NMOS gate. (c) Device cross-section following completion of double-level metal interconnect.¹⁰⁷ Reprinted with permission of Semiconductor International.

7.12.2 Process Integration of Analog/Digital BiCMOS

Many electronic systems employ both analog and digital circuits, and thus having the option of producing both types of circuit functions on the same chip can provide significant benefits. For example, CMOS can be used to minimize dc power dissipation

and provide high-impedance FET inputs for sample-hold operations. Bipolar devices can not only provide high current gain and extended gain-bandwidth capabilities, but they can also be used to minimize noise factors and provide good on-chip voltage references. This section describes the issues involved in designing BiCMOS processes that permit simultaneous fabrication of high-performance analog and digital functions. The applications of such chips will be described in section 7.12.3.4.

The differences between analog/digital BiCMOS and 5-V digital-BiCMOS processes stem primarily from the fact that analog functions generally operate over a much wider range of power-supply voltages (e.g., higher than 10 V) and power-dissipation levels. Since these higher voltages exceed some of the maximum operating limits of devices fabricated with 5-V-compatible BiCMOS processes (e.g., hot-carrier-degradation effects and gate-oxide breakdown and punchthrough voltages in MOS devices, and BV_{CEO} and BV_{EBO} values in bipolar devices), the device structures and process sequences must be significantly modified to make them suitable for analog/digital applications.

Two subcategories of analog-compatible BiCMOS processes have emerged to address these applications: (1) medium-voltage (10-30 V) processes that modify a baseline CMOS flow, and (2) high-power (>30 V and >1 A) processes based on power-bipolar process flows. Since the topic of power-IC technology is outside the scope of this text, our discussion will be limited to the process-integration issues of medium-voltage analog/digital BiCMOS ICs.

Medium-voltage analog/digital processes (10-30 V) seek to optimize the performance of both CMOS and bipolar devices in the face of voltages that are higher than those encountered in strictly digital applications.* Some CMOS speed performance must nevertheless be traded off to gain reliable operation at the increased voltage levels (e.g., thicker gate oxides are needed to withstand the higher gate voltages, and I_{DS} will thus be reduced.) A number of other necessary compromises are discussed in the following section, followed by details of a low-to-medium-voltage analog/digital BiCMOS process flow.

Analog functions also require passive circuit components (which preferably also exhibit small temperature and voltage coefficients), and the process sequence must be able to accommodate their fabrication. Resistors fabricated in polysilicon are preferred for precision applications because they exhibit smaller temperature coefficients than do resistors made in diffused regions of the substrate (e.g., 500-1500 ppm/°C versus 2500-3000 ppm/°C, respectively). Appendix A deals with the general aspects of IC resistor fabrication, and section 8.1.2.3 in chapter 8 gives additional details concerning the fabrication of high-valued polysilicon resistors.

High-performance capacitors are also key elements for analog CMOS and BiCMOS technology, especially in the areas of A/D converters and switched-capacitor filters. Polysilicon-polysilicon MOS capacitors find the widest use because they exhibit smaller

* Most high-performance, medium-voltage analog/digital BiCMOS circuits require 10-15-V operating voltages in order to maintain high signal-to-noise ratios and to permit the use of cascoding in analog designs, whereas lower-performance circuits used in automotive applications are typically operated between 15 and 30 V.

parasitic effects than do polysilicon-silicon capacitors. The polysilicon films must be heavily doped to minimize depletion effects in the capacitor electrodes.

A recent report, however, has described polysilicon-silicon capacitor structures for single-poly BiCMOS processes.¹¹³ These capacitors exhibit high capacitance per unit area, low voltage coefficients (e.g., lower than 50 ppm/V), and good dielectric quality. The capacitor interplate dielectric can be an oxide that is grown simultaneously with a gate oxide, or it can consist of oxide-nitride-oxide (ONO) layers that need one more mask than that of the oxide structure. These capacitor structures can also be integrated into a BiCMOS process in such a way that the capacitor implant (which is a high-dose, high-energy n^+ implant that forms a junction with a depth of about 1.2-1.6 μm) can also form the collector plug without the need for extra masks.

7.12.2.1 Process-Integration Issues of Medium-Voltage Analog BiCMOS. For analog-function circuit design, the most important device characteristics are those of the *npn* transistors. High gain is required to reduce input bias current into the transistor and to prevent the loading of a previous circuit stage. A value of 100 or more is usually satisfactory, and f_T values of at least 5 GHz for 10-V ICs and 2 GHz for 20-V ICs are desirable for high-frequency applications. (Such β and f_T values can be achieved with the example 20-V process sequence presented in section 7.12.2.2.) The transistors must also have low R_C values (e.g., less than 100 Ω) to allow high-current operation, as well as high Early-voltage (V_A) values (e.g., greater than 50 in a 10-V process and 25 in a 20-V process), since, as noted in section 7.3.2, the intrinsic small-signal voltage gain of an amplifier is proportional to V_A . Although low R_C values (e.g., less than 100 Ω) are achieved with the use of buried n^+ layers, thin epitaxial layers, and deep collector plugs, analog voltage requirements force the use of thicker epitaxial layers (which also reduce the f_T of the transistor). Similarly, large V_A values are obtained with the use of higher base Gummel numbers, but this also degrades the value of f_T .¹¹⁴ Tradeoffs must thus be made to arrive at optimum *npn* device designs.

Isolated *pnp* transistors are also needed in some analog circuits. Although lateral *pnp* transistors are usually "free" devices, their speed performance is consequently poor. That is, while they can exhibit values of β as high as 100 and large Early-voltage values, these transistors are slow ($f_T = 5$ -10 MHz) because of the need for wide base regions to prevent emitter/collector punchthrough. As result, their use must be avoided in high-frequency signal paths. (A process that allows vertical *pnp* transistors to be integrated into a 5-V analog/digital BiCMOS process is described in ref. 115.)

The higher operating voltages also limit the minimum gate lengths of MOS devices to 2-3 μm in analog/digital BiCMOS to prevent drain-induced barrier lowering (see chap. 5, section 5.5.2) and excessive channel-length modulation. Reduced junction capacitance, which implies the use of lightly doped wells, is thus desirable. Furthermore, since analog designs often use both positive and negative supply voltages, isolated CMOS structures are desirable. Finally, thicker gate oxides must be used (e.g., a minimum gate-oxide thickness of 30 nm must be used to allow reliable operation at 10 V).

Thicker field oxides are also usually needed to provide field-region threshold-voltage values that exceed the maximum operating voltage. Likewise, the CMOS devices must be protected against hot-carrier effects that arise as a result of exposure to higher supply voltages. Latchup is another concern due to the high substrate currents common in many analog BiCMOS designs.

7.12.2.2 An Example of an Analog/Digital BiCMOS Process. The following process sequence for fabricating medium-voltage analog/digital BiCMOS ICs, developed by Texas Instruments,^{116,117} is used to manufacture ICs capable of operating from 20-V supplies. The first-generation process uses 3- μm design rules (although these are predicted to be scaled to less than 2 μm by the third generation). Typical parameters for the devices are the following: β of 80, f_T of 2.5 GHz, and breakdown voltages of 25, 40, and 8 V, respectively, for BV_{CEO} , BV_{CBO} , and BV_{EBO} .

The process is a modified n -well CMOS process with four additional masking steps to permit the integration of high-performance bipolar transistors. The bipolar devices are of the collector-diffused isolation (CDI) type, which means that n^+ -buried layers and a p -epi layer are used. (Note that the buried layers are also placed beneath the n -wells to increase latchup immunity.) Bipolar transistors fabricated with CDI need no deep p^+ -isolation diffusions (unlike nnp transistors fabricated with the SBC process), yet they exhibit much lower R_C values than do 3-D bipolar transistors. They can therefore be made smaller than transistors produced with the SBC process. In CDI, the n -well region serves as the nnp collector region, and the p -epi layer provides the sidewall junction-isolation area for the collectors of the bipolar transistors and the n -wells.

The starting material is a p -type wafer with (100) orientation. The wafer surface is first oxidized. Mask #1 is then used to open windows in the oxide wherever n^+ buried layers are to exist. These buried layers are formed by means of an antimony implant and diffusion, and the oxide is then stripped. (Since the formation of the n^+ buried layers takes place first, it does not impact the CMOS device profiles.)

The p -epi layer is deposited next. The resistivity and thickness (t_{epi}) of this layer impact both the MOS and bipolar device properties, so some trade-offs must be made to achieve optimum characteristics in both device types. Typically, however, the resistivity of the epi layer is selected to be identical to that of the substrate doping employed in the n -well CMOS process, so that the CMOS device characteristics will not be impacted by the epi resistivity value.

The value of t_{epi} is more difficult to specify, since it depends on such device dimensions and processing factors as the n -well junction depth, the n -well doping level, the n^+ buried-layer updiffusion, the maximum operating voltage, the maximum β of the nnp transistors, and the thickness of the epi layer that is removed during subsequent oxidations.

The selection of the epi thickness is begun by satisfying the requirement that the n -wells must extend down to contact the n^+ buried layers (to establish low-resistance collector regions). This dictates the maximum epitaxial-layer thickness ($t_{epi(max)}$) for a given n -well depth (as illustrated in Fig. 7-66). The fact that the n^+ buried layers typic-

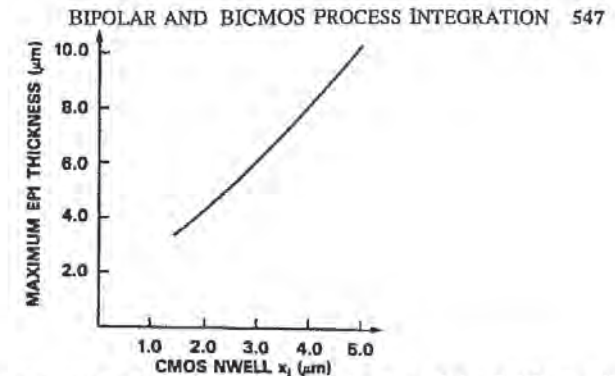


Fig. 7-66 The maximum p -epitaxial layer thickness is a function of the CMOS n -well diffusion depth. Approximately 1.0 μm of the p -epitaxial layer is removed through oxidations in this example.¹⁰⁷ Reprinted with permission of Semiconductor International.

ally up-diffuse a distance of 80-100% of the well depth, also plays a role in determining $t_{epi(max)}$.

Once $t_{epi(max)}$ has been determined from the well-formation design considerations, the minimum epi-layer thickness ($t_{epi(min)}$) must be determined. This value depends on three interrelated parameters: the maximum β that is to be exhibited by the nnp transistors, the maximum operating voltage, and the n^+ buried-layer updiffusion. We next illustrate how to determine $t_{epi(min)}$ based on these factors.

The maximum operating voltage of a bipolar circuit is usually limited by the BV_{CEO} breakdown mechanism, which depends on BV_{CBO} (plane) and β of the transistor, according to Eq. 7-13.

The value of BV_{CBO} (plane) is dependent on the epitaxial thickness. As the thickness is reduced, the vertical distance between the base and the edge of the up-diffused n^+ region is decreased, which reduces BV_{CBO} (plane). Figure 7-67a gives an example of the doping profile in a typical nnp transistor fabricated with a base-collector junction depth of 1.0 μm , a 4.5- μm -deep n -well, and an 8.0- μm -thick p -epi layer. Figure 7-67b then illustrates how BV_{CBO} in such a device varies as the epi thickness is decreased. With Eq. 7-13 and the information from Fig. 7-67b used, $t_{epi(min)}$ can be selected once the maximum operating voltage and maximum nnp β have been specified.

Example: Assume that an analog/digital BiCMOS process is used to produce bipolar nnp transistors with doping profiles as shown in Fig. 7-67a. The ICs must be able to withstand a maximum operating voltage of 15 V. The maximum β of the transistors is designed to be 150. Select an appropriate value for the epi-layer thickness, assuming that the parameter n in Eq. 7-13 has a value of 4.

Solution: Using Eq. 7-13, we calculate that in order for BV_{CEO} to exceed the maximum operating voltage of 15 V, the minimum value of BV_{CBO} (plane)

would have to be 53 V. From Fig. 7-67b, we see that the $t_{\text{epi(min)}}$ needed to achieve this BV_{CBO} value would be $7.0 \mu\text{m}$.

Following the epitaxial deposition, the wafer is oxidized again, and *Mask #2* is used to open windows in the oxide wherever n -well regions are to exist. The n -wells are formed by means of ion implantation and diffusion (Fig. 7-68a). The wells are partially diffused with one thermal cycle, and the subsequent deep n^+ collector cycle completes the n -well drive-in. The n -well doping levels are set by parameters that produce high-performance CMOS devices.

Mask #3 is used to produce the patterns of the deep n^+ collector regions, and these regions are then formed (e.g., through the diffusion of phosphorus alone, or through a high-energy, high-dose implant of phosphorus, followed by diffusion).¹¹³ Since a high temperature is needed for the diffusion, this step is also performed before the channel-stop steps to prevent adverse impact on CMOS device characteristics.*

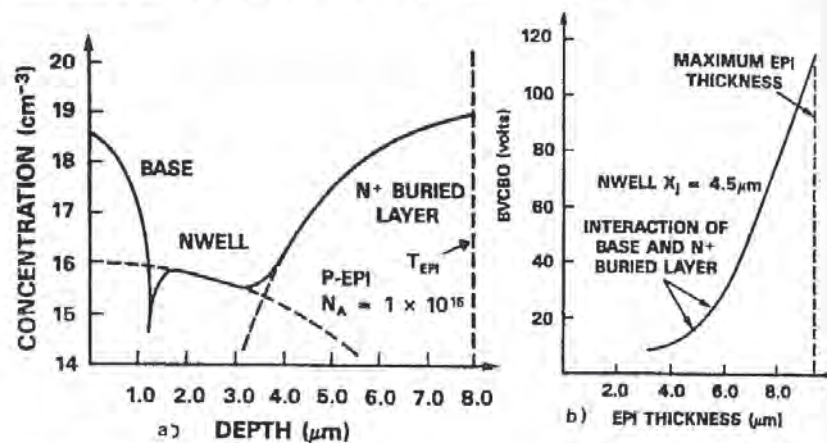


Fig. 7-67 (a) BiCMOS npn doping profiles for BV_{CBO} (plane) calculations. The solid line shows the net concentration while the dotted lines show the original profiles. (b) BV_{CBO} (plane) as a function of p -epitaxial layer thickness for a $4.5 \mu\text{m}$ n -well process. From A. R. Alvarez, Ed., *BiCMOS Technology and Applications*. Copyright, 1989, Kluwer Academic Publishers. Reprinted with permission.

* Note that the use of deep diffusions to form the buried layers, wells and, n^+ collector contact wastes silicon area due to lateral diffusion. Trench isolation could be utilized to reduce this loss of area and thereby increase packing density. However, the cost of the added complexity of would have to be weighed against the benefit gained.

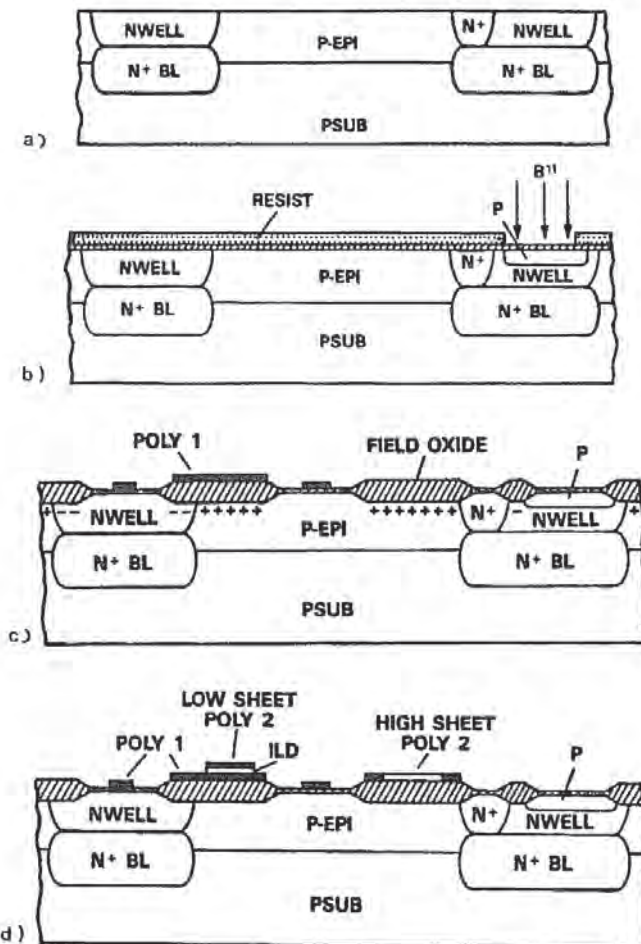


Fig. 7-68 (a) Formation of n -well and n^+ collector regions following the p -epitaxial layer deposition. (b) Definition of the base regions. (c) Deposition and definition of the 1st polysilicon layer for CMOS gates and MOS capacitors. (d) Formation of the 2nd polysilicon layer for resistors and capacitor plates.

