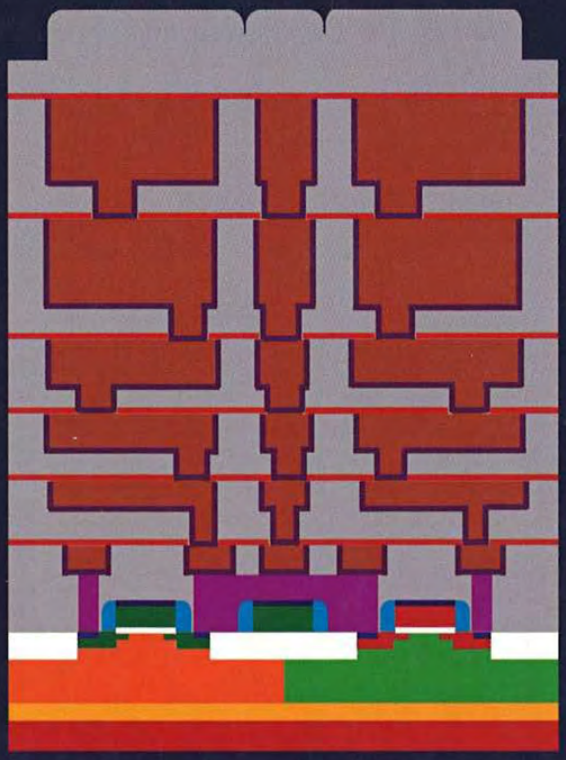
Kirkland & Ellis Library
5 0628 01064435 6

Silicon Processing

for the VLSI Era

Volume 4 - Deep-Submicron Process Technology



S. Wolf

SILICON PROCESSING FOR THE VLSI ERA

VOLUME 4:

DEEP-SUBMICRON PROCESS TECHNOLOGY

STANLEY WOLF Ph.D.

LATTICE PRESS

Sunset Beach, California

DISCLAIMER

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

Published by:

LATTICE PRESS
Post Office Box 340
Sunset Beach, California 90742, U.S.A.

Cover design by: S. C. Ross Wolf.

Copyright © 2002 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 4: Deep-Submicron Process Technology

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

ISBN 0-9616721-7-X

9 8 7 6 5 4 3 2 1

PRINTED IN THE UNITED STATES OF AMERICA

DETAILED TABLE OF CONTENTS

PREFACE

Chap. 1 - THE EVOLUTION OF THE STRUCTURE OF MOSFETS 1

1.1 THE STRUCTURE OF DEEP-SUBMICRON MOSFETS: (0.25-μm to 0.13-μm) - COMPARED TO THE STRUCTURE OF "CONVENTIONAL" MOSFETS (2.0-μm to 0.5-μm) 2

1.1.1 Evolution of the MOSFET Gate Stack and Contact Structure
1.1.2 Gate Dielectric Materials in Deep-Submicron MOSFETs
1.1.3 Doping-Concentration Profiles of the MOSFET Channel
1.1.4 Evolution of the Drain Structure of MOSFETs

1.2 DEEP-SUBMICRON CMOS STRUCTURES 10

1.2.1 Substrate Issues for Deep-Submicron CMOS
1.2.2 Well Formation in Deep-Submicron CMOS
1.2.3 Dual-Doped Poly in Deep-Submicron CMOS
1.2.4 Shallow Trench Isolation for Deep-Submicron CMOS
1.2.5 Silicon-On-Insulator (SOI) Technology

1.3 LIMITS TO CONVENTIONAL MOSFET SCALING 15

REFERENCES 16

Chap. 2 - 300-MM SILICON WAFERS 17

2.1 300-mm SILICON CRYSTAL GROWTH 17

2.2 GROWN-IN SILICON DEFECTS: 21
CRYSTAL-ORIGINATED-PARTICLES (COPS) & DISLOCATION LOOPS

2.3 DETAILS OF THE FORMATION OF CRYSTAL-ORIGINATED-PARTICLES (COPS) 23

2.3.1 The Radial Distribution of Grown-In Defects on the Wafer Surface

2.4 THE OXYGEN-STACKING-FAULT RING (OSF-RING) 26

2.5 MITIGATING EFFECTS OF COPS BY USE OF POST-CRYSTAL-GROWTH ANNEALING 27

2.6 ELIMINATION OF COPS WITH "PERFECT-SILICON" 29

2.7 MINIMIZING PERFORMANCE-DEGRADATION CAUSED BY COPS THROUGH THE USE OF HIGH-PULL-SPEED SILICON 30

2.8 GETTERING FOR ULSI PROCESSES 32

2.8.1 Basic Gettering Principles
2.8.2 Extrinsic Gettering
2.8.3 Intrinsic Gettering
2.8.4 Gettering with Oxygen Precipitates
2.8.5 Summary of Gettering

2.9 STATUS OF SILICON WAFER TECHNOLOGY & FUTURE TRENDS 45

2.9.1 Epitaxy-Optimized Substrate (EOS) Wafers
2.9.2 Comparing the Grown-In Defect Characteristics of the "New" Wafers
2.9.3 Gettering Methods & Denuded Zone Formation Techniques Used with the "New" Wafers
2.9.3.1 "Magic DeNuded Zone"