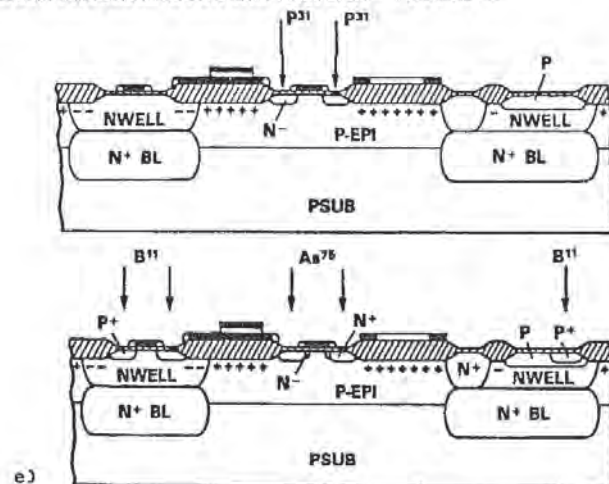


Fig. 7-68 (e) Formation of the n^+ emitter and collector regions. From A. R. Alvarez, Ed., *BiCMOS Technology and Applications*. Copyright, 1989, Kluwer Academic Publishers. Reprinted with permission.

After the oxide has been stripped, a thin pad oxide is grown. The base regions of the bipolar n_{pn} transistors are patterned with *Mask #4* and are then formed by means of ion implantation of boron (Fig. 7-68b). This is followed by an anneal in an inert atmosphere to activate the implant and reduce the implant damage (this step serves to reduce R_B).

A nitride layer is deposited over the pad oxide and *Mask #5* is used to define the active areas. After the nitride has been etched, a blanket phosphorus implant is carried out to produce a channel-stop dopant layer that will increase the threshold voltage of the parasitic field-oxide PMOS devices. Subsequently, another masking step is used (with *Mask #6*) to pattern the regions where a boron channel-stop implant is used to raise the threshold voltage of the parasitic NMOS field-oxide devices.

The resist is stripped, and a $1.0\text{-}\mu\text{m}$ -thick field oxide is grown. The nitride is then removed. A gate oxide of 35-50 nm is grown and a threshold-adjust implant carried out. The first level of polysilicon (*poly 1*) is deposited, doped, patterned (with *Mask #7*), and etched to produce the gate electrodes of the MOS devices and the bottom plates of the poly-to-poly capacitors (Fig. 7-68c).

Next, a 30-100-nm-thick capacitor interlevel-dielectric is formed over *poly 1* (by means of either an oxide or combination oxide-nitride-oxide [ONO] film). A second layer of polysilicon (*poly 2*) is then deposited and is appropriately doped to produce a high sheet-resistance film suitable for fabricating high-valued resistors. *Mask #8* is then used to protect these high-resistivity regions against the implant that follows. The unprotected regions are implanted with a heavy dose of arsenic so that they can be used

as the capacitor top plates and resistor contacts. *Mask #9* is then used to pattern *poly 2*. The subsequent etching of this layer completes the fabrication of the circuit resistor and capacitor structures (Fig. 7-68d).

The rest of the process follows conventional CMOS and bipolar process steps. *Masks #10* and *#11* are used to independently pattern the CMOS source and drain regions (with the p^+ -base-contact region of the n_{pn} transistors also being formed with one of these masking steps). NMOS LDD structures are formed, with a phosphorus implant used to form the lightly doped extensions of the drain and a heavy arsenic implant used to produce the source/drain regions of the NMOS devices. A boron implant is utilized to create the source/drain regions of the PMOS devices (and the p^+ base-contact regions).

Mask #12 is used to pattern the emitter and collector contact regions (Fig. 7-68e). A separate phosphorus implant is employed to produce these regions, since attempts to use the source/drain implants to also form the emitter regions have resulted in lower MOS g_m values.¹¹⁸ Furthermore, the deeper E-B junctions typically needed in analog devices are inconsistent with the requirements for high-density CMOS logic devices. This separate emitter-forming implant step allows both problems to be avoided, although at the cost of an additional masking step. It should also be noted that polysilicon emitters are also generally avoided in analog/digital processes because they yield higher (and relatively nonreproducible) emitter-contact resistance values than do directly-ion-implanted emitters. *Masks #13* and *#14* are used to pattern the contact holes and the metal interconnect layer, respectively. *Mask #15* is used to pattern an opening in the passivation layer. Note that if a double-level-metal process is used, the mask count would increase to 17.

7.12.3 BiCMOS Applications

7.12.3.1 Digital Logic Circuits and Gate Arrays. As noted earlier, one obvious way to integrate CMOS and bipolar devices on a single chip is to relegate the logic-function duties to high-density, low-power CMOS devices and to restrict the use of bipolar devices to the I/O circuits (Fig. 7-69).¹¹⁹ This approach was used in early digital BiCMOS ICs.

More advanced, higher-performance digital BiCMOS integrated the bipolar transistors into the logic gate (Figs. 7-70). Like conventional CMOS gates, these *BiCMOS logic gates* consume no dc power (except for leakage current), but they offer the possibility of smaller values of τ_{pd} when gates of comparable size are fabricated. Thus, although an area penalty is incurred when bipolar drivers are added to a CMOS gate (~20% increase in area for a CMOS inverter), the gate is actually denser in terms of its load-driving capability. Furthermore, the percentage increase in gate area declines as the fraction of the gate composed of CMOS components increases. Consequently, in some large, high-performance BiCMOS gate arrays (e.g., those containing 30 - 100K gates), bipolar devices are for the most part incorporated into such larger cells as adders, shift registers,

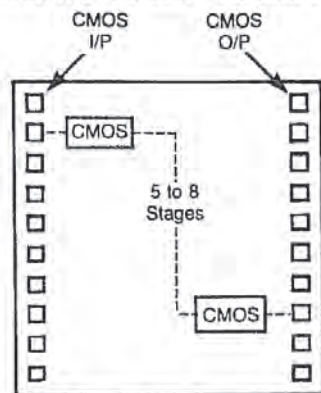


Fig. 7-69 BiCMOS gate array with CMOS gates and bipolar I/O.¹¹⁹ Reprinted with permission of VLSI Design.

and ST flip-flops. In these cells, the overhead penalty of using BiCMOS is decreased to as little as 5%.

BiCMOS logic gates are also attractive for gate-array applications because the fan-out in logic circuits implemented with such gate arrays is frequently 3-5 (and, in some instances, even larger; see Fig. 7-71, which illustrates the typical fan-out distribution in a modern VLSI gate array).^{120,121} Such large fan-out represents a heavy load capacitance, and BiCMOS gates can drive such loads faster than CMOS gates. Furthermore, the device size in gate arrays is typically uniform (for ease of physical design). Therefore, the delay degradation per unit load for CMOS circuits generally differs with the circuit function. In BiCMOS, the load degradation is practically the

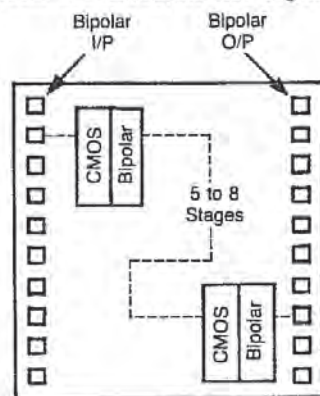


Fig. 7-70 BiCMOS gate array with BiCMOS gates and bipolar I/O.¹¹⁹ Reprinted with permission of VLSI Design.

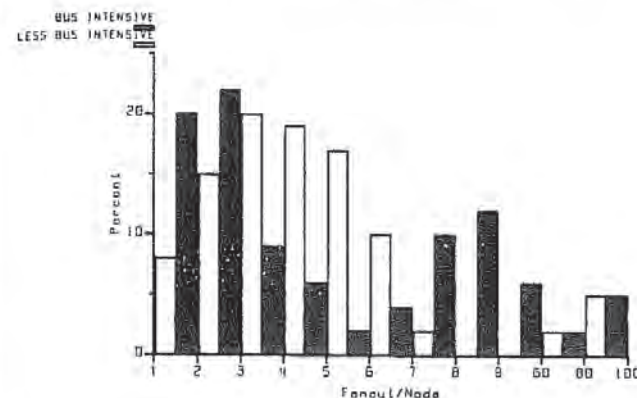


Fig. 7-71 Typical gate array fanout per internal output for options of low and high internal bus utilization.¹²⁰ (© 1988 IEEE).

same for all circuit functions, since the bipolar push-pull devices isolate the CMOS circuits from the loading. This makes the semicustom design task easier.

In some arrays, macrocells with uncommitted CMOS and bipolar drivers are offered (Fig. 7-72).¹²² The user can connect the transistors in each cell to provide an appropriate MOS and bipolar device mix for each logic gate. Many gate functions can be implemented by connecting MOS devices within a single macrocell through the use

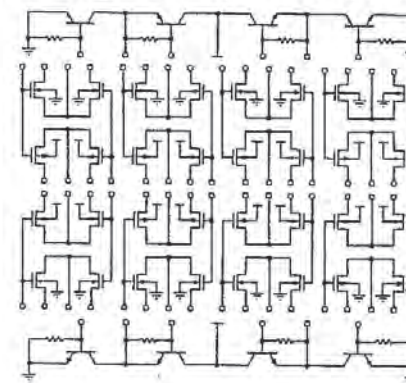


Fig. 7-72 Example of BiCMOS gate array macrocell.¹²² Reprinted with permission of Electronics.

of local interconnects. Since bipolar drivers are available for heavily loaded on-chip nodes, the MOS devices do not have to be designed to drive global interconnects (as must be done in pure CMOS arrays). Hence, gate widths in MOS gate arrays can be made much smaller than those in CMOS.

The large signal swing in CMOS often results in noise spikes. BiCMOS implemented with ECL outputs offers much smaller signal swings, so switching noise can also be better controlled.

BiCMOS gate arrays with gate delays of 0.8-0.9 ns have been fabricated using 1.5-2.0- μm technology. If 0.8- μm design rules are used, gate delays as small as 0.3 ns are expected.

7.12.3.2 Interface Driver Circuits. The latest generations of 32-bit microprocessors are being fabricated in CMOS. The off-chip buses that must be driven by these microprocessors represent large capacitive loads. *Interface driver circuits*, which have traditionally been implemented in bipolar technologies, are used to increase bus-driving speeds. A typical 32-bit microprocessor system, however, may contain 10 or more interface devices. Just one bus driver at a time is activated, causing a typical driver to be in the disabled state as much as 90% of the time. When the interface units are purely bipolar, the disabled drivers are all burning up power, since they remain in the high-impedance state awaiting their turns to be activated.

However, when BiCMOS bus drivers using cells similar to those described in the previous section are utilized, the current drawn by the disabled drivers is much lower. In some cases, power savings of nearly 100% can be achieved. Since studies indicate that traditional bus interfaces consume as much as 30% of a system's total current, such a savings can be significant. Propagation-delay times for the BiCMOS drivers are also comparable to those in advanced bipolar technology (e.g., 3-5 ns).

7.12.3.3 BiCMOS SRAMs. BiCMOS SRAMs perform better than CMOS or bipolar SRAMs in certain applications (for example, large, high-speed ECL SRAMs [up to 64 kbits] have power-dissipation and yield problems [see chap. 8, section 8.2.2.1]). On the other hand, CMOS alone cannot be used to build high-performance, higher-density SRAMs, since the driving capability of CMOS is inferior to bipolar and since it is practically impossible to design input and output buffer circuits in CMOS that have an ECL I/O capability.

BiCMOS SRAMs, however, offer densities and power-dissipation levels close to those of CMOS SRAMs, but with higher operating speeds. Figure 8-8 in chapter 8 shows the speed-versus-density characteristics of bipolar, CMOS, and BiCMOS SRAMs.³⁴ ECL I/O BiCMOS can be seen to offer higher speeds than bipolar ECL at middle and higher densities (because of the severe speed, power, and density trade-off in bipolar ECL). In addition, BiCMOS SRAMs are somewhat faster than CMOS SRAMs at equivalent power and density for the same geometry. BiCMOS thus offers a range of speed and power trade-offs (from very fast, low-density BiCMOS at high power, to slightly slower, mid-density memory circuits at moderate power).

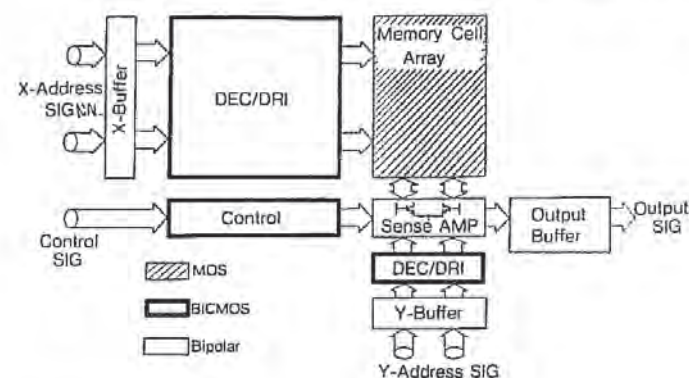


Fig. 7-73 Block diagram of a BiCMOS SRAM.⁹³ (© 1987 IEEE).

The memory array in a BiCMOS SRAM typically consists of NMOS cells in a p -well, so the die areas of CMOS and BiCMOS SRAMs are approximately the same. The decoders, word-line drivers, and write drivers are implemented with BiCMOS logic gates. (Control-logic functions with small fan-out, however, use CMOS.) The sense amplifiers, which require high input sensitivity, use pure bipolar circuits (Fig. 7-73).¹²³ In TTL BiCMOS SRAMs, the I/O buffers are typically BiCMOS gates, while in ECL BiCMOS SRAMs, bipolar ECL gates are used for I/O buffers.

7.12.3.4 Analog/Digital Applications. BiCMOS technology can also be exploited to build ICs that can perform both analog and digital functions on the same chip (e.g., ASIC chips with arrays of analog and digital cells). A large fraction of the chip area is typically set aside for the CMOS cells (mostly logic gates) that are used to perform digital-signal-processing functions. The remainder of the chip (~15-20%) is dedicated to the analog cells that permit the chip to interface with the analog world. These analog cells include the following: I/O blocks (containing resistors and nnp transistors); cells containing nnp and pnp transistors for building op amps, reference voltage sources, and comparators; and cells with large, high-breakdown-voltage (e.g., 60-V) nnp transistors. The latter are used to directly drive solenoids, lamps and LEDs and to rapidly discharge external capacitors. These ASIC chips can be used to implement such circuits as printer or camera interfaces, floppy-disk data separators, video-recorder controllers, and telecommunications circuits (e.g., PCM codec filters, ISDN transceivers, and modems).

Note that since the turn-on voltage of the bipolar transistor (V_{BE}) is much more precisely controlled than the threshold voltages of MOS devices (which are sensitive to manufacturing processes and device dimensions), well-matched pairs of bipolar transistors are much more easily obtained. Such well-matched pairs allow bipolar-IC

operational amplifiers to exhibit offset voltages that are more than an order of magnitude lower than those found in MOS op amps. Thus, BiCMOS analog/digital circuits can offer the low-input offsets and high gains of bipolar-circuit components, together with high packing density and low power of CMOS. Combining such attributes permits high-speed, low-power A/D and D/A converters to be implemented.

7.12.4 Trends in BiCMOS Technology

It is generally agreed that BiCMOS gates provide a factor of 2x improvement in speed performance (at nearly equal power dissipation) over CMOS gates of the same area when a 5.0 V power supply is used. This also implies that such circuits as SRAMs, gate arrays, and microprocessors will exhibit better performance when fabricated in 0.8- μm BiCMOS technology than with 0.5- μm CMOS. As BiCMOS is scaled to the 0.5 μm regime, however, several issues must be addressed. First, the smaller MOS devices in such BiCMOS gates will require lower power-supply voltages. At voltages smaller than 5.0 V, however, the performance of the BiCMOS gate is severely degraded.¹⁴³ Second, the scaling of the bipolar devices relative to the MOS devices in the BiCMOS gates must be optimized. Finally, new BiCMOS device structures will need to be developed. Progress in all of these areas has been reported.

7.12.4.1 BiCMOS Performance at 3.3 V. The problem of reduced BiCMOS gate performance at lower voltages has been attacked from two directions. The first uses on-chip voltage regulation to operate the on-chip 0.5- μm CMOS gates at lower voltages, while allowing the BiCMOS gates and the I/O interface to remain at the standard 5.0 V levels.¹⁴⁰ The other is to replace the conventional BiCMOS and ECL gates with new logic gates capable of operating at 3.3 V or less without suffering speed performance loss.^{141,142,144} Figure 7-74a shows examples of such novel gates, i.e., a complementary-BiCMOS inverter and a BiNMOS inverter as described in reference 144. The dependence of the delay times of these gates on supply voltage is compared to those of conventional BiCMOS and CMOS inverters in Fig. 7-74b.

7.12.4.2 Scaling BiCMOS to Retain the Advantage of Speed-Performance. The optimal scaling of bipolar devices in BiCMOS gates to allow such gates to retain their relative speed-performance advantage over scaled CMOS has been studied.¹⁴⁵ It is found that bipolar transistors in BiCMOS gates can be scaled according to the same rules formulated for bipolar devices in ECL gates,^{14,146} with the exception that bipolar transistors in BiCMOS require a higher collector doping (typically greater than $5 \times 10^{16} \text{ cm}^{-3}$) than those used in ECL gates (typically lower than $2 \times 10^{16} \text{ cm}^{-3}$).

7.12.4.3 New Process Advances in BiCMOS. With the scaling of devices into the submicron regimes, MOSFETs need increasing subsurface doping to control short-channel effects, while bipolar devices require greater collector doping to prevent the Kirk effect. Thus, the doping profile requirements of the well and collector regions will converge as BiCMOS circuits are scaled. This will relax some of the tradeoffs that have to be made when designing submicron-BiCMOS processes.