2.10 FROM INGOT TO FINISHED WAFER: SLICING, ETCHING, AND POLISHING 52

2.10.1 Ingot Evaluation
2.10.2 Ingot Surface Grinding
2.10.3 Grinding Flats or Notches on the Ingot for Orientation Purposes
2.10.4 Sawing the Ingot into Slices (Wafers)
2.10.5 Laser Marking the Wafers
2.10.6 Lapping and Grinding the Wafers
2.10.7 Removal of Surface Mechanical Damage by Chemical Etching
2.20.8 Rounding the Wafer Edge (and Notch)
2.10.9 Edge Polishing of the Wafers
2.10.10 Chemical-Mechanical Polishing of the Wafers
2.10.11 Cleaning the Wafers
2.10.12 Depositing Epitaxial Silicon Layers on the Wafers
2.10.13 Shipping 300-mm Wafers: Its Impact on the Evolution of Wafer Shipping Boxes

2.11 SPECIFICATIONS OF SILICON WAFERS FOR VLSI 63

2.12 THE ECONOMICS OF SILICON WAFERS 66

REFERENCES 69

Chap. 3 - GATE DIELECTRICS: THIN GATE OXIDES 75

3.1 REQUIRED CHARACTERISTICS OF GATE DIELECTRICS FOR DEEP-SUBMICRON MOSFETS 76

3.2 THE STRUCTURE OF THERMALLY GROWN SiO_2 AND THE PROPERTIES OF THE Si/SiO_2 INTERFACE 80

3.2.1 The Microscopic Structure of Thermally Grown SiO_2
3.2.2 The Si/SiO_2 Interface
3.2.2.1 Interface Trap Charge
3.2.2.2 Effect of Interface Traps on IC Characteristics
3.2.2.3 Oxide Trapped Charge
3.2.2.3 Effect of Oxide Trapped Charge on Device Characteristics
3.3 DIELECTRIC BREAKDOWN IN SILICON DIOXIDE FILMS 90
3.3.1 Electron Trapping in Silicon Dioxide:
3.3.2 The Electric-Field Driven Model of Oxide Degradation (*E* Model)
3.3.3 The Current-Driven Model of Oxide Degradation (1/E Model)
3.3.4 The Hole-Trapping Model that Describes How Holes are Injected & Trapped in SiO_2
3.3.5 Comparing the Electric-Field Driven and Current Driven Oxide Breakdown Models
3.3.6 Time to Breakdown (T_{BD}) and Charge to Breakdown (Q_{BD})
3.4 LEAKAGE CURRENTS IN SiO_2 FILMS (TUNNELING PHENOMENA) 99
3.4.1 Fowler-Nordheim (FN) Tunneling (Tunneling Into Silicon Dioxide)
3.4.2 Direct Tunneling (Tunneling Through Silicon Dioxide)
3.5 MODELS OF THIN OXIDE GROWTH 104
3.6 SINGLE-WAFER TECHNOLOGY OF THIN OXIDE GROWTH 109
3.6.1 Rapid Thermal Oxidation Tools
3.6.2 Wet RTO Processes
3.7 NITRIDED & FLUORINATED OXIDES AS GATE DIELECTRICS 112
3.7.1 Oxynitridation of Silicon in N_2O
3.7.2 Oxynitridation of Oxides in N_2O or NO
3.7.3 Fluorinated Gate Oxides
3.8 PROJECTIONS OF THICKNESS LIMITS OF GATE OXIDES 121
3.8.1 Minimum Oxide Thickness Due to Defects and Tunneling
3.8.2 Minimum Oxide Thickness: Stress Induced Leakage Current (SILC)
3.8.3 Soft-Breakdown of Oxide Films
3.8.4 Impact of Polysilicon Depletion
3.8.5 Impact of Process Induced Damage
3.8.6 Summary of Oxide Thickness Projections
3.9 MEASURING THIN GATE OXIDES 135
3.10 MANUFACTURING THIN GATE OXIDES 137
3.10.1 Process Control Issues of Growing Ultra-Thin Gate Oxides
3.10.2 Thin Gate Dielectric Stacks Based on Silicon Dioxide
REFERENCES 139

Chap. 4 - HIGH-*k* DIELECTRICS 145

4.1 HIGH-*k* DIELECTRICS AS THE GATE DIELECTRIC IN MOSFETS 145
4.1.1 Integrating High-*k* Dielectrics into MOSFET Structures:
4.2 HIGH-*k* DIELECTRICS AS THE CAPACITOR DIELECTRIC IN DRAMS 150
4.3 HIGH-*k* DIELECTRICS FOR FeRAM APPLICATONS 154
4.4 TANTALUM PENTOXIDE (Ta_2O_5) 154
4.4.1 Tantalum Pentoxide as a DRAM Capacitor Dielectric
4.4.2 Tantalum Pentoxide as a MOSFET Gate Dielectric
4.5 BARIUM STRONTIUM TITANATE (BST) 162
AS A DRAM CAPACITOR DIELECTRIC
4.5.1 Deposition of BST Films by CVD
4.5.2 Microstructure Effects on the Electrical Properties of BST
4.5.3 BST Integration Issues
4.6 USE OF OTHER HIGH-*k* MATERIALS AS GATE DIELECTRICS 169
4.6.1 Atomic Layer Deposition for Depositing Thin High-*k* Dielectric Films
4.6.2 Aluminum Oxide (Al_2O_3) as a Gate Dielectric
4.6.3 Zirconium Oxide (ZrO_2) as a Gate Dielectric
4.6.4 Hafnium Oxide (HfO_2) as a Gate Dielectric
4.6.5 Praseodymium Oxide (Pr_2O_3) as a High-*k* Gate Dielectric

REFERENCES 177

Chap. 5 - THE STRUCTURE OF DEEP-SUBMICRON MOSFETS: Retrograde-Wells; Super-Steep Retrograde Channel Doping; Drain Engineering; Punchthrough-Control Implants 181

5.1 WELL-FORMATION IN DEEP-SUBMICRON CMOS 181
5.1.1 Retrograde-Wells for Deep Submicron CMOS
5.1.2 Process Integration Issues Involving the Use of Retrograde-Wells
5.2 SUPERSTEEP RETROGRADE CHANNEL (SSR) PROFILES 191
5.2.1 Fabricating SSR Profiles for Sub-0.1 *m*m MOSFETs
5.3 SOURCE/DRAIN ENGINEERING IN DEEP-SUBMICRON CMOS 195
5.3.1 Parasitic Resistance of Deep-Submicron Source/Drains
5.3.1.1 Physical Meaning of R_{ac}, R_{sp}, and R_{SDE}
5.3.1.2 Comparing the Resistance of the SDE Region ($R_{sp} + R_{SDE}$) to R_{sat}
5.3.1.3 Parasitic Resistance of the Deep Contact Region of the Drain R_{co}
5.3.2 MOSFETs with Elevated Source/Drains

145

S 145

AMS 150

169

ilms

181

1

195

5.3.2 MOSFETs with Elevated Source/Drains
5.3.3 Shallow-Junction Formation for Deep-Submicron Source/Drains
5.3.3.1 Ultra-Low-Energy Implants
5.3.3.2 Transient-Enhanced Diffusion
5.3.3.3 Rapid Thermal Annealing of Shallow Junctions
5.3.3.4 Dopant Loss for Sub-keV Implants from Self-Sputtering, Out Diffusion, and Surface Oxide
5.3.4 Effect of the Overlap of the SDE Region and Gate Edge on MOSFET Performance
5.3.5 SDE Junction Lateral Abruptness
5.4 ANTI-PUNCHTHROUGH STRUCTURES IN DEEP-SUBMICRON CMOS 219
REFERENCES 223

Chap. 6 - ADVANCED LITHOGRAPHY I: DEEP-SUBMICRON RESISTS 227

6.1 CHEMICALLY-AMPLIFIED DEEP-UV RESISTS FOR OPTICAL LITHOGRAPHY 227
6.1.1 248-nm Photoresists
6.1.1.1 Amine-Contamination of DUV CA Resists
6.1.1.2 Current State of 248-nm Resist Technology
6.1.2 193-nm DUV Photoresists
6.1.2.1 Problems of 193-nm Resists
6.1.3 157-nm DUV Photoresists
6.2 ANTI-REFLECTIVE COATING (ARCs) 244
6.2.1 Organic BARCs
6.2.2 Inorganic Dielectric BARCs
6.2.3 PVD-Deposited BARCs
6.2.4 BARCs for Dual-Damascene Applications
6.3 PHOTORESIST PROCESSING SYSTEMS 251
REFERENCES 257

Chap. 7 - ADVANCED LITHOGRAPHY II: OPTICS AND HARDWARE 259

7.1 EXCIMER LASER DEEP-UV LIGHT SOURCES 259
7.1.1 KrF Excimer Lasers
7.1.2 ArF Excimer Lasers
7.1.3 F_2 Excimer Lasers
7.2 EXPOSURE TOOLS For DUV LITHOGRAPHY 265
7.2.1 Exposure Tools for 248-nm Lithography

7.2.2 Exposure Tools for 193-nm Lithography
7.2.3 Exposure Tools for 157-nm Lithography
7.2.4 Calcium Fluoride Optical Elements for DUV Lithography
7.2.5 300-mm Lithography Tools
7.2.6 Mix-an-Match Lithography
7.3 RESOLUTION ENHANCEMENT TECHNOLOGIES (RETs) 274
7.3.1 Off-Axis Illumination
7.3.2 Optical Proximity Correction (OPC)
7.3.3 Phase Shift Masks (PSM)
7.4 MASK ERROR FACTOR (MEF, or MEEF) 292
7.5 EXTENDING THE LIMITS OF OPTICAL LITHOGRAPHY 296
7.6 NON-OPTICAL (or NEXT GENERATION) LITHOGRAPHIC TECHNOLOGIES (NGL) 301
7.6.1 Extreme Ultra-Violet Reflective Projection Lithography (EUV)
7.6.2 Electron Beam Projection Lithography (SCALPEL and PREVAIL)
REFERENCES 309

Chap. 8 - CHEMICAL MECHANICAL POLISHING (CMP) 313

8.1 TERMINOLOGY AND EVOLUTION OF PLANARIZATION PROCESSES FOR ICS 313
8.1.1 Defining the Degree of Planarization
8.1.2 The Need for Dielectric Planarization:
8.1.3 Design Rules Related to Intermetal Dielectric-Formation and Planarization Processes
8.1.4 Step-Height Reduction of Underlying Topography as a Technique to Alleviate the Need for Planarization
8.1.5 Planarization through Sacrificial-Layer Etchback
8.1.5.1 Sacrificial-Etchback Process Problems
8.2 INTRODUCTION TO CHEMICAL MECHANICAL POLISHING (CMP) 322
8.3 THE TERMINOLOGY USED TO CHARACTERIZE CMP PROCESSES 325
8.3.1 CMP Removal Rate (RR)
8.3.2 The Degree of Planarization (DOP)
8.3.3 Within-Die Non-Uniformity (WIDNU)
8.3.4 Within-Wafer Non-Uniformity (WIWNU)
8.3.5 Wafer-to-Wafer Non-Uniformity (WTWNU)
8.3.6 Efficiency of Planarization (EOP)