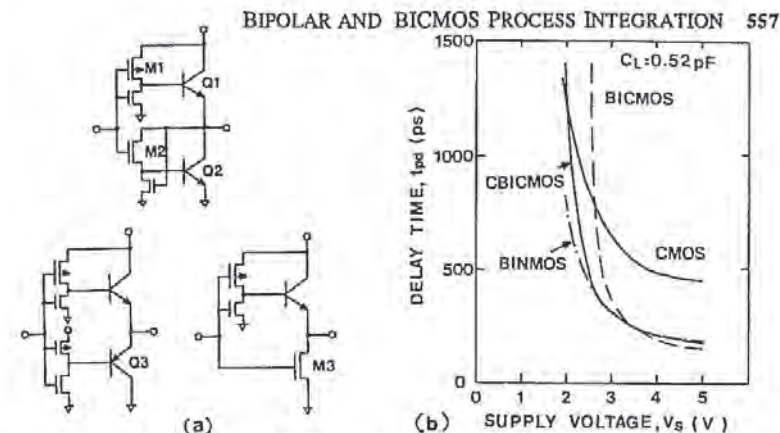


Fig. 7-74 (a) Conventional BiCMOS, CBiCMOS, and BiNMOS inverter circuits. (b) Dependence of delay time on supply voltage for various types of inverter circuits.¹⁴⁴ (Copyright 1989 IEEE).

Trench isolation will also be incorporated into more BiCMOS structures. Such trench-isolated BiCMOS processes have already been reported,¹⁰² and a cross section of one structure is shown in Fig. 7-75a.

One of the ways of making BiCMOS gates that can operate at lower supply voltages is to decrease the E-B and C-B junction capacitances of the bipolar devices. This will allow the smaller drain currents provided by the MOS devices driven with lower gate voltages to still charge these capacitances in a short time. The E-B and C-B junction capacitances can be decreased by using polysilicon-self aligned bipolar devices, as described in section 7.9. Several BiCMOS processes incorporating PSA bipolar devices have been reported (Fig. 7-75b).^{144, 147}

7.12.4.4 Analog-Digital BiCMOS Trends. In analog-digital BiCMOS circuits, the push to smaller feature sizes in the digital sections of the chip will call for shallow heavily-doped twin wells and very thin gate oxides. These conflict with the demands of higher operating voltages of the analog sections. Thus, separate wells and dual gate oxides may be needed. This may lead to very high mask count processes.

7.13 COMPLEMENTARY BIPOLAR (CB) TECHNOLOGY

Up to this point we have primarily considered bipolar (and BiCMOS) ICs fabricated with only one type of transistor (*npn* type). However, it is possible to design circuits which contain both *npn* and *pnp* transistors, and ICs having both such types are called *complementary bipolar (CB)* circuits. CB circuits have advantages for both digital and analog applications.

Digital CB circuits offer the potential of providing the high-speed performance of bipolar logic gates, but which are analogous to CMOS in that they dissipate very little

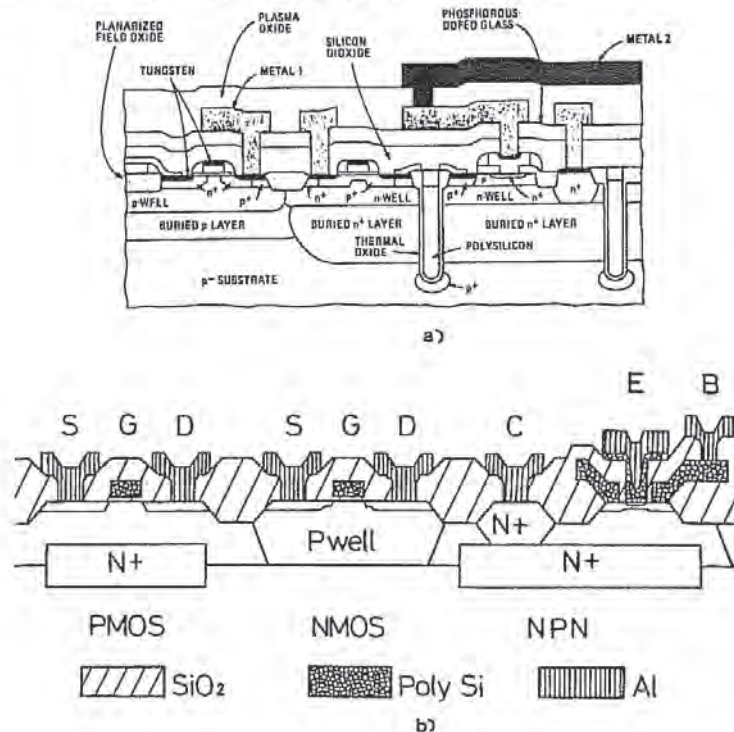


Fig. 7-75 (a) Cross section of a 0.8-μm BiCMOS process using trench-isolated bipolar devices.¹⁰² (Copyright 1987 IEEE). (b) Cross section of a BiCMOS structure incorporating a PSA bipolar device.¹⁴⁷ (Copyright 1989, IEEE).

standby power. (An earlier bipolar logic family, *integrated injection logic* [I²L], merged both the *pnp* and the *nnp* transistors into a single structure, but used the *pnp* transistor only as a load device). Since both the *pnp* and the *nnp* are involved in pull-up or pull-down during switching, the circuit performance is very much limited by the slower of the transistors.

Analog *CB* circuits would be able to improve upon the following limitations of an *nnp*-only, high performance digital bipolar technology: the lack of high-speed, substrate-isolated *pnp* transistors, the lack of capacitors and low and high sheet resistance diffused resistors, and the inability to provide precision matching of components.

Because of these deficiencies, if such analog circuits as operational amplifiers (Op Amp) are fabricated with an *nnp*-only bipolar technology, they exhibit severely limited

frequency response behavior. If a high-performance *CB* process is used, the level shift and gain of the output stage of the Op Amp can be achieved with a *pnp* transistor (the bandwidth of the Op Amp is typically limited by the output stage), and significantly improves the amplifier performance. A host of other linear circuits, such as voltage regulators, phase-locked loop circuits and D/A converters also benefit from being fabricated with a *CB* process.

Until recently, the major drawback of implementing *CB* technology has been that they have been "lopsided" with slow *pnp* devices. The conventional lateral *pnp* devices that can be integrated as "free" devices in an otherwise *nnp* technology (Fig. 7-76a) have very much lower f_T 's than the *nnp* transistors (e.g., 5-10 MHz), due to the large spacing that must be used to prevent emitter-collector punchthrough. Several other *pnp* structures have therefore been explored.

The first group of these are also lateral devices. Of them, the *epi-base lateral pnp* device (Fig. 7-76b) is available "free" in the self-aligned bipolar process.¹⁴⁸ At 2.2 μm, typical values of f_T are about 120 MHz. The *double-diffused lateral pnp* device (Fig. 7-76c) has a base whose width is given by the difference of the diffused emitter and base lateral-junction "depths." The double-diffused lateral *pnp* exhibits better performance than the *epi-base lateral pnp*, but is more complex to fabricate. Lateral *pnp* devices based in the *SICOS* structure described in section 7.9.3 have the smallest parasitic capacitance

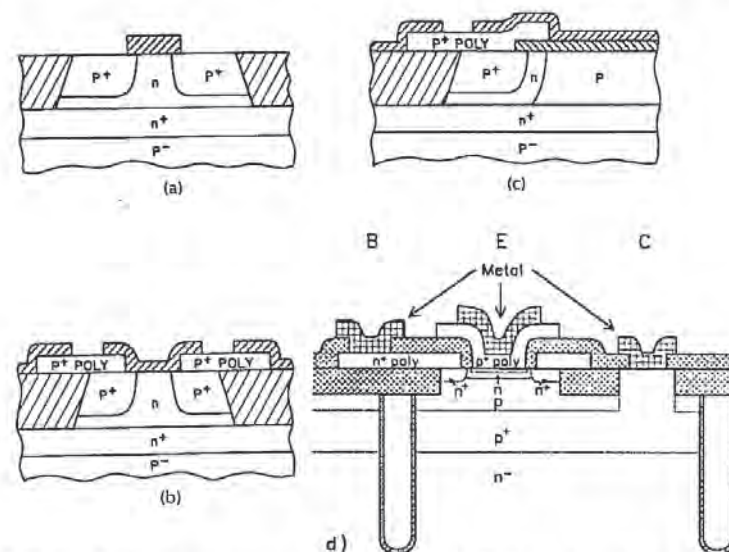


Fig. 7-76 (a) Conventional lateral *pnp* transistor. (b) *Epi-base lateral pnp* with a polysilicon emitter contact. (c) Double-diffused lateral *pnp* transistor. (d) Cross section of the 27-GHz vertical *pnp* transistor.¹⁴⁹ (Copyright 1989, IEEE).

and parasitic current component, and hence, the best current gain and performance among the lateral structures.

The second group of *pnps* are vertical devices. Although some have been built as non-self-aligned devices, the self-aligned vertical *pnp* transistors have the potential of providing the highest performance, but at the price of requiring the most complex process.

The real challenge has been to integrate vertical *pnp* devices into the high performance *npn*-only processes without significantly degrading the performance of the *nnp* devices. Recently, a 27-GHz 20 ps double-polysilicon self-aligned vertical *pnp* transistor has been announced. It has features that allow it to be integrated into a more general CB process (Fig. 7-76d).¹⁴⁹ That is, these devices have 80-nm-wide implanted bases, 0.5 μm -wide emitters, optimized emitter and collector dopant profiles, and are fabricated on a thin *p*-type epi layer over a buried *p*⁺ layer to provide high current-driving capability. A deep polysilicon-filled trench and field oxide provide device isolation.

AT&T has developed several CB processes for 5-V analog and mixed analog-digital applications. The most recent of these are the CBIC-V and CBIC-U (complementary bipolar IC) processes. In them, *pnp* transistors are isolated from the common *p*-substrate by means of junction isolation. The more-advanced CBIC-V process features vertical *pnps* with *f_T*s of 5 GHz and *npns* with *f_T*s of 13 GHz.¹⁵⁰ The process uses two layers of metal and selective epitaxial growth. Its predecessor, CBIC-U has *npns* operating at 4.5 GHz and *pnps* with an *f_T* of 3.75 GHz.

Another analog CB process designed for higher operating voltages (i.e., $\text{BV}_{\text{CBO}} = 36$ V) has been developed by Analog Devices. Its *nnp* and *pnp* transistors have *f_T*s of 600 MHz. For applications which require ultra-precise Op Amps that operate faster and produce wider bandwidth, dielectrically isolated CB processes have been developed.

REFERENCES

1. R. S. Muller and T. I. Kamins, *Device Electronics for Integrated Circuits*, 2nd Ed., New York: John Wiley & Sons, 1986.
2. E. S. Yang, *Microelectronic Devices*, New York: McGraw-Hill, 1988.
3. D. L. Pulfrey and N. G. Tarr, *Introduction to Microelectronic Devices*, Englewood Cliffs, N.J., Prentice-Hall, 1989.
4. A. R. Alvarez, Ed., *BiCMOS Technology and Applications*, Norwell, MA, Kluwer Academic Publishers, 1989.
5. D. A. Hodges and H. G. Jackson, *Analysis and Design of Digital Integrated Circuits*, New York, McGraw-Hill, 1983.
6. J. Lohstroh, "Devices and Circuits for Bipolar (VLSI)," *Proc. of the IEEE*, vol. 69, no. 7, July, 1981, p. 812.
7. T. -Z. Chen et al., *IEEE Electron Dev. Lett.*, August 1989, p. 364.
8. S. Konaka et al., *Ext. Abs. 16th Int. Conf. on Solid-State Devices and Materials*, p. 2019, 1984.
9. J. Woo and J. Plummer, *IEEE Trans. Electron Dev.*, August 1988, p. 1311.
10. P. R. Gray and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, New York, John Wiley and Sons, 1984, Eqs. 1-91, 1.113, 1.114, and 3.9.
11. A. Grove, *Physics and Technology of Semiconductor Devices*, New York: John Wiley & Sons, 1967, pp. 230-234.
12. T. Ikeda et al., *IEEE Trans. Electron Dev.*, June 1987, p. 1304.
13. P. -F. Lu and T. -C. Chen, *IEEE Trans Electron Dev.*, June 1989, p. 1182.
14. P. M. Solomon and D. D. Tang, "Bipolar Circuit Scaling," *Abs. IEEE Intl. Solid-State Ckts. Conf.*, 1979, p. 86-87.
15. A. R. Alvarez et al., *IEEE Trans. Computer-Aided Design*, vol. CAD-7, No. 2, p. 272, February 1988.
16. J. M. Stork, *Tech. Dig. IEDM*, 1988, p. 550.
17. J. Buie, "Improved Triple Diffusion means Densest ICs Yet," *Electronics*, August 1975, p. 101.
18. D. Fuoss and T. Yuzuriha, *Dig. of IEEE Custom IC Conf.*, Portland, OR, 1987, p. 348.
19. D. Fuoss, *Ext. Abs. of Electrochem. Soc. Meeting*, Spring, 1989, p. 291.
20. L. C. Parrillo et al., *Tech. Dig. IEDM*, 1977, p. 265A.
21. R. S. Payne et al., *IEEE Trans. Electron Devices*, ED-21, 273, (1974).
22. *ibid.*, Reference 1, chapter 2.
23. C. T. Kirk, *IRE Trans. Electron Devices*, ED-9, 164 (1962).
24. C. C. Allen, L. H. Clevenger, and D. C. Gupta, *Journal Electrochem. Soc.*, May 1966, p. 508.
25. J. Borland and J. Hann, *Ext. Abs. of Electrochem. Soc. Meeting*, Spring, 1989, p. 218.
26. T. H. Ning and D. D. Tang, "Bipolar Trends," *Proceedings of the IEEE*, December 1986, p. 1669.
27. T. H. Ning and R. D. Isaac, *IEEE Trans. Electron Devices*, ED-27, 1980, p. 2051.
28. T. H. Ning and R. D. Isaac, *Tech. Dig. IEDM*, 1979, p. 473.
29. T. H. Ning, D. D. Tang, and P. M. Solomon, *Tech. Dig. IEDM*, 1980, p. 61.
30. A. Tahara et al., *IEEE Bipolar Circuits and Technology Meeting*, 1989, p. 169.
31. L. Parrillo et al., *IEEE Trans. Electron Dev.*, ED-28, p. 1508, (1981).
32. H. Schaber and T. F. Meister, "Technology and Physics of Polysilicon Emitters," *IEEE Bipolar Circuits and Technology Meeting*, 1989, p. 75.
33. P. Ashburn, "Polysilicon Emitter Technology," *IEEE Bipolar Circuits and Technology Meeting*, 1989, p. 90.
34. P. Ashburn and B. Soerowirjo, *IEEE Trans. Electron Dev.*, ED-31, 1984, p. 853.
35. C. C. Ng and E. S. Yang, *Tech. Dig. IEDM*, 1986, p. 32.
36. T. F. Meister et al., *IEEE Bipolar Circuits and Technology Meeting*, 1989, p. 86.
37. B. Landau et al., *IEEE Bipolar Circuits and Technology Meeting*, 1988, p. 117.
38. P. Ashburn, D. Roulston and C. R. Sevakumar, "Comparison of Experimental and Computed Results on Arsenic- and Phosphorus-Doped Polysilicon Emitter Bipolar Transistors," *IEEE Trans. Electron Dev.*, June 1987, p. 1346.
39. G. L. Patton, J. C. Bravman, and J. D. Plummer, "Physics, Technology, and Modeling of Polysilicon Emitter Contacts for VLSI Devices", *IEEE Trans. Electron Dev.*, November, 1986, p. 1754.