8.5.1 The Mechanical Aspects of Silicon Dioxide Removal by CMP: (Preston's Law)
8.5.1.1 Model of WIWNU Effects in Rotary CMP Tools Based on Preston's Law
8.5.2 Models Which Describe Factors that Impact Oxide RR in CMP Through the Preston Constant
8.5.2.1 Impact on the Oxide RR from the Electrochemical Interactions Between the Slurry Particles and the Oxide Surface
8.5.2.2 Dependence of Oxide RR in CMP on Dielectric Hardness
8.5.3 Models Which Describe Chemical-Mechanical Phenomena that Give Rise to Planarization Phenomena Associated with CMP of Oxides
8.5.3.1 Model of the Mechanism That Produces Planarization of Steps in Oxide CMP
8.5.3.2 Pattern Dependence of the Planarization Rate in CMP Processes
8.5.4 Metal-CMP Mechanisms
8.6 CMP EQUIPMENT 352
8.6.1 CMP Polishing Tools:
8.6.1.1 CMP Polishing Tool Configurations
8.6.1.2 Rotary CMP Tools:
8.6.1.3 Orbital CMP Tools
8.6.1.4 Linear CMP Tools
8.6.1.5 Fixed Table CMP Tools
8.6.1.6 300-mm CMP Tools
8.6.2 Wafer-Carriers in CMP Polishing Tools (and the CMP Edge Effect)
8.6.2.1 CMP Edge-Effect
8.6.3 CMP Consumables (Slurries):
8.6.3.1 Slurries for Oxide-CMP
8.6.3.2 Slurries for Metal-CMP
8.6.3.3 Multi-Step Slurries
8.6.4 Slurry-Distribution Systems
8.6.5 Environmental, Safety, and Handling Issues of CMP Slurries
8.6.6 CMP Consumables (Polishing-Pads)
8.6.7 Pad-Conditioners
8.6.8 Slurry-Free Pads
8.6.9 Endpoint-Detection in CMP
8.7 CLEANING ISSUES IN CMP 392
8.7.1 Mechanisms of Particle Adhesion and Removal from Surfaces
8.7.1.1 Electrical Repulsion of the Particle and the Wafer Surface
8.7.1.2 Chemically Assisted Particle Removal
8.7.1.3 Mechanically Assisted Particle Removal
8.7.1.4 Steric Stabilization for Suspending Particles in a Solution

8.7.2 Practical Particle Removal Processes: Brush Scrubbing
8.7.2.1 Brush Scrubbing to Remove Particles in Post-Oxide-CMP Processes
8.7.2.2 Brush Scrubbing to Remove Particles in Post Metal-CMP Processes
8.7.3 Vibrational Scrubbing to Remove Particles (Megasonic)
8.7.4 Secondary (or Buff) Polish Procedure as a CMP Cleaning Step
8.7.5 Post-CMP Cleaning of Metallic Contaminants
8.7.6 CMP Cleaning Equipment
8.8 CMP METROLOGY 407
8.8.1 Detection of Defects Associated with the CMP Process
8.9 CMP POLISHER TOOL RELIABILITY 411
8.10 CMP SYSTEMS AND PROCESS INTEGRATION 412
8.11 CMP OF VARIOUS MATERIALS 413
8.11.1 CMP of Oxide Interlevel & Intermetal Dielectrics (ILDs & IMDs)
8.11.2 CMP of Tungsten
8.11.3 CMP of Copper
8.11.4 CMP of Aluminum
8.11.5 CMP of Polysilicon
8.11.6 CMP of Photoresist
8.11.7 CMP of Low-*k* Dielectrics
8.12 MISCELLANEOUS ISSUES CMP 421
8.12.1 Dishing and Erosion
8.12.2 Thickness Non-Uniformity Within a Wafer After CMP
REFERENCES 428

Chap. 9 - SHALLOW TRENCH ISOLATION (STI) 433

9.1 EARLY SHALLOW TRENCH ISOLATION STRUCTURES 435
9.2 STI-ENABLING PROCESSES 438
9.3 SHALLOW TRENCH ISOLATION FOR CMOS 439
9.4 DETAILS OF THE PROCESS FLOW TO FORM A BASELINE SHALLOW-TRENCH-ISOLATION (STI) STRUCTURE 446
9.5 ISSUES INVOLVED WITH STI PROCESS INTEGRATION 455
9.5.1 Issues Involved With Etching Trenches for STI
9.5.2 Corner-Rounding Techniques in STI Structures
9.5.3 Trench-Fill Dielectrics for STI
9.5.4 Dishing Problem Associated with the CMP of STI Trench Dielectrics.
9.5.5 Corner Engineering and Its Effects on MOSFET TURN-OFF Characteristics of MOSFETs Fabricated with STI
REFERENCES 472

Chap. 10 - SILICON-GERMANIUM (Si-Ge) TECHNOLOGY FOR HIGH-PERFORMANCE TELECOMMUNICATIONS ICs 475

10.1 INTRODUCTION TO HETEROJUNCTION BIPOLAR TRANSISTORS 475
10.2 HETEROJUNCTION BIPOLAR TRANSISTORS WITH LARGE-BANDGAP EMITTERS 480
10.3 HETEROJUNCTION BIPOLAR TRANSISTORS WITH SMALLER BANDGAP BASE REGIONS (Si-Ge BASE) 482
10.4 PROCESS TECHNOLOGY FOR FABRICATING SI-GE HBTS 484
10.5 PROCESS INTEGRATION ISSUES FOR BiCMOS TECHNOLOGY WITH Si-Ge HBT 486
10.6 HIGH-QUALITY SI-GE EPITAXIAL FILM GROWTH ON PATTERNED WAFERS 491
10.7 THERMAL STABILITY PROCESS INTEGRATION ISSUES 492
10.8 COMPATIBILITY WITH ION IMPLANTATION PROCESSES 492
10.9 COMPATIBILITY WITH THERMAL OXIDATION PROCESSES 493
10.10 VERTICAL SCALING OF Si:Ge HBTS 493
10.11 BASE-DURING-GATE AND BASE-AFTER-GATE BiCMOS PROCESS FLOWS 495
10.12 EXAMPLES OF THE MODULAR INTEGRATION OF Si-Ge HBTS INTO A BiCMOS PROCESS FLOW 497