40. J. Stork and J. Cressier, "The Impact of Non-Ohmic Polysilicon Emitter Resistance on Bipolar Transistor Performance," *Symp. VLSI Technol., Dig. Tech. Papers*, p. 47, 1986.
41. S. -Y. Yung and D. E. Burk, *IEEE Trans. Electron Dev.*, September 1988, p. 1494.
42. G. R. Wolstenholme et al., "An investigation of the thermal stability of the interfacial oxide in polysilicon emitter bipolar transistors by comparing device results with high resolution electron microscopy observations," *Int. Appl. Phys.*, vol. 61, p. 225, (1987).
43. J. E. Brighton et al., *Proc. IEEE Bipolar Circuits and Technology Mtg.*, 1989, p. 121.
44. Y. Niitsu et al., *IEEE Bipolar Circuits and Technology Meeting*, 1989, p. 98.
45. R. Bagri et al., *IEEE Bipolar Circuits and Technology Meeting*, 1989, p. 63.
46. Y. Yamamoto et al., *IEEE Trans. Electron Dev.*, July 1985, p. 1231.
47. J. E. Lary and R. L. Anderson, "Effective Base Resistance in Bipolar Transistors," *IEEE Trans. Electron Dev.*, November 1985, p. 2503.
48. A. K. Kapoor, M. E. Thomas, and M. B. Vora, *IEEE Trans. Electron Dev.*, June 1986, p. 772.
49. P. Asbeck, "Heterojunction Bipolar Transistors: Status and Directions," *IEEE Bipolar Circuits and Technology Meeting*, 1989, p. 65.
50. S. L. Wright, H. Kroemer, and M. Inada, "Molecular beam epitaxy growth of GaP on Si," *J. Appl. Phys.*, vol. 55, p. 2916, April 1984.
51. Y. H. Kwark and R. M. Swanson, "N-type SIPOS and poly-silicon emitters," *Solid State Electronics*, vol. 30, p. 1121, November 1987.
52. M. Takahashi, M. Tabe, and Y. Sakakibara, *IEEE Electron Dev. Lett.*, October 1987, p. 475.
53. T. Sugii et al., *IEEE Electron Dev. Lett.*, February 1988, p. 87.
54. H. Fujioka et al., *Tech. Dig. IEDM*, 1988, p. 574.
55. S. S. Iyer et al., *IEEE Trans. Electron Dev.*, October 1989, p. 2043.
56. B. S. Meyerson, "Low-temperature silicon epitaxy by ultrahigh vacuum chemical vapor deposition," *Appl. Phys. Lett.*, vol. 48, no. 12, p. 797 (1986).
57. J. F. Gibbons, C. M. Gronet, and K. E. Williams, "Limited Reaction Processing: Silicon Epitaxy," *Appl. Phys. Lett.*, vol. 47, p. 721 (1985).
58. T. Won and H. Morkoc, *IEEE Electron Dev. Lett.*, vol. 10, no. 1, p. 33, 1989.
59. D. D. Tang et al., *IEEE J. Solid-State Circuits*, SC-15, p. 444, (1980).
60. T. -C. Chen et al., "A Submicron High-Performance Bipolar Technology," *IEEE Electron Dev. Lett.*, August, 1989, p. 364.
61. G. P. Rosseel et al., "Delay Analysis for BiCMOS Drivers," *Proc. IEEE Bipolar Circuits and Technology Mtg.*, 1988, p. 220.
62. C. -T. Chuang, G. P. Li, and T. H. Ning, *IEEE Electron Dev. Lett.*, July 1987, p. 321.
63. T. Yuzuhira, T. Yamaguchi, and J. Lee., *Tech. Dig. IEDM*, 1988, p. 748.
64. H. Schaber et al., *Tech. Dig. IEDM*, 1987, p. 170.
65. G. P. Li et al., *IEEE Trans. Electron Dev.*, November 1988, p. 1942.
66. J. N. Burghartz et al., *IEEE Electron Dev. Lett.*, May 1988, p. 259.
67. H. Takemura et al., *Tech. Dig. IEDM*, 1987, p. 375.
68. T. H. Ning and D. D. Tang, *IEEE Trans. Electron Dev.*, ED-31, p. 409, 1984.
69. T. Tahiro et al., *Tech. Dig. IEDM*, 1984, p. 686.
70. H. K. Park et al., *IEEE Electron Dev. Lett.*, December 1986, p. 658.

71. G. P. Li et al., *Tech. Dig. IEDM*, 1987, p. 174.
72. C. -T. Chuang, *Tech. Dig. IEDM*, 1987, p. 178.
73. D. P. Verret and J. E. Brighton, "Two-Dimensional Effects in the Bipolar Transistor," *IEEE Trans. Electron Dev.*, November 1987, p. 2297.
74. C. T. Chuang, D. D. Tang, and E. Hackbarth, *IEEE Trans. Electron Dev.*, July 1987, p. 1519.
75. S. Sawada, *Proc. IEEE Bipolar Circuits and Device Mtg.*, 1988, p. 206.
76. T. Gomi et al., *Tech. Dig. IEDM*, 1988, p. 744.
77. T. -C. Chen et al., *Tech. Dig. IEDM*, 1988, p. 740.
78. A. Cuthbertson and P. Ashburn, *IEEE Trans. Electron Dev.*, ED-32, p. 242 (1985).
79. J. L. de Jong et al., *Proc. IEEE Bipolar Circuits and Device Mtg.*, 1988, p. 202.
80. B. van Schravendijk et al., *Proc. IEEE Bipolar Circuits and Device Mtg.*, 1988, p. 202.
81. B. Y. Hwang et al., *Proc. IEEE Bipolar Circuits and Device Mtg.*, 1988, p. 28.
82. T. I. Kamins, *IEEE Electron Dev. Lett.*, September 1989, p. 401.
83. T. Nakamura et al., *IEEE Trans. Electron Dev.*, vol. ED-29, p. 596 (1982).
84. T. Nakamura et al., *Tech. Dig. IEDM*, 1986, p. 472.
85. T. Yamaguchi and T. H. Yuzuriha, *IEEE Trans. Electron Dev.*, May 1989, p. 890.
86. A. Tuhara et al., *Proc. Bipolar Cts. and Technol. Conf.*, 1989, p. 169.
87. T. -C. Chen et al., *IEEE Trans. Electron Dev.*, ED-35, August 1988, p. 1322.
88. T. Yamaguchi and T. H. Yuzuriha, *IEEE Trans. Electron Dev.*, May 1989, p. 890.
89. M. P. Brassington et al., *IEEE Trans. Electron Dev.*, ED-36, April 1989, p. 712.
90. G. P. Rosseel and R. W. Dutton, *IEEE J. Solid-State Circuits*, February, 1989, p. 90.
91. A. R. Alvarez, "BiCMOS Technology," 1987 IEDM Short Course on BiCMOS Technol., December 1987.
92. A. R. Alvarez, "Introduction to BiCMOS," Chap. 1, p. 1, in *BiCMOS Technology and Applications*, A. R. Alvarez, Ed., Kluwer Academic Publishers, Norwell, MA., 1989.
93. K. Miyata, "BiCMOS Technology Overview," 1987 IEDM Short Course on BiCMOS Technology, December 1987.
94. K. Shibayama and Y. Akasaka, "Laboratory and Factory Automation for ULSI Developments and Mass Production," *Tech. Dig. IEDM*, 1988, p. 736.
95. R. A. Chapman et al., "0.5- μ m CMOS for High Performance at 3.3 V," *Tech. Dig. IEDM*, 1988, p. 52.
96. A. Alvarez and D. W. Schucker, "BiCMOS Technology for Semi-Custom Integrated Circuits," *Proceedings Custom Integ. Cir. Conf.*, 1988, p. 22.1.1.
97. H. Klose et al., *IEEE Bipolar Circuits and Technology Meeting*, 1989, p. 98.
98. A. R. Alvarez et al., *Semiconductor International*, April 1989, p. 226.
99. R. Richards, "Fujitsu's BiCMOS Process," *Semiconductor Internat.*, June 1989, p. 104.
100. T. Ikeda et al., "High Speed BiCMOS VLSI Technology," *Tech. Dig. IEDM*, 1986, p. 408.
101. S. Cosentino, "Motorola's BiCMOS Process," *Semiconductor Internat.*, June 1989, p. 102.
102. R. H. Havemann et al., "An 0.8- μ m BiCMOS SRAM Technology," *Tech. Dig. IEDM*, 1987, p. 841.

103. T. Ikeda et al., *Proc. 1st. Int. Symp. on BiCMOS Technologies and Circuits*, Spring Meeting, Electrochem. Soc., May 1987.
104. T. Ikeda et al., *IEEE Trans. Electron Dev.*, June 1987, p. 1304.
105. R.A. Haken et al., "BiCMOS Process Technology," in *BiCMOS Technology and Applications*, A. R. Alvarez, Ed., Norwell, MA, Kluwer Academic, 1989.
106. R. A. Chapman et al., "Submicron BiCMOS Well Design for Optimum Circuit Performance," *Tech. Dig. IEDM*, 1988, p. 756.
107. R. A. Haken et al., *Semiconductor Internatl.*, June 1989, p. 96.
108. R. A. Chapman et al., *Tech. Dig. IEDM*, 1987, p. 362.
109. S. B. Kulkarni and A. A. Kozul, *Extended Abstr. Electrochem. Soc.*, 80-2, 1351 (1980).
110. J. Borland et al., "Silicon epitaxial growth for advanced device structures," *Solid-State Technology*, January 1988, p. 111.
111. J. Borland and J. Hann, *Extended Abstr. Electrochem. Soc. Meeting*, Spring 1989, p. 21.
112. H. Iwai et al., *Tech. Dig. IEDM*, 1987, p. 28.
113. T. -I. Liou and C.-S. Teng., *IEEE Trans. Electron Dev.*, September 1989, p. 1620.
114. M. Nanba et al., *IEEE Trans. Electron Dev.*, July 1988, p. 1021.
115. D. de Lang et al., *Proc. IEEE Bipolar Circuits and Tech. Conf.*, 1989, p. 190.
116. S. Weber, "TI Soups up LinCMOS Process with 20-V Bipolar Transistors," *Electronics*, February 4, 1988, p. 59.
117. R. A. Haken et al., "BiCMOS Process Technology," in *BiCMOS Technology and Applications*, A. R. Alvarez, Ed., Norwell, MA, Kluwer Academic, Chap. 3, p. 63.
118. C. Anagnostopoulos et al., *Tech. Dig. IEDM*, 1984, p. 588.
119. S. -C. Lee, D. W. Schuecker, and P. T. Hickman, *VLSI Design*, August 1984, p. 98.
120. J. McDonald et al., *Proc. IEEE Custom Integ. Circ. Conf.*, 1988, p. 13.4.1.
121. K. Deierling, "Digital Design," Chap. 5, in *BiCMOS Technology and Applications*, A. R. Alvarez, Ed., Norwell, MA, Kluwer Academic, Chap. 3, p. 63.
122. B. C. Cole, "AMCC's BiCMOS Array Hits Record Gate Utilization," *Electronics*, February 4, 1988, p. 65.
123. M. Kubo et al., "Perspective on BiCMOS VLSI's," *J. Solid-State Ckts.*, February 1988, p. 8.
124. M. R. Goulding et al., "Low vs High Temperature Epitaxial Growth," *Semiconductor International*, May 1988, p. 90.
125. C. C. Allen, L. H. Clevenger, and D. C. Gupta, *J. Electrochem. Soc.*, 113, 508 (May 1966).
126. H. Camenzind, *Electronic Integrated Systems Design*. Copyright 1972 by Litton Educational Publishing, Inc., New York.
127. W. C. Till and J. T. Luxton, *Integrated Circuits: materials, devices, and fabrication*, Prentice-Hall, Englewood Cliffs, N. J., 1982.
128. K. H. Nicholas, *Solid-State Electron.*, 9, 35, (1966).
129. D. Rice, "Isoplanar-S Scales Down for New Heights in Performance," *Electronics*, 52, 137 (1979).

130. S. M. Sze, *Semiconductor Devices - Physics and Technology*, John Wiley & Sons, New York, 1985.
131. E. F. Labuda and J. T. Clemens, "Integrated Circuit Technology," in R. E. Kirk and D. F. Othmer, Eds., *Encyclopedia of Chemical Technology*, John Wiley & Sons, New York, 1980.
132. R. People and J. C. Bean, *Appl. Phys. Lett.* 48, 538. (1986).
133. A. Wieder, *Tech. Dig. IEDM*, 1986, p. 8.
134. D. D. Tang et al., *IEEE Electron Dev. Lett.*, Vol. EDL-8, April 1987, p. 174.
135. A. R. Alvarez et al., *Semiconductor International*, May 1989, p. 226.
136. J. N. Burghartz et al., *Tech. Dig. IEDM*, 1989, p. 229.
137. M. Sugiyama et al., *Tech. Dig. IEDM*, 1989, p. 221.
138. J. van der Velde et al., *Tech. Dig. IEDM*, 1989, p. 233.
139. C. T. Chuang and P. F. Lu, *Tech. Dig. IEDM*, 1989, p. 799.
140. H. Fukuda et al., *Dig. Tech. Papers IEEE Int. Solid-State Circuits Conf.*, 1989, p. 176.
141. W. Heimsch et al., *Dig. Tech. Papers IEEE Int. Solid-State Circuits Conf.*, 1989, p. 112.
142. Y. Nishio et al., *Dig. Tech. Papers IEEE Int. Solid-State Circuits Conf.*, 1989, p. 117.
143. H. Momose et al., *Symp. VLSI Tech.*, 1989, p. 55.
144. A. Watanabe et al., *Tech. Dig. IEDM*, 1989, p. 429.
145. G. P. Rosscel and R. W. Dutton, *Tech. Dig. IEDM*, 1989, p. 795.
146. T. Ning et al., "Scaling Properties of Bipolar Transistors," *Tech. Dig. IEDM*, 1980, p. 550.
147. T. Yoshimura et al., *Tech. Dig. IEDM*, 1989, p. 241.
148. D. Tang et al., "73 ps bipolar ECL circuits," *Tech. Dig. ISSCC*, 1986.
149. J. Warmock et al., *Tech. Dig. IEDM*, 1989, p. 903.
150. A. Feynson et al., *Proc. IEEE Bipolar Circuits and Tech. Conf.*, 1989, p. 173.
151. D. J. Roulston, *Bipolar Semiconductor Devices*, McGraw-Hill, New York, 1990.

PROBLEMS

7.1 A Si n^+ - p - n transistor has impurity concentrations of 5×10^{19} , 10^{17} , and $5 \times 10^{15} \text{ cm}^{-3}$ in the emitter, base, and collector, respectively. The base width, W_B , is $1.0 \mu\text{m}$, and the device cross-sectional area is 3 mm^2 . When the emitter-base junction is forward-biased to 0.5 V and the base-collector junction is reverse-biased to 5 V, calculate: (a) the neutral base width, (b) the minority carrier concentration at the emitter-base junction, and (c) the minority carrier charge in the base region. (Note, to solve (b) and (c) it may be necessary to consult a device physics text.)

7.2 Calculate the percentage of the total device area that consists of the active area of the four transistor structures shown in Fig. 7-14.

7.3 Explain why deep n^+ collector diffusions are often used in digital $n\text{pn}$ transistors.

7.4 Consider using an n -type diffusion into silicon as an alternative to growing an epitaxial layer. Why are epitaxial layers used instead?

7.5 Explain each of the following nonidealities in a bipolar transistor: (a) Early effect, (b) transistor saturation at low voltage, (c) breakdown at high voltage, (d) falloff of β at high

currents. Suggest at least one process parameter that can be manipulated to reduce each of these effects.

7.6 Calculate the minimum design spacings for the following bipolar process that is to be used for linear circuits operating with a ± 15 V power supply.

Substrate: p -type, $10\ \Omega\text{-cm} \pm 20\%$, (111) orientation

Buried layer: $40\ \Omega/\text{sq}$, $6\ \mu\text{m}$ deep.

Epitaxial layer: $5\text{-}\Omega\text{-cm} \pm 10\%$, $10 \pm 1\text{-}\mu\text{m}$ thick

Base diffusion: p -type boron diffusion, nominally $200\text{-}\Omega/\text{sq}$, $4 \pm 0.5\ \mu\text{m}$ deep.

Emitter diffusion: n -type phosphorus diffusion, with n^+ surface concentration
base width is nominally $1\ \mu\text{m}$.

Isolation diffusion: p^+ boron diffusion, nominally $5\text{-}\Omega/\text{sq}$.

7.7 Using the cross-section in Fig. 7-14b as a guide, estimate the percentage reduction that can be achieved in the linear dimension of a bipolar transistor by the following approaches: (a) Replacement of junction isolation with oxide isolation (allowing $0.5\ t_{\text{epi}}$ per side, for lateral etching and oxidation), (b) Using washed emitter contacts instead of patterned and etched contacts, (c) Implantation of the base and emitter, instead of diffusing them. Assume that lateral straggle during implant is equal to 50% of the junction depths.

7.8 Consider an npn bipolar transistor with a $0.2\ \mu\text{m}$ wide uniformly doped intrinsic base (excluding the emitter- and collector- base space charge layers). Under maximum applied collector to emitter voltage, $5\ \text{V}$ will be applied across the active base (excluding built-in potentials and the voltage dropped in the collector). (a) Calculate the minimum doping level in the base to prevent collector-emitter punchthrough. (b) Calculate the base Gummel number. (c) Assuming that the emitter is degenerately doped and the collector doping is $1 \times 10^{16}\ \text{cm}^{-3}$, calculate the total base charge (including emitter-base and collector-base space-charge layers) required to produce the desired Gummel number.

7.9 What is the maximum collector-base breakdown voltage of a bipolar transistor with $X_{\text{BL}} - X_{\text{BC}} = 5\ \mu\text{m}$? What range of epitaxial layer dopings may be used to obtain this breakdown voltage?

7.10 If the collector doping density in a transistor is $6 \times 10^{15}\ \text{atoms/cm}^3$, and is much less than the base doping, find BV_{CEO} for $\beta = 200$ and $n = 4$. Use $E_{\text{crit}} = 3 \times 10^5\ \text{V/cm}$.

CHAPTER 8

SEMICONDUCTOR MEMORY

PROCESS INTEGRATION

Memories store digital information (or data) in terms of *bits*, or binary digits (ones or zeros). Modern digital systems use *memory devices* to store and retrieve large quantities of digital data at electronic speeds. Early digital computers used magnetic-cores as the devices in fast-access memories. With the introduction of semiconductor memory chips in late 1960s, however, magnetic cores began to be replaced by integrated circuits (which implemented a much higher-density digital-memory function). This not only increased the performance capabilities of the memory, but also drastically decreased its cost. By the end of the 1970s, magnetic-core memories had been completely displaced as high-speed memory devices.

8.1 TERMINOLOGY OF SEMICONDUCTOR MEMORIES

Memory capacities in digital systems are usually expressed in terms of bits, since a separate storage device or circuit is used to store each bit of data. Each storage element is referred to as a *cell*. Memory capacities are also sometimes stated in terms of *bytes* (8 or 9 bits) or *words* (32 – 80 bits). Each byte typically represents an alphanumeric character. Every bit, byte or word is stored in a particular location, identified by a unique numeric address, and only a single bit, byte, or word is stored or retrieved during each cycle of memory operation.

Memory-storage capability is expressed in units of kilobits and megabits (or kilobytes and megabytes). Since memory addressing is based on binary codes, capacities that are integral powers of 2 are typically used. As a result, a memory device with a 1-kbit capacity can actually store 1,024 bits, and a 64-kbit device can store 65,536 bits.

In digital computers, the number of memory bits is usually 100 to 1000 times greater than the number of logic gates, which implies that the memory cost per bit must be kept very low. In addition, it is desirable for the memory devices to be as small as possible (since this will allow the highest density of cells on a chip), to operate at a high speed, to have a small power consumption, and to operate reliably.

APPENDIX A

IC RESISTOR FABRICATION

Resistors in ICs are implemented using either diffused regions in the silicon substrate, or thin-films deposited on the wafer surface. In either case, it is useful to describe their resistance value in terms of the *sheet resistance* parameter, R_s .

Sheet Resistance Definition

Assume that a rectangular block of conducting material with a uniform resistivity (as a function of its depth) is given as shown in Fig. 1. The resistance, R , of this structure is calculated from

$$R = \rho L / A \quad (1)$$

where L and A represent the length and cross-sectional area of the block, respectively, and ρ is the resistivity of the material. (Note that if the resistivity of the block is not constant with depth, ρ in Eq. 1 is considered to be the *average* resistivity of the layer. A technique for accurately calculating ρ in such non-uniformly doped layers is outlined in Muller and Kamins, *Device Electronics for Integrated Circuits*, 2nd Ed., John Wiley & Sons, 1986, p. 112. In this reference, it is pointed out that detailed numerical analysis may be necessary to obtain an accurate value of ρ in regions of doped silicon formed by diffusion, if the diffusion profile deviates from a simple Gaussian distribution.)

Using W as the width of the conducting region, and t as its thickness, the resistance of the block can be written as

$$R = (\rho/t) (L/W) = R_s (L/W) \quad (2)$$

where $R_s = (\rho/t)$ is called the *sheet resistance* of the layer of material. Note that R_s

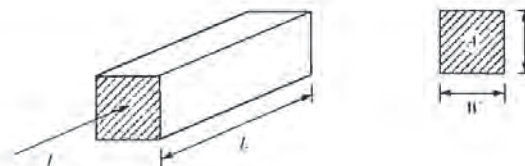


Fig. 1 Resistance of a block of material having uniform resistivity. The ratio of resistivity to thickness is called the *sheet resistance* of the material.

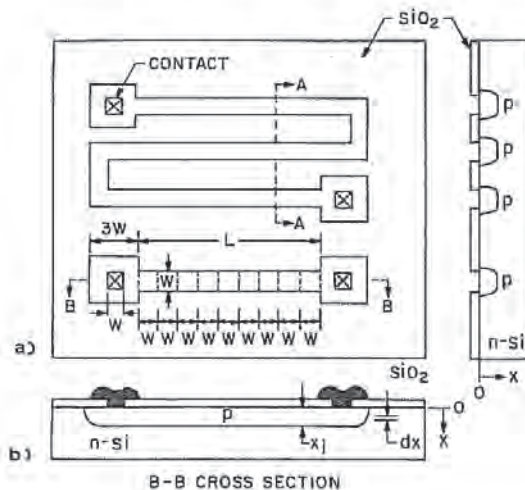


Fig. 2 (a) Top view of a zig-zag and a straight diffused resistor. (b) Side view of the straight diffused resistor.

measured for a given layer is numerically the same for any size square. It should also be noted that the unit for sheet resistance is the ohm, since the ratio of L/W is unitless. To avoid confusion between R and R_s , however, sheet resistance is given the special descriptive unit of ohms per square (Ω/sq).

If the sheet resistance value is known, a designer need only specify the length and width of the resistor to define its value. The ratio L/W can be interpreted as the number of unit squares of material in the resistor. If it is necessary to use a large number of squares to obtain a desired resistance value, a folded or zig-zag structure is normally employed (Fig. 2a).

Figure 2b depicts the side views of the straight resistor structure (in this example, a diffused IC resistor is shown). The layout of the resistor has a dumbbell shape, with the contacts made at each end. In this illustration, the resistor body is nine "squares" long. If the sheet resistance of the diffusion were $50 \Omega/\text{sq}$, the resistor body would have a resistance of 450Ω . Each end of the resistor adds approximately 0.65 squares to the resistor, and the total resistance of the resistor would therefore be approximately 515Ω . The resistance contributed by a corner square can be estimated by taking it to be approximately 65% of the straight-path value. Figure 3 gives the number of squares contributed by various end and corner configurations.

Diffused Resistors

Various diffused resistor structures can be formed in either bipolar or MOS processes. Diffused IC resistors in *npn bipolar processes* can be formed by using the shallow diffusion for the transistor base and emitter regions, or by using legally doped epitaxial collector regions. In *MOS processes*, diffused resistors are formed using the process steps that produce the source/drain regions. In *CMOS processes*, the diffused regions that form the wells can also be used to form diffused resistors.

Use of diffused resistors is popular because of their compatibility with the remainder of the rest of the planar process. That is, they can be formed at the same time as the other circuit elements, and hence do not add to the fabrication cost. To keep the diffused resistor structure isolated from the rest of the circuit, the isolation scheme is usually the same as that used for the transistors. Proper biasing must also be observed. For example, when the base region of the bipolar process is used to form resistors, the n -type region surrounding the p -diffused region must be tied the most positive potential in the circuit to keep the resultant pn junction from ever becoming forward biased. To conserve space, diffused resistors may therefore be put into a common isolation region, since only one contact to the isolation region is needed to establish the appropriate bias.

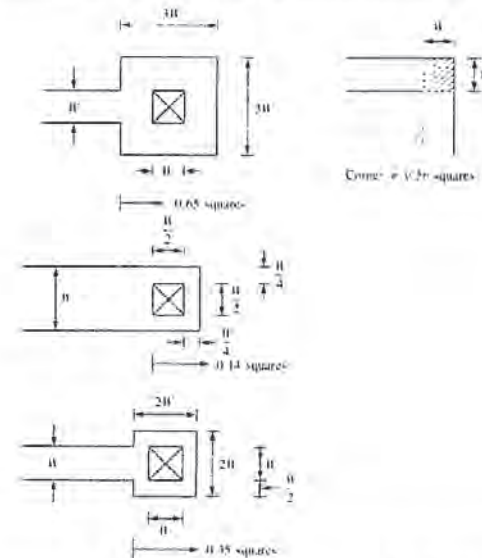


Fig. 3 Effective square contributions of various resistor end and corner configurations. From R. C. Jaeger, *Introduction to Microelectronic Fabrication*. Copyright, 1988, Addison-Wesley. Reprinted with permission.

The resistance of the structure depends on the length, width, depth of the diffusion, as well as on the resistivity of the material. For large-width resistors, the dimension W is simply determined by the mask dimensions. For very narrow resistors, however, the effective width may differ substantially from the mask dimension because impurities diffuse laterally under the oxide as well as vertically. Since minimization of the die area is important, such narrow resistors are frequently encountered. When they are used, the effective width must be determined in order to accurately estimate the resistance value.

The sheet resistance value may vary somewhat from run to run, and it is thus difficult to precisely predict the resistor values in a given circuit. However, it is easy to obtain matched resistors on a given chip, and significantly greater precision can be maintained in the ratios of paired resistors than in the resistance values themselves. For this reason, ICs are often designed so that their critical behavior depends on the ratio of two resistor values rather than on the absolute value of a specific resistor in the circuit.

In addition, since the depletion region of the isolation junction varies with voltage, diffused resistors exhibit more nonlinear I-V characteristics than do thin-film IC resistors. Finally, it should be noted that diffused resistors suffer from two additional limitations: (1) A parasitic junction capacitance is associated with the resistor and the underlying region; and (2) they exhibit a high temperature coefficient.

Bipolar IC Diffused Resistors: Base Resistors. Resistors in bipolar ICs are most commonly formed using the p -type base diffusion of an npn transistor, with a contact at each end (Fig. 4a). The resistor structure is created within a separate n -type island so that it is isolated from other devices on the chip. The resistivity of the base diffusion is usually determined during the design of the npn transistor (typically ~ 100 - to $200\text{-}\Omega/\text{sq}$). Resistors formed by the base diffusion are usually limited to a range from approximately $50\text{ }\Omega$ to $100\text{ k}\Omega$. Note that a pnp parasitic transistor exists between the resistor and the base resistor and the substrate. The n^+ buried layers, however, help to reduce the gain of this parasitic transistor.

Bipolar IC Diffused Resistors: Emitter Resistors. Diffused resistors can also be fabricated by forming an n -type emitter diffusion entirely within a p -type base region, and then providing contacts at both ends. Since the emitter region has a higher doping concentration than the base, the sheet resistance of emitter resistors will be smaller (~ 2 - to $10\text{-}\Omega/\text{sq}$). These resistors can be used to advantage where very low resistance values are required.

Bipolar IC Diffused Resistors: Epitaxial Resistors. By making contact to the lightly doped epitaxial layer, larger resistor values can be obtained than with the base or emitter resistors ($\sim 1000\text{-}\Omega/\text{sq}$ — the buried layer, however, must be omitted in order to obtain such high resistance values.)

Bipolar IC Diffused Pinch Resistors. For even higher values of sheet resistance, a double-diffused structure, known as *pinched resistor*, is sometimes used. In pinched resistors, the n^+ -region is formed by the emitter over the p -base region reduces

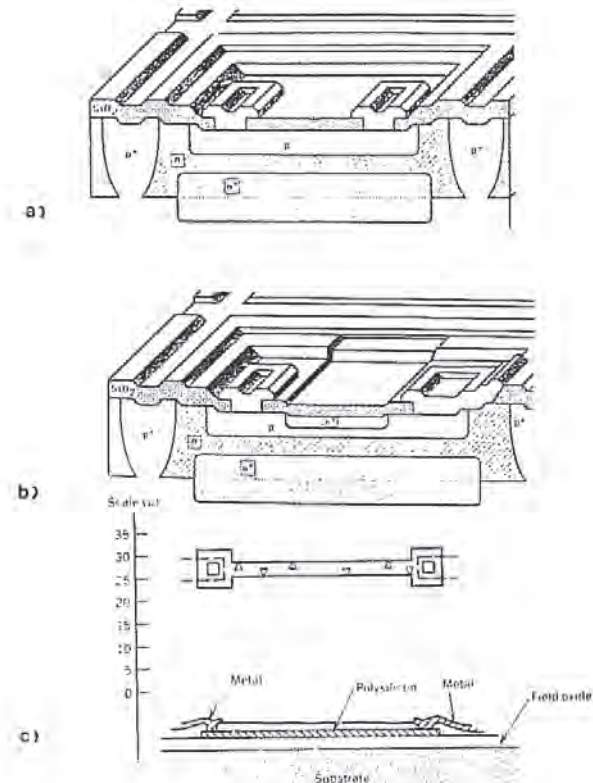


Fig. 4 (a) Base diffused resistor. (b) Base pinched resistor. (c) Thin-film polysilicon resistor. From W. Maly, *Atlas of IC Technologies*, Copyright 1987 by the Benjamin/Cummings Publishing Company. Reprinted with permission.

the vertical dimension of the diffused base resistor (Fig. 4b). This can increase the base resistor sheet resistance to 2-k- to $10\text{-k}\Omega/\text{sq}$. The epitaxial layer pinched between the p -type base and the p -type substrate can also serve as pinch resistor structure ($\sim 4\text{-k-}$ to $10\text{-k-}\Omega/\text{sq}$). The reproducibility of the sheet resistance of pinched resistors, however, is typically poor (e.g., $\pm 50\%$). Another significant drawback is that the maximum voltage that can be applied across base pinched resistors is limited to $\sim 6\text{ V}$ because of the breakdown voltage between the emitter-diffused top layer and the base diffusion. The epitaxial pinched resistors can tolerate higher applied voltages.

MOS IC Diffused Resistors: Source/Drain Resistors. The diffused layer used to form the source and drain of the n -channel and p -channel devices can be

used to form a diffused resistor. The resulting resistor structure and properties are very similar to the bipolar emitter diffused resistor structures described above.

MOS IC Diffused Resistors: Well Resistors. The well diffused region in CMOS ICs can also be used as the body of a diffused resistor. Since the well is a relatively lightly doped region, the sheet resistance is on the order of $10\text{-k}\Omega/\text{sq}$. Its properties are similar to those of the bipolar epitaxial diffused resistors described above.

Thin-Film-Deposited IC Resistors

As noted, diffused resistors have the drawbacks of high temperature coefficient, poor tolerance, and must be junction isolated. The latter implies that a parasitic capacitance exists along with each diffused resistor, and exposure to radiation causes photocurrents to flow across the reverse-biased isolating junction. These disadvantages can be reduced by using thin-film resistors deposited on the top surface of the substrate. After the resistor material is deposited, the individual resistors are patterned in the conventional manner. They are then interconnected to the rest of the circuit using the standard Al interconnect process.

Thin-film resistors exhibit smaller parasitic capacitance values than do diffused resistors, and also do not suffer nonlinear behavior from depletion effects.

Polysilicon Thin-Film Resistors. Polysilicon films are used in silicon-gate MOS processes, as well as in advanced bipolar, polysilicon self-aligned structures, and in BiCMOS technology. Such polysilicon films are often used to form resistors. The nominal sheet resistance of heavily doped polysilicon films is on the order of 20 to $80\ \Omega/\text{sq}$. The matching properties of polysilicon resistors are similar to those of diffused resistors. A cross sectional view of a polysilicon resistor is shown in Fig. 4c. High-valued, lightly doped polysilicon resistors (as described in chap. 8) are used as load devices in MOS SRAMs.

Nichrome, Tantalum, and Cermet Thin-Film IC Resistors. Three other materials have found use as thin film resistor materials in IC applications. The value of these resistors can be more reproducibly fabricated than can diffused resistors or polysilicon thin-film resistors, and they exhibit much lower temperature coefficients. These are *nichrome* (80% nickel and 20% chrome), *tantalum*, and *Cermet* (Cr-SiO) films. The sheet resistance ranges of these films are respectively 10 to $1000\ \Omega/\text{sq}$, 10 to $1000\ \Omega/\text{sq}$, and 30-2500 Ω/sq . The added cost of depositing and patterning another layer of resistive material is the chief drawback of such thin-film resistors.

MOS Devices as Resistors

The MOS transistor biased in the triode region can be used in many circuits to perform the function of a resistor. The effective sheet resistance is a function of the applied gate bias but can be effectively much higher than polysilicon or diffused resistors, allowing large amounts of resistance to be implemented in a small area.

APPENDIX B

PROPERTIES OF SILICON at 300°K

Quantity	Silicon
Atomic weight	28.09
Atoms, total (cm^{-3})	4.995×10^{22}
Crystal structure	Diamond
Density (g/cm^3)	2.33 (solid) 2.53 (liquid, 1421°C)
Density of Surface atoms (cm^{-2})	
(100)	6.78×10^{14}
(110)	9.59×10^{14}
(111)	7.83×10^{14}
Dielectric constant	11.8
Effective density of states in conduction band, N_C (cm^{-3})	3.22×10^{19}
Effective density of states in valence band, N_V (cm^{-3})	1.83×10^{19}
Energy gap (eV)	1.12
Index of refraction	3.42
Intrinsic carrier concentration (cm^{-3})	1.38×10^{10}
Intrinsic Debye length (μm)	24
Intrinsic resistivity ($\Omega\text{-cm}$)	2.3×10^5
Lattice constant ($\text{\AA}$)	5.43
Linear coefficient of thermal expansion, $\Delta L/L \Delta T$ ($^\circ\text{C}^{-1}$)	2.6×10^{-6}
Melting point ($^\circ\text{C}$)	1421
Mobility (drift) ($\text{cm}^2/\text{V s}$)	
μ_n (electrons)	1500
μ_p (holes)	475
Specific heat ($\text{J}/\text{g}\cdot^\circ\text{C}$)	0.7
Thermal conductivity ($\text{W}/\text{cm}\cdot^\circ\text{C}$)	
Solid	1.5
Liquid	4.3
Youngs modulus (kg/mm^2)	10,890

APPENDIX C

FUNDAMENTAL PHYSICAL CONSTANTS

CONSTANT	SYMBOL	SI - UNIT	DEFINITION
Angstrom unit	Å		$1\text{Å} = 10^{-1}\text{ nm} = 10^{-4}\text{ }\mu\text{m} = 10^{-10}\text{ m}$
Atomic mass	$m_u = 1\text{ u}$	kg	$m_u = 1.6605 \times 10^{-27}\text{ kg}$
Avogadro's constant	N_A	mol^{-1}	$N_A = 6.022 \times 10^{23}\text{ mol}^{-1}$
Bohr radius	a_B		0.529 Å
Boltzmann constant	k	$\text{J } ^\circ\text{K}^{-1}$	$k = 1.38066 \times 10^{-23}\text{ J } ^\circ\text{K}$
Electron rest mass	m_e	kg	$m_e = 9.11 \times 10^{-31}\text{ kg}$
Elementary charge	q	C	$1.60218 \times 10^{-19}\text{ C}$
Electron volt	eV		$1\text{ eV} = 1.60218 \times 10^{-19}\text{ J}$ $= 23.053\text{ kcal /mole}$
Molar gas constant	R	$\text{J mol}^{-1} ^\circ\text{K}^{-1}$	$R = 1.987\text{ cal /mole } ^\circ\text{K}$
Permeability in vacuum	ϵ_0		$\epsilon_0 = 8.854 \times 10^{-14}\text{ F /cm}$
Planck constant	h	J s	$h = 6.626 \times 10^{-34}\text{ J-s.}$
Proton rest mass	M_p	kg	$1.672 \times 10^{-27}\text{ kg}$
Speed of light in vacuum	c	m s^{-1}	$2.9979 \times 10^8\text{ m /s}$
Standard (atmospheric) pressure		$\text{N m}^{-2}, \text{ Pa}$	$1.013 \times 10^5\text{ Pa}$
Thermal voltage at 300° K	kT/q	V	0.0259 V

INDEX

NOTE: FOR MAJOR SUBJECT HEADINGS ALSO CHECK THE TABLE OF CONTENETS. THE LISTING ARE NOT ALWAYS DUPLICATED IN THE INDEX.