REFERENCES 499

Chap. 11 - SILICON-ON-INSULATOR (SOI) TECHNOLOGY 501

11.1 WHAT IS SILICON-ON-INSULATOR (SOI)? 501
11.1.1 Isolation Issues in CMOS in Bulk Silicon
11.1.2 Isolation in SOI CMOS
11.2 THE "END OF THE ROAD FOR BULK CMOS" SCENARIO PREDICTED BY THE 1999 INTERNATIONAL TECHNOLOGY ROADMAP FOR SEMICONDUCTORS 505
11.2.1 Source/Drain Junction Scaling
11.2.2 Dopant Activation in Shallow Source/Drain Junctions
11.2.3 Doping Profile Abruptness in Shallow Source/Drain Junctions
11.3 WHY SOI? 510
11.3.1 The Path of Using SOI May Allow Extension of the ITRS Roadmap
11.4 HISTORY OF SILICON-ON-INSULATOR (SOI) TECHNOLOGY 516
11.5 SILICON-ON-INSULATOR DEVICES 518
11.5.1 Partially-Depleted Thin SOI MOSFETs (PD-SOIs)

11.5.2 Fully-Depleted Thin SOI MOSFETs (FD-SOIs)
11.5.3 Process Technologies Used to Fabricate Fully-Depleted SOI Devices
11.5.3.1 FD-SOI: Thinning Si-SOI Surface Layer by Thermal Oxidation
11.5.3.2 FD-SOI Fabrication by LOCOS Recess of the MOSFET Channel Region
11.6 FABRICATION TECHNOLOGIES USED TO PRODUCE SOI-STARTING-WAFERS 527
11.6.1 Silicon-on-Sapphire (SOS)
11.6.2 SOI by Separation by Implanted Oxygen (SIMOX)
11.6.2.1 High-Dose SIMOX
11.6.2.2 Low-Dose SIMOX
11.6.3 Wafer Bonding
11.6.3.1 Bond and Etch-back SOI (BESOI)
11.6.3.2 Hydrogen-Implantation-Induced Layer Splitting to Create SOI Substrates (SMART-CUT®)
11.6.3.3 NANOCLEAVE® Wafer Bonding SOI
11.6.3.4 ELTRAN® Wafer Bonded SOI
11.7 LATERAL ISOLATION IN SOI 550
11.8 SOI PRODUCT EVOLUTION 556
11.9 PROBLEMS ASSOCIATED WITH SOI 559
11.9.1 Floating Body Effects in Partially-Depleted SOI MOSFETs
11.9.1.1 Kink Effect in Partially Depleted SOI NMOSFETs
11.9.1.2 Transient I_D Effects
11.9.1.3 Effect of Building Partially Depleted SOI MOSFETs Without Body Contacts
11.9.2 Short-Channel Effects in SOI MOSFETs
11.9.3 Self-Heating Effects in SOI MOSFETs
11.9.4 Reliability Issues of SOI
REFERENCES 568

Chap. 12 - MULTILEVEL INTERCONNECTS FOR ULSI 573

12.1 EVOLUTION OF THE STRUCTURE AND MATERIALS USED IN IC INTERCONNECTS 575
12.1.1 Single-Level Metal Interconnects in Early ICs
12.1.2 Double-Level Metal Interconnects in Early Bipolar ICs
12.1.3 Double-Level Metal Interconnects in CMOS ICs (Partial Planarization of the ILD)
12.1.4 Three-Level Metal Interconnects in CMOS ICs (CMP of the ILD & W-Plugs)
12.1.4.1 Problems of W-plug/Al-line Interconnect Structures
12.1.5 Dual Damascene Cu/Low-*k* Interconnects

12.1.6 Other Geometrical Aspects of Multilevel ULSI Interconnects
12.2 PERFORMANCE REQUIREMENTS OF ULSI INTERCONNECT SYSTEMS 586
12.2.1 Propagation Delays in ULSI Multilevel Interconnects
12.2.1.1 Case When the *k*-Value of the ILD is Constant Throughout the Interconnect System
12.2.1.2 Case When the *k*-Values of the Interlevel and Intralevel ILD are Different
12.2.1.3 Minimizing the *RC* Delay in the Local Interconnect Region
12.2.1.4 Minimizing the *RC* Delay in the Global Interconnect Regions
12.2.2 Power Dissipation Issues in ULSI Interconnects
12.2.3 Crosstalk in ULSI Interconnects
12.2.4 Contact Resistance in ULSI Interconnects
12.2.5 Electrostatic Discharge Protection Issues of Interconnects
12.3 SUMMARY OF USLI INTERCONNECT ISSUES AND FUTURE TRENDS 599
REFERENCES 601