Absorption constant, resist, 702
Access time, 571, 597
 in advanced DRAMs, 650
Address transition detection, 577
Adhesion layers for W, 247-8, 282
Aerial image, 699, 700, 703
Al:Cu alloys, 191, 264, 267, 269, 271
Al-gate MOS, 416-8
AlF₃, 252, 263
Alpha particles, 615-17
Aluminum metallization, 113-4, 116-7, 191
 Al:Cu alloys (see Al:Cu)
 corrosion, 267
 CVD, 253
 hillocks in (see Hillocks)
 laser planarization of, 255-6
 nitrogen incorporation in, 261, 266
 oxygen incorporation in, 261
 stress induced voids in, 265
Al:Si alloys, 116-7
 Si precipitates in, 116
 solubility of Si in Al, 113-4
Al-Si contacts, 111-6
 contact electromigration failure in, 117
 diffusion of Si in Al films, 114-6
 junction spiking, 114-6, 502
 sintering of, 111-6
Al:Ti alloys, 191, 264, 271
Al-TiN-Ti-Si contacts, 132
Al-Ti;W-PtSi-Si contacts, 132, 495
Amorphous silicon, in single polysilicon,
 self-aligned bipolar transistors, 517
Analog bipolar transistors, 484-6
 in BiCMOS technology, 543
 process integration issues of, 545-6
Analog/digital BiCMOS process, 546
 applications, 556
Anodized Al, deposition, 273
Arsenic,
 clustering, 666
 diffusion, simulation of, 665-6
 in buried collectors, 488
 in emitters of bipolar transistors,
 495, 511
Asymmetrical implant effects,
 in LDD-MOSFETs, 359-60
 in self-aligned bipolar structures, 511-13
Avalanche breakdown, 471, 472, 488
 in bipolar transistors, 492
Average interconnect length vs. device
 density, 182
Au (see Gold)
Autodoping, 387
 in BiCMOS, 539
 simulation of, 674-6
AZ protective coating, 278-9

Base-emitter voltage, V_{BE} , 479
Base pushout (Kirk) effect, 467-8, 472,
 474, 477, 481
Base region formation in SBC transistor,
 491-2
Base transit time, τ_B , 477
Base transport factor, α_T , 464
Base width, W_B , 453, 458-9, 464, 465,
 477, 492, 493
 modulation (Early effect), 466-7
 in self-aligned bipolar structures, 511
BASIC bipolar structure, 521
BF₂ 154-5, 439
Bias sputtered quartz
 (see Bias sputtered SiO₂)
Bias-sputtered SiO₂, 199, 217-20
BICEPS process simulator, 647, 648
BiCMOS, 529-58
 analog-digital, 530, 545-51, 556
 applications, 551-6
 capital equipment cost savings, 529
 evolution of, 529-31
 high-performance digital, 536-43
 device design issues in, 536-7
 process flow, 538-43
 low-cost digital,
 triple diffused, 483, 531-4, 536
SRAMS, 554-5, 585-7
trends in, 556-7

- Bipolar PROMs, 622-3, 627
- Bipolar transistors (see also chap. 7)
 - analog device structures (see Analog bipolar transistors)
 - breakdown voltages in (see Breakdown voltage)
 - characteristic curves, 457
 - Early voltage, 492 (see also, Early voltage)
 - heterojunction, 506-10 (see also Heterojunction bipolar transistors)
 - Kirk effect (see Kirk effect)
 - IC transistor structures, 458
 - junction isolated,
 - CDI, 16-7, 482, 483
 - standard-buried collector (SBC), 482
 - characteristics, 484-6
 - doping profile of, 485, 492
 - fabrication process, 486
 - limitations of, 486-98
 - triple-diffused, 482-3,
 - logic families, 460-2
 - oxide isolated process, 482, 489-500
 - pn*p (see *pn*p transistors)
 - saturation, 457
 - self-aligned device structures (see Self-aligned bipolar structures)
 - series resistances in (see Parasitic series resistance)
 - SICOS, 520-2
 - switching speed, 478-81
 - trench isolated, 522-3
- Bipolar transistor action, 454-8
- BIPOLE, device simulator, 645, 649, 720-1
- BIPS, isolation structure, 64
- Bird's beak, 24, 331, 334, 498, 499, 691-4
- Bird's head, 29, 498, 499
- Bit line, def., 568
- capacitance, 594, 596
- Black silicon, 52-3
- Body factor, def., 311
- Boltzmann transport equation (BTE), 658, 660
- BOMOS, isolation structure, 153
- Bonding pads, 337
- Boron diffusion, 665, 667
 - in the presence of high arsenic concentrations, 513
- Boron-encroachment in LOCOS, 23
- narrow width effect, 23, 340-1
- suppression by
 - chlorine implant, 23, 433
 - high-pressure oxidation, 23, 423, 433
 - Ge co-implant 23, 433
- Borosilicate glass, (see BSG)
- BOX isolation, 45-8, 66, 522, 604
- BOXES, 47-8
- PROPS, 47
- Sidewall doping by diffusion, 47-8
- BPSG, 104-5, 195-6, 335, 396
- Breakdown voltage,
 - gate oxide, 314
 - in bipolar transistors,
 - collector-base junction, BV_{CBO} , 468, 469, 470, 484, 490, 516, 546-8
 - collector-emitter junction, BV_{CEO} , 468, 470, 471, 477, 483, 484, 485, 489, 506, 537, 547-8
 - effect of junction curvature on, 468
 - collector-substrate junction, 490
 - emitter-base junction, BV_{EB0} , 469, 493, 501, 506
 - limits in bipolar transistors, 472
 - pn* junction, 313-4
 - punchthrough voltage, 471, 492, 507
 - reachthrough, 468, 469, 489-9
- BSE DRAM cell, 605-6
- BSG, 198
- BSQ (see Bias-sputtered SiO_2)
- Buried-channel PMOS, 394
- Buried contacts, 160-2, 326, 332
- Buried layers,
 - in BiCMOS, 538-9
 - n*-type, 538
 - p*-type, 538
- in bipolar devices, 459, 469, 472-3, 487-8, 499
- formation of, 487-8
- antimony vs. arsenic, 488, 499
- Buried-storage electrode (BSE) DRAM cell, 605-6
- Butted contacts, 160-2
- CADDET, device simulator, 649, 719-21
- Capacitance,
 - bit line, in DRAMs, 594, 596
 - charge-storage, in bipolar transistors, 475
 - gate, in MOS devices, 303, 479
 - oxide, 303
 - input capacitance in CMOS, 378

- interconnect, parasitic, 184-5
 - side-by-side, 185
 - parasitic, in bipolar devices, 475
 - collector-base, C_{CB} , 475, 477, 479, 500, 520
 - collector-substrate, C_{CS} , 475, 477, 479
 - emitter-base, C_{EB} , 475, 477, 493, 500, 501, 506, 51
 - parasitic, in MOS devices, 303
 - pn* junction, 177
 - storage capacitance, in DRAMs, 595
 - wiring, C_W (see Interconnect capacitance)
 - Capacitors,
 - in analog IC processes, 544-5, 550
 - in DRAM cells,
 - planar, 590-1, 596-600
 - stacked, 610-15
 - trench, 600-9
 - process integration of IC capacitors, 5
 - CBIC-U, complementary bipolar, 560
 - CBIC-V, complementary bipolar, 560
 - Cell, memory, def., 567
 - characteristics of, 568
 - ferroelectric, 636
 - Channel length, (see Gate length)
 - Channel modulation factor, 345
 - Channel resistance, 138-9
 - Channel-stop implant, 18, 22-3, 329-31
 - in CMOS, 422-3, 433
 - dual, self-aligned, 389, 424, 433
 - in NMOS, 329
 - in oxide isolated bipolar process, 500
 - in trench isolation structures, 54
 - Channeling tail, causing wider base in self-aligned bipolar transistors, 512
 - methods to prevent, 512-3
 - Chemical-mechanical polishing,
 - in dielectric isolation, 68
 - in SEG, to remove corned facets, 60.
 - for planarizing dielectrics, 226, 238-9
 - Chlorine, Cl, 267, 433
 - implant, 23, 433
 - Cluster tool, 221
 - Clustering effects, with arsenic, 103, 666
 - CMOS (see also chap. 6)
 - epitaxial, 385-7, 423
 - p* on *p*⁺, 386
 - n* on *n*⁺, 386
 - historical evolution of, 370-3
 - inverter operation, 373-6
 - isolation technology, 419-28
 - latchup in, (see Latchup)
 - n*-well, 384-5, 428-31
 - power supply voltages in, 446-7
 - p*-well, 383-4, 428, 31
 - ratio of *p*-channel to *n*-channel
 - saturation currents, 377
 - retrograde-well, 389-92,
 - n*-well type, 390-1
 - in BiCMOS circuits, 532
 - p*-well type, 390-1
 - in BiCMOS circuits, 534
 - three dimensional, 447
 - to suppress latchup, 415-6, 427
 - twin-well, 387-8, 432-41
 - in BiCMOS circuits, 535-543,
 - process, 538-43
 - vertical punchthrough in, 384-5
 - with gate materials having symmetrical work functions, 398-9
 - with *p*-doped poly, 397-8
 - with W gate, 399
- Coefficient of thermal expansion,
 - Al, 268
 - Si, 268
- Coercive voltage, 636
- Coherently-strained epi layers, def., 508
- Collector current vs. V_{BE} in bipolar transistors, 480
- Collector diffused isolated bipolar process, 16-7, 482, 483
- used in BiCMOS circuits, 533-4
- in analog/digital BiCMOS, 546-8
- Collector resistance (see Parasitic series resistance, bipolar transistors)
- Common-base current gain, 456, 464
- Common-emitter configuration, 457, 459
- Common-emitter current gain, 456, 464-5, 491, 493, 501, 515, 527
- Complementary BiCMOS, 556, 557
- Complementary bipolar technology, 557-60
- Conductors, VLSI, requirements, 189-90
- Contact chain structure, 101
- Contact electromigration failure, 117, 127, 149, 264-5
- Contact resistance, R_{co} (see Parasitic resistances)
- Conventional metal-Si contacts, 101-121, 335
- dirty surfaces, effect of, in, 108
- etching contact openings, 103-7
- damage to Si during dry etching of,

104, 195, 335
 end-point detection
 during dry-etching of, 107
 sidewall sloping, by
 multi-step etch process, 107
 resist erosion, 106
 resist flow, 105
 ion-beam mixing to form
 intimate metal-Si interface, 111
 metal deposition, 109-10
 multilayered contacts,
 Al-Ti-W-PtSi-Si, 124, 132
 Al-TiN-Ti-Si, 132
 Mo-Ti-Si, 132
 Mo-Ti-W-Si, 132
 native oxide growth control, 108fn
 native oxide removal from Si, 107
 sintering, 110-11, 336-7
 wider than minimum width, 143-4
 Contrast-enhancement layers,
 simulation of, 705-6
 Conventional via filling (see Vias)
 Copper, 193
 Corrosion of Al lines, 267, 274
 CoSi₂, 146, 147, 150-2, 156, 163, 166
 diffusion barrier needed between
 Al and CoSi₂, 151
 formation process, 151
 in CMOS, 438
 resistivity of, 150
 silicon consumption
 during formation, 151
 CREEP, oxide growth simulator,
 648, 649, 694-6, 717
 Crevices, 207
 Critical upset charge, 616
 Current crowding, 474-5
 Cutoff frequency (see Unity-gain frequency, f_T)
 Cutoff mode of operation, 456
 CVD Al, 253
 CVD W (see Tungsten)
 Cu (see Copper)
 Cycle time, 571

 DASH, DRAM cell, 614
 Deep-collector contact,
 formation of,
 in BiCMOS,
 polysilicon plug, 540
 in SBC bipolar process, 490-1
 Defects,
 in emitter-base junction, due to
 ion implanting the emitter, 50-2

 in Si, stress induced
 due to nitride,
 in LOCOS, 22
 in fully recessed LOCOS, 30-1
 due to oxide growth in LOCOS, 30-1
 due to SWAMI isolation, 41
 in TiSi₂ contacts, stress induced, 150
 in trenches, during oxidation, 55
 in ZMR wafers, 76
 stress induced voids in Al lines, 265
 Depletion-mode MOSFETs (def.), 301, 317
 implant to create, 322, 331
 Depletion-region width in *pn* junctions, 343-4
 Depth-of-field limitations, 65, 203, 214
 DEPICT, process simulator, 648, 649, 710-2
 Device simulators, 718
 BIPOLE, 645, 649, 720-1
 CADDET, 649, 719-21
 CANDE, 649, 719
 FIELDAY, 649, 721-2
 GEMINI, 645, 649, 719
 MINIMOS, 719-20
 PISCES, 645, 649, 721-22
 SEDAN, 649, 720-21
 SIFCOD, 649, 721-2
 DIBL, 341-4
 Dielectric isolation, 67-9
 Dielectric materials for
 multilevel interconnects, Table, 195
 Diffusion barriers, 121-31
 defined, 122
 passive (def.), 123
 sacrificial (def.), 123
 stuffed (def.), 123
 experimental materials, 131
 polysilicon, 126-7
 Ti, 127-8
 TiN, 128
 Ti:W (see Ti:W)
 W, 128-31
 DIET DRAM cell, 607-8
 Diffusion coefficients of charge carriers,
 in base, 465, 477
 simulation of,
 in SUPREM III, 663-9
 in SUPRA, 685
 Diffusion overlap (def.), 95
 Digital circuits,
 BiCMOS, 531-42
 bipolar, 459-63, 476, 479-82, 486
 DM1 (def.), 189
 Double-level metal systems,
 bipolar, 276

CMOS, 277-83
 submicron, 279-83
 Doubly-ionized boron, 391, 428
 Drain engineering, 354, 356
 Drain-induced barrier lowering (see DIBL)
 DRAMs, 587-620
 access times of, 597, 600
 cell size, evolution of, 597
 CMOS, 596-7
 design and economic constraints on,
 597-600
 Hi-C cell, 596, 601, 603, 605
 leakage current of cells in, 592
 one transistor cell, 589-97
 operation of, 591-3
 plate electrode of, 591
 product cycle of, 597
 process complexity vs. DRAM size,
 599-600
 soft errors in, 615-8
 stacked capacitor cells, 610-5
 DASH cell, 614
 fin structure in, 614-5
 SCC cell, 614-4
 storage capacitance in, 595
 three-transistor cell, 588
 trench-capacitor cell, 600-9
 BSE cell, 605
 DIET cell, 607-8
 DSP cell, 607-8
 FASIC cell, 605
 folded capacitor cell, 604
 HPSAC cell, 607
 IVEC cell, 604-5
 SEOT cell, 609
 SGT cell, 609
 SPT cell, 605-6
 trench-transistor cell, 609
 TSAC cell, 608-9
 Video RAM (VARAM), 572
 DSP DRAM cell, 607-8
 Dual-dielectric storage capacitor, 598
 in trench capacitor, 601
 Dynamic logic, 587
 Dynamic RAMs (see DRAMs)

 Early effect, 466-7
 Early voltage, 466-7, 492, 508, 545
 ECL, 461-2, 478, 481
 EIAJ machine model test method, 444
 Electrically erasable PROMs (EEPROMs),
 570, 627, 628-35
 Flash, 632-5

 FLOTOX, 628-9
 MNOS, 628-9
 Textured poly, 631-2
 ELO (epitaxial lateral overgrowth), 63
 for forming extrinsic base poly regions in
 self-aligned bipolar transistors, 513
 Electroless deposition of Ni, 256-7, 260
 Electron-cyclotron-resonance (ECR)
 plasma CVD, 237-8, 273, 285
 Electromigration, 191, 264, 378
 failure in contacts, 117, 127, 149, 264-5
 Electrostatic discharge protection, 441-6
 Elevated source/drains, 157-8, 439
 Emitters,
 polysilicon, (see Polysilicon emitters)
 push-effect, 494-5
 walled, 496, 500, 522
 washed, 496
 Emitter-base junction capacitance
 (see Capacitance)
 Emitter-coupled logic (see ECL)
 Emitter-injection efficiency, 464-5, 501
 dependence on band-gap difference, 507
 Endurance, EEPROM, 631
 of Flash-EEPROMs, 635
 Enhancement-mode MOSFETs (def.), 301, 307
 Enhancement-depletion-mode process (def.),
 324
 Epitaxy
 autodoping in, 387, 488, 539, 674-6
 in BiCMOS, over dual-doped buried layers,
 539
 in analog/digital BiCMOS, 546-7
 in bipolar devices,
 doping concentration to prevent
 Kirk effect, 467-8, 488-9
 layer thickness,
 469-71, 484, 485-6, 490
 thickness minimum
 to prevent reachthrough, 468
 in heterojunction bipolar transistors,
 coherently strained, 508
 Ge-Si layers, 508
 limited reaction processing to form,
 508
 MBE-formed, 508
 psuedomorphic, 508
 ultra-high-vacuum, formed in, 508
 in high-performance BiCMOS technology,
 537,
 in BOMOS, 153
 in CMOS, 385-7, 423
 to suppress latchup in, 414-5