Chap. 13 - POLYCIDES AND SALICIDES OF $TiSi_2$, $CoSi_2$ AND NiSi 603

13.1 POLYCIDES AND SALICIDES 603
13.2 POLYCIDES FABRICATED BY CVD OF TUNGSTEN SILICIDE (WSi_x) 605
13.3 SALICIDE (SELF-ALIGNED SILICIDE) STRUCTURES 608
13.4 THE EFFECTS OF SALICIDES ON MOSFET PERFORMANCE 612
13.5 SALICIDES FORMED WITH TITANIUM SILICIDE ($TiSi_2$) 614
13.5.1 Reaction Kinetics of $TiSi_2$ Formation (and the Problem of *Bridging*)
13.5.2 The $TiSi_2$ Narrow-Line Effect
13.5.3 Proposed Process Modifications to Mitigate the $TiSi_2$ Narrow-Line Effect
13.5.4 Junction Dopant Redistribution and Silicon Consumption During $TiSi_2$ Formation
13.6 SALICIDES FORMED WITH COBALT SILICIDE ($CoSi_2$) 624
13.6.1 Advantages and Drawbacks of Cobalt Salicides
13.6.2 Details of the Cobalt Salicide Fabrication Process
13.6.3 Diode Leakage Issues Involving Cobalt Salicides
13.6.4 Sensitivity of Cobalt Salicide Films to Degradation by Oxygen Ambients: TiN Capping of Cobalt Salicide Films
13.7 EPITAXIAL $CoSi_2$ 630
13.8 $CoSi_2$ AS A SOURCE/DRAIN DIFFUSION SOURCE 631
13.9 NICKEL SILICIDE FILMS FOR SALICIDE APPLICATIONS 634
REFERENCES 635

Chap. 14 - LOW-*k* DIELECTRICS **639**

14.1 INTRODUCTION TO LOW-*k* DIELECTRICS 635
- 14.1.1 Spin-On versus CVD Methods for Forming Low-*k* Dielectrics
- 14.1.2 Silicon-Based versus Carbon-Based Low-*k* Dielectrics
- 14.1.3 Status of Low-*k* Dielectrics in 2000

14.2 DESIRED CHARACTERISTICS OF LOW-*k* DIELECTRIC FILMS 646

14.3 GENERAL PROCESS INTEGRATION ISSUES INVOLVING LOW-*k* FILMS 648

14.4 FIRST GENERATION LOW-*k* DIELECTRICS ($2.8 < k < 3.5$) 650
- 14.4.1 Hydrogen Silsesquioxanes (HSQ)
- 14.4.2 Spin-On Methyl Silsesquioxanes (MSQ)
- 14.4.3 Fluorinated Silicate Glass (FSG) Films

14.5 SECOND-GENERATION LOW-*k* DIELECTRICS ($2.5 < k < 2.8$) 655
- 14.5.1 2nd-Genereation Spin-On-Dielectrics: ($2.5 < k < 2.8$)
- 14.5.2 2nd-Generation CVD-Dielectrics: ($2.5 < k < 2.8$)
  - 14.5.2.1 2nd Generation Low-*k* CVD-Films: Organo-Silicate Glasses
  - 14.5.2.2 2nd Generation Low-*k* CVD-Films: *FlowFill* CVD Films
  - 14.5.2.3 2nd Generation Low-*k* CVD-Films: Black Diamond & Coral CVD Films
  - 14.5.2.4 2nd Generation Low-*k* CVD-Films: Parylene CVD Films

14.6 ULTRA-LOW-*k* DIELECTRICS: ($k < 2.3$) 663

REFERENCES 669

Chap. 15 - DUAL-DAMASCENE INTERCONNECTS **671**

15.1 DAMASCENE VERSUS SUBTRACTIVE INTERCONNECT STRUCTURES 671

15.2 DAMASCENE INTERCONNECTS AS ALTERNATIVE DEEP-SUBMICRON INTERCONNECTS 672

15.3 SINGLE-DAMASCENE INTERCONNECT STRUCTURES

15.4 INTRODUCTION TO DUAL-DAMASCENE INTERCONNECT PROCESSES 674

15.5 THE THREE DUAL-DAMASCENE PROCESS SEQUENCES 676

15.6 THE EVOLUTION OF DUAL-DAMASCENE FROM SiO_2 TO LOW-*k* DIELECTRIC STACKS 679

15.7 TRENCH-FIRST DUAL-DAMASCENE PROCESS FLOW 683

15.8 VIA-FIRST DUAL-DAMASCENE PROCESS FLOW WITH AN SiO_2 DIELECTRIC STACK 685

639

; 646

648

650

655

) Films

671

2

4

676

679

15.9 PROCESS INTEGRATION ISSUES OF THE VIA-FIRST DUAL-DAMASCENE PROCESS FLOW WITH A LOW-*k*-DIELECTRIC STACK 687
15.9.1 Via-First-Damascene Integration Issues: Lithography
15.9.1.1 Via-Patterning Lithography Issues
15.9.1.2 Trench-Patterning Lithography Issues – DUV Resist Poisoning
15.9.1.3 Trench-Patterning Lithography Issues – Trench CD Control Issues
15.9.1.4 Trench-Patterning Lithography Issues – BARC Veils
15.9.1.5 Bi-Layer Resist Process
15.9.2 Via-First-Damascene Integration Issues: Dielectric Etching
15.9.2.1 Via-Etching for Low-*k* Processes
15.9.2.1 Trench-Etching
15.9.2.1 Etching the SiN Barrier at the Bottom of Vias
15.9.3 Via-First-Damascene Integration Issues: Resist Stripping
15.9.4 Via-First-Damascene Integration Issues: Cleaning
15.9.4.1 Post-Dielectric-Etch-and-Strip Residue-Removal
15.9.4.2 Pre-Metal-Deposition Clean
15.9.4.3 Post-CMP Clean
15.9.4.4 Clean of Top-Surface of Cu Prior to Diffusion-Barrier-SiN-Deposition Process
15.9.4.5 Supercritical Fluids as Wafer Cleaners
15.10 VIA-FIRST DUAL-DAMASCENE DIELECTRIC STACK WITH NO EMBEDDED ETCH-STOP LAYER 708
15.11 PROCESS INTEGRATION ISSUES OF DUAL-DAMASCENE DIELECTRIC STACK WITH ULTRALOW-*k* DIELECTRICS 708
REFERENCES 709