- in SBC bipolar process, 484-5, 488-9
- MBE, 508
- photo-enhanced, 159-60
- selective (see SEG)
- simulation of, 674-6
- Erasable (UV) PROMs (EPROMs), 570, 623-8
- Error detection and correction, 571, 631
- Etchback of dielectric
 - CVD bias etchback, 22-24
 - without sacrificial layer, 311-2, 222-9
 - with sacrificial layer, 220, 222-229
 - integrated process, with via patterning, 229
- Etchback of SOG (see SOG)
- Etchback of W (see Tungsten)
- Etching
 - selective
 - PtSi, 120,
 - CoSi₂, 150
 - TiSi₂, 148, 164, 165
 - trench, 51-4, 55
 - damage during, 55
 - W (see Tungsten)
- Extrinsic-base region, 473-4
- formation of, 486
- in polysilicon,
 - self-aligned bipolar structures, 511
 - in SBC, 493
- FABRICS, process simulator, 649, 718
- FAMOS structure, 624-8
- Fanout,
 - BiCMOS, 524, 551
- FASIC, 605
- Ferroelectric, def., 636
- Ferroelectric RAMs (FRAMs), 635-8
- FELT/SOI, isolation structure, 69
- Field-shield isolation, 63-4
- FIG, interconnect structure, 259-60
- Filled-in-grooves (see FIG)
- Fin-structure, DRAM cell, 614-5
- FIPOS, 76-6
- Fixed oxide charge, 319
- Flash EEPROMs, 632-5
- Floating polysilicon gate, 623
- FLOTOX EEPROMs, 629-31
- Fluorine,
 - impact on hot-carrier degradation, 362
 - impact on boron penetration
 - thru thin oxides, 398
- Fluorinated nitride, 362
- Folded-capacitor trench DRAM cell, 604
- Forward-active mode, def., 455
- Four-level metal process, 285-6
- Fowler-Nordheim tunneling,
 - 597, 601, 630, 632, 633
- Fully depleted SOI transistors, 74
- Functional density (def.), 181
- FUROX, isolation structure, 41-2
- Fusible links, 623
- GaP, in heterojunction
 - Si transistors, 508
- Gas-immersion laser doping (see GILD)
- Gate arrays,
 - in BiCMOS, 551
 - in CMOS, 379
- Gate bird's beak, 438
- Gate-controlled breakdown devices, 443
- Gate/drain overlap, 358
- Gated-diode structure in DRAMs, 606
- Gate-induced drain leakage current, 363
- Gate length, 315
 - effective, L_{eff} , 315-6, 326, 334
 - evolution of, 355
 - physical, L , 315
- Gate oxide
 - growth, 331, 434, 435
 - sacrificial, 27, 331, 431
 - ultrathin, novel growth method, 435
- GEMINI, device simulator, 645, 649, 719-20
- Ge-Si-base heterojunction transistors, 508-10
- GILD (gas immersion laser doping), 158-9
- Global interconnects (def.), 162-3
- Gold, 193
- Graft-base (see Linkup region)
- Gridding, 652
- Grow oxide-and-etch isolation structures, 18
- Guard rings, 371-2
 - around PtSi Schottky contacts, 497
 - to suppress latchup in CMOS, 416-9
- Gummel number, def., 465
 - base, 465, 468, 491, 492, 545
 - emitter, 465, 493
- Half- V_{CC} , in DRAM cells, 599
- Halo LDDs, 363
- Hard fails in DRAMs, 615-6
- Heterojunction bipolar transistors (HBTs), 506-10
 - coherently strained epi layers in, 508
 - Gap emitter, 508
 - Ge-Si base HBTs, 508-10

- processing of, 509-10
- microcrystalline Si emitters, 508
- OXSEF emitter, 508
- SIOPS emitter, 508
- Hi-C DRAM cell, 596, 601, 603, 617
- High-energy implanter,
 - for channel-stop, 38
 - for collector implant in advanced bipolar, 391, 415, 428
 - for PMOS punchthrough prevention, 396
- High-level injection, 468
- High-pressure oxidation (HIPOX),
 - 23, 423, 434
- Hillocks, 262, 268-73
 - suppression techniques, 271-3
- Holding current, in latchup, 403, 410-11
- Holding voltage, in latchup, 403, 410-11
- Hot-carrier effects in MOSFETs
 - (see MOSFETs and LDDs)
- Hot-electron induced punchthrough (HEIP), 362
- Hot-electron injection into gate, 624-5, 633
- HPSAC DRAM cell, 164-6
- Hydrogen
 - diffusion into polysilicon resistors, 582
 - in silicon nitride, 274
 - effect on hot electron degradation
 - in MOSFETs, 362
 - oxidation of Mo in wet H₂, 399
- Hygroscopic (def.), 215
- I_D (def.), 306, 479
- I_{Dsat} (def.), 306
- In situ doping of poly, 540, 602
- Inhibitor concentration, in resist,
 - simulation of, 699, 704-5
- Interface trap charge, 314
- Intralevel shorts, 270
- Integrated-injection logic (I²L), 558
- Interface driver circuits, BiCMOS, 554
- Interfacial oxide in polysilicon emitters, 503
- Interlevel shorts, 270
- Intrinsic base region, 85, 474-5, 493
- Ion implantation,
 - simulation of,
 - in PEPPER, 662
 - in SUPRA, 684-5
 - in SUPREM III, 658-63
- Isolation technologies, chap. 2
 - summary of submicron technologies, 65-6
- Isoplanar isolation process, 498
- IVEC DRAM cell, 604
- Junction isolation, 13-17
- Kirk effect, 467-8, 472, 474, 477, 481, 484, 488-9, 513, 537
- Knee current, I_k , 470
- Kooi effect, 27, 39, 315
- Latchup in
 - bipolar ICs, 486
 - CMOS ICs, 51, 400-19
 - circuit/layout techniques to suppress, 416-9
 - guard rings, 416-8
 - conditions needed to induce, 406
 - current gain of vertical parasitic bipolar transistors, 407
 - current gain of lateral parasitic bipolar transistors, 407-8
 - effects that induce latchup, 408
 - elimination in SOI, 67
 - ESD protection structure that
 - relies on latchup, 444
 - processing techniques to suppress, 413-16
 - test methods for characterizing, 410-12
- Laterally sealed LOCOS, 35-6
- Laser planarized Al films, 255-6
- LDD (lightly doped drain) MOSFETs, 355-61
 - asymmetrical I-V characteristics of, 359
 - effect on I-V characteristics, 356-7
 - fabrication sequence, 355
 - gate/drain overlap, 358
 - halo LDD, 363
 - oxide spacers, 354
 - reduction of E_{max} , 357
 - sloped junction LDD, 359
- Lead-zirconate titanate (see PZT)
- Leakage currents in DRAM cells, 592
- Lengths of interconnect lines on a logic chip, 184
- Lift-off, 258, 259-60, 286
- Lift-off using edge detection (see LOPED)
- Limited-reaction processing epitaxy, 508
- Line notching, 698
- Linear region of operation in MOSFETs, 306
- Linkup region formation, 516
- Local collector implant to reduce channeling
 - tail in self-aligned bipolar devices, 511, 516
- Local interconnects, 162-9, 178-9, 190-1
 - polycide, 179
 - polysilicon, 167-8, 178-9
 - dual-doped, 167-8

- selectively formed TiSi_2 (HPSAC)
- TiN over TiSi_2 , 166
- Ti:W over CoSi_2 , 166
- LOCOS isolation, 17-45, 323
 - etchback, 25, 31
 - fully recessed, 19, 28-31, 39-44, 498
 - bird's head, 30
 - FUOX, 41-2
 - in MOS evolution, 323
 - Kooi effect, 27
 - laterally sealed LOCOS, 35-6
 - limitations of, 27
 - OSELO, 43-4
 - oxide thinning in, 26
 - poly-buffered, 32
 - POP-SILO, 37
 - semi-recessed, 19-27
 - details of fabrication sequence, 20-27
 - SILQ, 33-9
 - simulation of 691-4
 - SLOCOS, 38
 - SPOT, 41
 - SWAMI, 39-41
 - double-silicon etched, 41
- Logic-in-memory circuits, 571
- LOPED, lift-off process, 259-60, 262
- Lot (def.), 6
- Lot traveller, 6
- Low-temperature CMOS, 447
- Manufacturing-based process simulators, 718**
- Mask-stack engineering, 38
- Masked ROMs, def., 570, 619-20
 - bipolar type, 620, 622
 - MOS NAND type, 620-1
 - MOS NOR type, 620-1
- Melting point, of conductors, 193
- Memories, semiconductor (see chap. 8)
 - architecture, 568
 - DRAMs (see DRAMs)
 - ROMs
 - EEPROMs (see EEPROMs)
 - masked ROMs (see Masked ROMs)
 - PROMs (see PROMs)
 - SRAMs (see SRAMs)
- Metal pitch (def.), 177
- Microcrystalline silicon emitters, 508
- Mil. Std., human body model test method, 444
- Micro-loading,
 - during etchback of resist, 228, 235
 - during etchback of SOG, 235
- during etchback of W, 250
- Microprocessors, 369
- MNOS EEPROMs, 628-9
- Molybdenum (Mo), 134, 192, 273, 279
 - as gate material, 398
- Mo-Ti-Si contacts, 134
- Mo-Ti:W-Si contacts, 134
- Mobile ionic charge, 319
- Mobility degradation effect, 345
- Monte Carlo analysis, 659, 660-62
- MoSi_2 Schottky contacts, 497
- MOSFETs
 - hot carrier effects in,
 - g_m degradation due to, 352
 - impact ionization, 349
 - in CMOS, vs. NMOS, 378
 - in PMOS devices, 362-3
 - injection of hot carriers into gate oxide, 350-1
 - latchup inducing, 410
 - lifetime reduction due to, 352-4
 - lightly doped drains (see LDDs)
 - maximum E-field, 348
 - snapback breakdown, due to, 350
 - substrate currents, 349-50, 352-4
 - V_T shifts, due to, 352-4
 - long channel device behavior of, 301-15
 - NMOS process sequence, 324-37
 - scaling of, 316
 - short-channel effects in, 323, 338-47
 - DIBL, 341-4
 - effect on I-V characteristics, 345
 - engineering model of, 346
 - narrow width effect, 23, 340-1
 - punchthrough, 341-4
 - V_T , effects on, 338-41
- Multi-chamber tool, 221
- Multi-layer (def.), 189
- Multi-level (def.), 189
- Na (sodium), 319, 335, 337
- Narrow width effect, 23, 340-1
- Navier-Stokes equation, 694
- Nested vias (def.), 242-3
- Ni, 256-7
- NiSi_2 , 146, 147
- Nitrogen incorporation in Al films, 261, 266
- NMOS
 - gate oxide growth, 331
 - isolation, 17-20
 - polysilicon gate, 332-3
 - source/drain formation, 154-60, 333-4
 - versus CMOS in power dissipation, 442-3

- Node-to-node punchthrough,
 - ESD protection device, 442-3
- Noise margin (def.), 375-6
- Non-threshold logic (NTL), 463
- npn* Transistors, 453, 454
 - design of, 477-8
- Ohmic contacts, 84, 87
- One-time programmable ROMs,
 - (see [OTP] ROMs)
- Open-circuit voltage gain, 466
- [OTP] ROMs, 623, 627
- OSELO, 43
- Oxidation-enhanced diffusion,
 - simulation of, 669
- Oxidation, rounding off, 602
- Oxidation, simulation of,
 - one-dimensional, 669-74
 - intrinsic rate constants of, 670
 - under high doping concentration conditions, 670-2
 - two-dimensional, 690-6
- CREEP, 648, 649, 694-6
- SOAP, 693-4
- SUPRA, 686
- SUPREM IV, 689, 692, 694
- Oxide growth,
 - field oxide, 23
 - gate oxide, 331, 434, 435
 - rounding off oxide, 53-4
 - trench sidewall oxidation, 54-5
 - viscoelastic growth, 31
- Oxide-isolated bipolar transistors, 498-500
- Oxide spacers, (see Spacers)
- Oxide thinning in submicron LOCOS
 - structures, 26
- OXSEF bipolar structure, 508
- Oxygen implantation,
 - doping of short, high-valued, polysilicon resistors, 581
 - in SIMOX, 72, 73
- Oxynitride films, 274
- Pad mask, 273, 337
- Pad oxide, 21, 327
- Parasitic resistances;
 - accumulation layer, R_{ac} , 85, 86, 138, 140-3
 - channel resistance, R_{ch} , 138-9
 - collector resistance, R_C , 472-3, 483, 484, 485, 527, 532, 545
 - contact resistance, R_{co} , 85, 86, 91-4, 140-3, 146
- (see also Specific contact resistivity)
 - between CVD W and Al, 252, 263
 - between CVD W and TiSi_2 , 150
- $R_{co(min)}$, 141-3
- diffusion sheet resistance in MOSFETs, R_{sh} , 85, 138-43, 145
- emitter resistance, R_E , 475
- extrinsic base resistance, R_{B1} , 86
- impact on devices, 142-3
- in bipolar transistors,
 - base, R_B , 469, 473-5, 477, 486, 492, 502, 532
 - extrinsic, R_{B1} , 473-4, 511
 - in self-aligned, bipolar transistors, 514
 - silicided poly to reduce R_{B1} , 514
 - intrinsic, R_{B2} , 86, 474-5
- MOSFET series resistance, R_S , 85, 86, 137-43
- spreading resistance, in MOS devices, R_{sp} , 85, 86, 138, 140-43
- substrate and well resistances in latchup structures, R_w and R_{sub} , 386
- Passivation layers, 273-6, 337
- PBM (planarization blocking mask), 225
- PEPPER, process simulator, 648, 662
- Perimeter punchthrough,
 - in bipolar transistors, 511
- Perimeter punchthrough voltage, 516
- PETEOS, 198, 273, 279
- Phenomenological models, 650-1
- Phosphorus
 - diffusion, simulation of, 665, 667-8
 - doping polysilicon with, 504-5
 - in situ, 540, 602
- Phosphosilicate glass (see PSG)
- Photoactive compound, 701
- Pillars, 245, 258-9, 286
- Pinholes, 261, 315
- PISCES, 91, 645, 649, 721-2
- Planar process for bipolar ICs, 458
- Planarity, need for, 65
- Planarization
 - of field oxides (see chap. 2)
 - of interlevel dielectrics (see also listings under individual planarization processes)
 - degree of planarization, 200-1, 224-5
 - achieved with etchback planarization, 224-5
 - design rules related to, 206
 - need for, 201-4
 - price paid when

- planarization is achieved, 204
- Plate electrode, 591, 599
- Plug, collector (see Deep collector contact)
- PMD (def.), 188-9
- PMOS in CMOS, 392
 - with p^+ poly, 397-9
 - boron penetration through gate oxide, 398-9
 - counterdoping of NMOS devices, 169, 397
- pnp Transistors, 453, 454
 - (see also chap. 6, Latchup)
 - load in bipolar SRAMs, 584
- pnpn Diodes (latchup structure),
 - in analog/digital BiCMOS, 545
 - in bipolar structures, 498
 - in complementary bipolar, 559-60
 - in CMOS, 400-8
- Poisoned vias, 235-6, 263
- Polyimides, 199, 214-7, 223, 275, 276, 283, 284
- Polybuffered LOCOS (see LOCOS)
- POP-SILO, 37-8
- Polysilicon
 - deposition, 332-3
 - diffusion barrier, 126-7
 - dual-doped, 169, 398-9
 - gate, 318, 332-3
 - in CMOS, 456
 - in self-aligned bipolar transistors (see Self-aligned bipolar structures)
 - p-doped, 397-9
 - boron penetration of gate oxide, 398-9
 - reoxidation of, 438
 - via filling with, 253
 - Polysilicon diodes, in SRAM cells, 585
 - Polysilicon emitters, 501-6, 511-2
 - arsenic doped, 504-5
 - models that describe their physics, interface dopant segregation model, 503-4
 - interface oxide tunneling model, 503
 - reduced mobility model, phosphorus doped, 504-5
 - process technology of, 504
 - in self-aligned bipolar structures, 511-2
 - Polysilicon load cells, 575
 - Poly-poly capacitors, 550
 - Post-exposure baking of resist, simulation of, 704, 710
 - Posis (see Pillars)
 - Power consumption
 - in BiCMOS vs CMOS gates, 528
 - in CMOS vs NMOS, 368-9
 - Power-supply voltages
 - in analog applications, 485-6, 544-6
 - in BiCMOS, 529, 545
 - in early PMOS, 317
 - in ECL, 461-2
 - in CMOS, 446-7
 - in TTL, 461-2
 - Preamorphization implant, 680
 - PREDICT, process simulator, 648, 652, 679-80
 - Process integration,
 - technology development flowchart, 10
 - levels of, 7
 - simulation in, 9
 - Profile interchange format (PIF), 653, 716
 - PROLITH, lithography simulator, 648, 649, 710
 - Propagation delay,
 - in bipolar logic families, 462, 463, 479, 480
 - interconnect lines, 183-6
 - minimum,
 - in digital bipolar circuits, 471-2, 478-9, 481
 - in digital MOS circuits, 481
 - PROPS, 47
 - PROXIS, 43
 - PSG, 197, 335, 337
 - Pseudomorphic epi layer, 508
 - PtSi-Si contacts,
 - 117-21, 135-6, 146, 147, 497-8
 - spiking in, 120
 - Al-Ti:W-PtSi-Si contacts, 124
 - in multilevel processes, 277, 278
 - PtSi salicide in bipolar ICs, 514, 519
 - Punchthrough
 - between adjacent DRAM cells, 603
 - prevention of, 605
 - in bipolar transistors, 471
 - in heterojunction bipolar transistors, 507
 - in MOS devices, 341-4, 388
 - in PMOS devices which are part of CMOS, 362
 - hot-electron induced, in PMOS, 362
 - node-to-node ESD prevention, 442
 - perimeter, in bipolar transistors, 511
 - punchthrough voltage, 516
 - prevention in adjacent MOS devices, 20
 - prevention implant, 388
 - that induces latchup in CMOS, 410
 - vertical, in CMOS devices
 - Pyrophoric, (def.), 198