Chap. 16 – COPPER INTERCONNECT PROCESS TECHNOLGY 711

16.1 WHY COPPER FOR DEEP-SUBMICRON IC INTERCONNECTS? 712
16.1.1 The Advantages of Copper Interconnects
16.1.2 Technological Challenges of Using Copper Interconnects
16.2 OVERVIEW OF COPPER PROCESS TECHNOLOGY 723
16.2.1 The Transition from Aluminum to Copper Interconnects
16.2.2 The Key Goal of Depositing Copper into Damascene Recesses
16.3 BARRIER LAYERS FOR COPPER INTERCONNECTS 728
16.3.1 The Impact of the Barrier-Layer Deposition Process on the Filling of Damascene Recesses with Copper
16.3.2 Introduction to the Technology of Depositing Barrier-Layers for Cu Interconnects

16.3.3 Ionized Magnetron Sputter Deposition
16.3.4 Hollow-Cathode Magnetron Sputtering Source
16.3.5 Self-Ionized-Plasma Sputtering Source
16.3.6 Material Properties of Tantalum and Refining Tantalum Metal For Semiconductor Applications
16.3.6.1 Material Properties of Tantalum
16.3.6.2 Refining and Forming Tantalum Metal For Sputtering Targets
16.3.6.3 The Properties of Tantalum (and TaN & TaSiN) as Cu-Barrier Layers
16.3.7 Other Candidate Materials for Cu-Barrier Layers
16.4 SEED-LAYER TECHNOLOGY FOR COPPER INTERCONNECTS 740
16.5 COPPER INTERCONNECT FILM DEPOSITION 747
16.5.1 Electrolytic-Plating of Copper
16.5.2 Filling High-Aspect-Ratio Recesses Without Voids Using a Cu Electrolytic Electroplating Process
16.5.2.1 The Electroplating Process Sequence
16.5.2.2 Additives in Plating Baths for Damascene Cu Interconnects
16.5.2.3 Models of How Additives in Plating Baths Impact "Bottom-Up" Filling
16.5.2.4 Electroplating Bath Control
16.5.2.5 Pulsed Plating
16.5.2.6 Copper Thickness Variation Within a Wafer After Electroplating
16.5.3 Post-Electroplating Cu Annealing and Recrystallization
16.5.4 Electrochemical Mechanical Deposition (ECMD™)
16.5.5 Electroless-Plating of Copper
16.5.6 CVD of Copper
16.6 WET-CHEMICAL ETCHING AND ELECTROPOLISHING COPPER 775
16.7 CONTAMINATION CONTROL IN COPPER TECHNOLOGY 778
16.7.1 Sources of Copper Cross-Contamination in a Wafer Fab
16.7.2 Effects of Copper-Cross Contamination at the Back-End-of-Line (BEOL)
16.7.3 Cu Removal from the Wafer Bevel/Edge and Backside
16.8 RELIABILITY OF COPPER INTERCONNECTS 784
16.8.1 Reliability of Copper Interconnects: "Missing-Metal" Defects
16.8.2 Reliability of Copper Interconnects: Plated vs. CVD Cu Films
16.9 ENVIRONMENTAL, HEALTH, & SAFETY ISSUES OF CU TECHNOLOGY 787
16.9.1 Copper Waste Handling
16.9.1.1 Handling Concentrated Waste from Plating and Wafer Reclaim
16.9.1.2 Handling Copper CMP Effluents
REFERENCES 789

INDEX **795**

ayers

'S 740

Filling

PER 775
778

ne (BEOL)

CHNOLOGY 787

795

PREFACE

With the dawn of the new millennium, IC technology-nodes have evolved beyond 0.18-micron. New processes also had to be developed to meet the needs of this *Deep-Submicron Era,* including: copper/low-*k* dual-damascene interconnects; high-*k* MOSFET-gate and DRAM dielectrics; 300-mm wafers; chemical-mechanical polishing (CMP); excimer-laser-based DUV-lithography; chemically-amplified (CA) resists; silicon:germanium (Si:Ge); and silicon-on-insulator (SOI).

The semiconductor industry, however, also experienced an economic decline of unprecedented severity at the start of the 2000 decade. It seams clear that to fuel the recovery, chip manufacturers will have to exploit the above innovations. Such processes are a path to achieving the performance goals of future IC technology-nodes that will be needed for manufacturing new and profitable electronic products.

This is the *first available book dedicated exclusively to describing such deep-submicron process technologies* (including those listed above). It is intended to provide semiconductor engineers and researchers with a comprehensive, state-of-the-art reference about these emerging and leading-edge processes. While many other books have been written about the "traditional" IC fabrication processes, none offer what this volume contains. It's sole focus is those techniques needed to *manufacture chips for 0.18-μm (and smaller) generations*. Thus, it belongs on the shelf of every microelectronic professional whose livelihood depends on having a thorough grasp of the latest advances in IC fabrication techniques. There is no faster way to gain this information. As with our other volumes, owners will find themselves referring to this text again and again!

Readers can still consult Vol.1, 2nd Edition for background material on basic IC fabrication modules, as this book (Vol. 4) does not repeat that information. Instead, it presents over 800 pages of new content about the process technologies of the 2000's. It also has over 900 references (with 200 of these being published since 2000), and more than 500 illustrations. It is written in the same lucid, user-friendly style that has been widely praised in the other volumes of *Silicon Processing for the VLSI Era.*

As with the previous volumes, this text was produced using an accelerated book-production schedule. By employing revolutionary publishing techniques made available in the 1990's, Lattice Press can turn a finished manuscript into a published text in less than 3 months (a task that takes most publishers of technical books 15-18 months). This allows our books to be released faster, and to contain the most-up-to-date information at the time of publication.