- PZT, 637
- Quasi-saturation,
 - in bipolar transistors, def., 532
- RACPLE, process simulator, 718
- RCA-clean interfacial layers in poly emitters, 505
- Reachthrough, 465, 468, 488-9
- Reentrant angles on metal lines, 209
- Reflow, 104, 195, 335, 431
- Refresh signal, in DRAMs, 588, 594
- Refresh time, in DRAMs, 592
- Removable sidewall spacers to form self-aligned bipolar transistors, 518
- Resist etch rate during development, models of, 705
- Resist exposure, parameters, 699, 701-2, 710
- Resistance, parasitic series (see Parasitic resistance)
- Resistivity,
 - conductors, (Table) 186, 193
 - polysilicon
 - heavily doped, 178
 - lightly doped, 581-3
- Resistors (see also Appendix A)
 - in analog ICs, 544
 - polysilicon load, 575, 581-2
- Retrograde-well CMOS (see CMOS)
- Reverse narrow width effect, 60, 61
- ROMANS, process simulator, 647, 648
- Rounding off oxidation, 53-54
- in trench corners, 601-2
- Run sheet, 6
- Sacrificial-layer etchback,
 - of dielectrics (see Etchback, planarization of dielectrics)
 - of SOG (see Etchback of SOG)
 - of CVD W (see Etchback of W)
- Salicides (see Self-aligned contacts, MOSFETs)
- SAMPLE, process simulator, 644, 648, 649, 650, 683, 697-710, 717
 - Deposition and etch portion, 706-10
 - Lithography simulation portion, 709-10
 - Optical imaging subprogram, 700-1
 - Resist development subprogram, 704-5
 - Resist exposure subprogram, 701-3
- Saturation region of operation
 - in bipolar transistors,
 - in MOSFETs, 404, 457, 462
- Saturation velocity, 467, 489
 - SBC BiCMOS, 534
- Schottky clamp, 481, 496
- Schottky contacts, 87, 90, 134-7, 497
- Schottky diodes, 90 118, 134-7
 - device structures, 136-7, 497-8
 - in SRAM cell, 584
 - MoSi₂, 497
 - PtSi, 135-6, 497
 - Al-Si, 136
- Schottky TTL, 461
- SEDAN, bipolar device simulator, 645, 720
- SEG (see selectively-deposited epitaxy)
- Segregation of dopants at Si-SiO₂ interface, 669
- Segregation of dopants to grain boundaries
 - in polysilicon, 676
 - in silicides, 157
- Selective-epitaxially deposited base transistor, 513
- Selectively-deposited epitaxy, 56, 58-63, 66
 - CMOS, 388, 426-8, 439
 - corner faceting during, 59
 - epitaxial lateral overgrowth (ELO), 62-3
 - in DRAM cell, SEOT, 609
 - refill of oxide windows by SEG, 59-60
 - SELFOX, 61
 - SSSPD, 60
 - via filling with, 254-5
- Self-aligned bipolar structures, 510-12
 - double-poly self-aligned structures, 510-16
 - BSG as diffusion source in, 514
 - channeling tail causing wider base in, 511-2
 - process sequence to form, 511
 - selective epitaxial base transistor, 513
 - sidewall injection into the base, 514-5
 - reduction in current gain due to, 515
 - single-poly self-aligned structures, 516-7
 - amorphous poly to avoid channeling tail, 517
 - removable spacer for forming, 518
 - SPOT, 518
- Self-aligned contact structures,
 - in bipolar devices,
 - in MOS devices, 143-53, 157-8, 333
 - BOMOS, 153
 - elevated source/drain, 157-8
 - salicides, 143, 144-52
 - CoSi₂ (see CoSi₂ listing)
 - in CMOS, 436
 - silicon consumption during salicide

- formation, 147, 156
- TiSi₂ (see TiSi₂ listing)
- Self-isolation of MOS, 17
- SELFOX, 61
- Semi-recessed LOCOS (see LOCOS)
- Sense amplifiers in DRAMs, 569, 594
- Sensitivity analysis in process simulation, 649
- SGT DRAM cell, 609-10
- SEOT DRAM cell, 609, 611
- SEREPI, 62, 426-7
- Shallow angle implants, 47
- Shallow junctions, 154-60
 - in bipolar transistors,
 - in MOSFETs
 - BF2 implant, 154, 5
 - GILD, 158-9
 - implanting into an already formed silicide layer, 156, 157
 - lateral channeling, 155fn
 - photo-enhanced epitaxy, 159
 - preamorphization, 155
 - RTP, 155
- Sheet resistance, conductors (Table - 186)
- Short-channel effects in MOSFETs (see MOSFETs)
- Sidewall inversion in trenches, 51, 388
- SiCl₄, 489, 539
- SICOS, 520-1
- SIFCOD, device simulator, 645, 649, 721-2
- SiH₄ (silane), 489, 539
- SiH₂Cl₂, 489
- Silicon dioxide
 - boron doped, 198
 - ECR plasma CVD, 237-8
 - CVD, 194-99
 - phosphorus doped, 197, 273
- Silicon-gate MOS, 318-9
- Silicon gel, 275
- Silicon nitride
 - fluorinated, 362
 - hydrogen in, 362
 - interlevel dielectric, 199
 - masking layer in LOCOS, 21, 327
 - passivation layer material, 274, 337
 - under polyimide, 216
- Silicon-on Insulator (SOI), 66-79
 - by implanting nitrogen, 75
 - dielectric isolation (DI), 67-9
 - FIPOS, 76-7
 - separation by implanted oxygen (SIMOX) 72-3
- Silicon-on -Sapphire (SOS), 72
- Wafer bonding, 70-2
- ZMR, 75-6
- SILO, 53-4, 696
 - POP-SILO, 37-8
 - Planarized SILO, 38-9
 - SLOCOS, 38
- SIMBAD, process simulator, 713-4
- SIMOX, 72-3
 - thin-film SIMOX, 74
- SIMPL, layout/process simulator, 645, 648, 649, 714-8
 - SIMPL-2, 716
 - SIMPL-DIX, 648, 649, 716-8
- Sinker (see Deep-collector contact)
- Sintering contacts, 110-1, 111-6
- SIPOS, 508
- SLOCOS, 38
- Sloped-junction LDD, 354
- Snapback breakdown, 350
- Snowplowing, 147
- SOAP, 679, 693-4
- Soft errors,
 - in DRAMs, 603, 615-18
 - in SRAMs, 583-4
- Soft-error rate, 617-8
- SOG, 199, 222, 229-36, 275
 - etchback process, 235-6
 - in multilevel metal processes, 278, 280, 281, 283, 286,
 - non-etchback process, 235-6
 - process integration, 232-3
 - silicates, 230-2
 - siloxanes, 230-2
 - used in lieu of resist
 - in planarization etchback, 227
- Source-bias effect in MOS transistors, 304-5
- Spacers
 - for forming self-aligned bipolar transistors, 518
 - for LDDs, 354-5
 - removable, in CMOS, 437
 - for planarization, 212
 - for silicides, 144-5
 - in self-aligned structures
 - in bipolar devices,
- Specific contact resistivity, 91-101, 140-3
 - CVD W to Si, 129, 129fn
 - extracting accurate values of, 91-101
 - with CBKR structure, 94-100
 - with CER structure, 94-100
 - with TLTR structure, 94-10
 - measuring the value of,
 - in silicide contacts, 152

- table of values, 100
- TiSi₂ to Si, 148-9
- Ti:W to Si, 126
- Split poly process, 541
- Spin-on-glass (see SOG)
- SPOT, bipolar structure, 518
- SPT DRAM cell, 605-6
- SRAMs, 570, 572-87
 - bipolar, 573, 584-5
 - BiCMOS, 573, 585-7
 - CMOS, 573
 - alternative cell designs, 576
 - full CMOS, 575-6
 - layout of, 578
 - poly-load, 575
 - NMOS, 573, 575, 576-8
 - circuit operation of, 576-7
 - layout of, 577-8
 - I/O capability, 570, 574
- SSPD, 60
- SST DRAM cell, 605-6
- Stacked capacitor DRAM cells, 610-15
 - DASH cell, 614
 - fin structure cell, 614-5
 - SST cell, 614-5
- Stacked vias (def.), 242
- Staggered vias (def.), 242
- Standard buried collector bipolar process, 13, 14-6
 - testing the isolation structures of, 15
- Standing wave effects, in resist, 703-4
- Static column decoding, 597
- Step coverage (def.), 202
 - in filling vias, 240
- Storage electrode, 591
- Stress-induced voids in Al lines, 265
- String advance algorithm, 706-7
- Stringers (def.), 202
- STT DRAM cell, 605-6
- Studs (see Pillars)
- Subthreshold current
 - in MOSFETs, I_{DS}, 308, 311-3, 341
 - in PMOS, 394
- Subthreshold swing (def.), 312, 341, 344-5
 - in PMOS, 394-5
- SUPRA, 648, 649, 650, 684-7, 688, 689-90, 711-2
 - diffusion models in, 685,
 - epitaxial models in, 685
 - input file in, 685-7
 - ion implantation models in, 684-5
 - oxidation models in, 686
- SUPREM III, 655-679
 - diffusion models in, 663-9
 - epitaxy models in, 674-5
 - input file format, 677-9
 - ion implantation models, 658-63
 - oxidation models in, 666-74
 - polysilicon models of in, 675-7
- SUPREM IV, 648, 649, 650, 652, 655, 687-91, 694
 - diffusion models in, 688-9
 - input file format, 689, 690
 - ion implantation models in, 689
 - oxidation models in, 689, 692, 694
- Surrounding-gate DRAM cell (SGT), 609
- Ta₂O₅, 435
 - in SRAM cell, 585, 597
- TBP, 602
- Tetra-ethyl-ortho-silicate, (see TEOS)
- Tetra-methyl-cyclo-tetrasiloxane (see TMCTS)
- TEOS, 194, 195, 211, 234, 279, 281, 286
 - PECVD, 194, 198, 273
 - Low-temperature, 198
- Textured-poly EEPROMs 631-2
- Three-level metal systems, 283-5
- Three-transistor DRAM cell, 588
- Threshold loss, 379
- Threshold voltage in MOS transistors,
 - controlling through process
 - and circuit design techniques, 309
 - body biasing, 309
 - ion implantation, 311, 321-2, 331
 - with single implant to adjust both NMOS and PMOS V_T in CMOS, 393, 394
 - hot electron effects on, 351-2, 362-3
 - in early PMOS IC technology, 317
 - in field regions, 17-8
 - in silicon-gate NMOS, 318-22
 - long channel transistors, 301-5
 - reduction in field regions in CMOS, 422
 - short-channel effects on, 338-40
- Ti, 127
- TiF₃, 130
- TiN
 - as an adhesion layer for W films, 247-8, 282
 - as a diffusion barrier, 128, 166, 285
 - as a local interconnect,
 - in BiCMOS, 542
 - in CMOS SRAM, 580
 - TiN over TiSi₂, 166-7
 - TiN/Al interconnect, 286

- TiSi₂, 146, 146fn, 147-50, 156, 163, 164-6, 166-7, 184
 - contacts, 146, 147-9, 151
 - contact electromigration in, 149
 - dopant depletion effects in, 199
 - specific contact resistivity of, 148-9
 - stress-induced defects in, 150
 - CVD W in contact with TiSi₂, 150
 - implanting dopants into, 156
 - ion beam mixing formation of, 156
 - local interconnect, 164-6, 166-7, 168
 - resistivity of, 146
 - salicides, 147-50
 - bridging during formation of, 148
 - prevention of, 148
 - selective wet etchant for, 148, 164
 - silicon consumption during the formation of, 152
 - silicon "suck out," problem, 165
- Ti:W
 - as a diffusion barrier, 123, 124-6
 - nitrogen stuffed, 124-5
 - thermal decomposition of, 125-6
 - direct contacts to Si with, 126
 - particulate generating qualities of, 126
 - under W to prevent high contact resistance to Al, 263
 - use in multi-level metal schemes, 276, 278, 279, 280
- Titanium, 127
- Titanium-tungsten (see Ti:W)
- TMCTS, 198-9
- Transconductance
 - of bipolar transistors, 527-8
 - of MOS transistors, g_m , def., 308, 527-8
- Transconductance parameter in MOSFETs, k , def., 306
- Transfer characteristics, 373, 375
- Transfer length, l_T , 93, 142
- Transit time, 477
- Transit time, base, 477, 479, 480, 481, 492
- Transistor-transistor logic (see TTL)
- Trench capacitors, 56
- Trench capacitor DRAM cells, 600-9, 618 (see also DRAMs)
- Trench isolation, 51-8, 66
 - deep trench
 - etching, 51-4
 - filling, 54-6
 - oxidation of poly, 58
 - planarization after refill, 56-8
 - rounding oxidation, 53-4
 - SEG refill, 48-50
 - sidewall oxidation, 54-5
 - to suppress latchup in CMOS, 425
 - moderate depth trenches, 48-50
 - Toshiba process, 49-50
 - U-groove, 48-9
 - shallow trenches, 45-8
 - BOX, 45-8
- Triboelectricity, 441
- Triggering current, latchup, 410
- Triple-diffused bipolar transistor, 13
 - process flow, 482-3
 - in BiCMOS, 532-4, 536
- Triple-polysilicon structure, 631-2, 633-5
- TTL, 317, 460-2
- TTL-to-ECL translators, 462, 528
- Tungsten
 - as a gate material, 398
 - CVD deposition of,
 - blanket, 246, 247-51, 259
 - adhesion layer for, 246-7
 - by hydrogen reduction, 130, 246
 - by silane reduction, 130-1, 246-7
 - by silicon reduction, 246
 - effect of Si surface on deposition process, 129fn
 - selective, 169, 191, 251-2, 259
 - on poly, 169
 - on Al, 191, 271
 - CVD W in contact with TiSi₂, 150
 - CVD W in contact with Al, 252, 263
 - etching, 192, 248-9, 251
 - leakage of junctions under W contacts, 130-1
 - specific contact resistivity to Si, 129
 - use in multi-level metal systems, 278, 280, 282-3, 284
- Tunneling,
 - in metal-Si contacts, 88
 - through gate oxide, 434
- Twin-well CMOS (see CMOS)
- Ultra-high speed SRAMs, 584-5
- Ultra-high vacuum epitaxy, 508
- Ultra-thin gate oxides, 435
- U-groove isolation, 48-9
- Unity gain frequency, f_T ,
 - 470, 472, 477, 478-9, 481
 - in analog/digital BiCMOS processes, 545
 - in heterojunction bipolar transistors, 506, 508
 - in polysilicon emitter transistors, 502
 - in self-aligned bipolar transistors, 512
- U-V light, 623, 626-7, 632, 633, 701

- V_{CEsat} , 472
- Velocity saturation factor, 345
- Vertical punchthrough in CMOS, 384-5
- Vias
 - defined, 189
 - design rules related to sloped vias, 241-3
 - filling
 - advanced, vertical-sided, 244-57
 - with bias-sputtered Al
 - with CVD W
 - blanket, 247-51
 - selective, 251-2
 - with electroless-deposited Ni or Pd, 256-7
 - with laser-planarized Al, 255-6
 - with polysilicon, 253
 - blanket, 253
 - selective, 253
 - with silicide, 253
 - conventional approach, 240-4
 - required degree of filling, 245
 - high resistance in
 - CVD W to Al, 252, 263
 - sloped, 204
 - varying depths, 205
- Video RAMs (VRAMs), 572
- Viscoelastic flow during oxide growth, 602
- Voids,
 - formed during
 - CVD SiO₂ deposition, 197, 207, 211, 214, 228, 275
 - CVD W deposition, 247-51
 - passivation layer deposition, 275
 - trench refill with polysilicon, 56
 - prevention of,
 - using CVD/etch, 222
 - using SOG, 234
- Wafer bonding, 70-2
- Wafer-scale integration, 186
- Walled-emitter process, 496
- Washed emitter process, 496
- White ribbon (Kooi effect), 27
- WIT, 259, 261
- Word line, def., 568
- Work-function difference,
 - in MOS devices, 302-3, 309
 - in metal-Si contacts, 89
- Write speed in ferroelectric RAMs, 635
 - compared to UV-PROMs or EEPROMs, 635