SILICON PROCESSING FOR THE VLSI ERA is a multi-volume treatise designed to provide a comprehensive and up-to-date treatment of this important and rapidly changing field. The volume in hand is the fourth in the series. Volume 1 (*Process Technology*, 2nd Edition) was published in 2000 and deals with individual processes employed in the fabrication of silicon ULSI circuits (such as silicon wafering, thermal oxidation of silicon, CVD and PVD of amorphous and polycrystalline films, diffusion and ion implantation of dopants in silicon, microlithography, and patterning technology). Volume 2 (*Process Integration*) was released in 1990, and describes how the individual processes of Volume 1 are combined in various ways to produce MOS and bipolar VLSI and ULSI circuits. The task of integrating these various fabrication processes together is referred to as *process integration.* In Volume 3 (*The Submicron MOSFET,* published in 1995) the topics of submicron MOSFET device physics and the relationship between such device physics and submicron MOSFET fabrication are addressed.

A book of this length and diversity would not have been possible without the indirect and direct assistance of many others. From an indirect perspective, virtually all of the information presented here is based on the research efforts of countless numbers of scientists and engineers. Their contributions are recognized to a small degree in citations of their articles given in the references at the end of each chapter. The direct help was also provided by many people, and the resulting text is thus significantly better. The author expresses heartfelt thanks to those who gave their time, energy, and intellect.

Important direct help came from members of my immediate family. My wife, Carrol, read the manuscript for grammatical and stylistic correctness. My son, Ross, designed the cover and improved many of the graphics. My daughter, Jennifer, helped in marketing the book. Thus, it truly represents a family enterprise.

Stanley Wolf

For additional copies of the books contact:

LATTICE PRESS
Post Office Box 340
Sunset Beach, CA 90742 USA
www.latticepress.com

For your convenience an order form is provided on the final leaf of the book.
It can also be ordered from the **LATTICE PRESS** website.

, Formed
247.
romise,"

acitors,"

)RAMs,"

tric Film

'roperties

l CMOS

yer HfO_2
'ech. Dig.

electric,"

h-*k* Gate

ıd – The

Chapter 5

THE STRUCTURE OF DEEP-SUBMICRON MOSFETS: Retrograde-Wells; Super-Steep Retrograde Channel Doping; Drain Engineering; Punchthrough-Control Structures

In Chapter 1 we discussed the evolution of MOSFET structures down to the submicron technology nodes in production at the time publication of this text. In this chapter we delve further into the details of the MOSFET regions that are fabricated in such deep-submicron MOSFETs, namely: **1**) retrograde wells for CMOS twin-well technology; **2**) super-steep retrograde (SSR) channel doping profiles; **3**) drain engineering issues in deep-submicron MOSFETs; and **4**) punchthrough-control structures.

5.1 WELL-FORMATION IN DEEP-SUBMICRON CMOS

Twin-well CMOS is now the dominant CMOS technology in IC production. In twin-well CMOS each well is formed separately by implanting the well dopants into a lightly doped substrate region. Having two well implants allows each well-type to be tailored separately, permitting independent device performance adjustments for both PMOS and NMOS transistors. When first introduced, twin-well CMOS structures were fabricated by using moderate-energy implants to place both types of well dopant near the wafer surface. These dopants then had to be driven into the substrate by a long (12-24 hour), high-temperature (>1100°C) diffusion process (Fig. 5-1). Such a deep drive-in step was needed to prevent punchthrough from the MOSFET drain regions at the surface to the substrate under the well. This type of CMOS-well-structure is referred to as a *diffused-well.* However, use of diffused-well structures gives rise to some problems which limits their use when CMOS is scaled below 0.5 μm:

1. Lateral (as well as vertical) diffusion occurs during the diffusion step, which has the effect of reducing packing density. For example, since the lateral diffusion is ~0.7 times the junction depth, a typical diffused well having a vertical junction depth of 3-4-μm, results in a lateral spread of well dopant of 2-3-μm. Consequently, the maximum aspect ratio of a diffused well is limited to about 1.6:1 (i.e., wells narrower than about 1.6 times their depth would intermingle with the dopant of adjacent wells).

2. If LOCOS isolation is used, boron from the *p*-well-dopant implant-step segregates into the oxide during the long, high-temperature field-oxide growth step. Compensating for this effect requires a separate field-threshold-adjust implant (also referred to as a *channel-stop-implant*), prior to the field-oxidation step. This additional step adds process complexity.

3. Diffusion of the channel-stop-implant dopant into the active regions during field oxidation adversely affects both the junction-capacitance and the turn-off characteristics of the NMOSFETs (see Vol. 3).

4. The lengthy, high-temperature drive-in-step increases process cycle time, and increases the risk of silicon wafer warpage, bow, and flatness degradation (and the likelihood of metal contamination from furnace heating elements).

5. The diffused-well profile for a particular technology node is the net result of compromises between the need to control threshold-voltage, to suppress hot-carrier effects, to control punchthrough and soft-error effects, and to provide latch-up immunity. For example, a high dopant concentration is desirable deep in the well to strengthen latch-up immunity. This cannot be achieved in a diffused-well because doping at the bottom of such a well must be lower than at the surface, and the maximum surface doping is set by limits on junction breakdown voltage and junction capacitance.

For all of the above reasons, the diffused-well approach to forming a twin-well CMOS process becomes less effective for each smaller technology node. As a result, an alternative method of forming the wells was sought. The technique adopted to replace the older diffused-well structure when CMOS technologies are scaled below about 0.5-μm involves *retrograde-well formation* with *high-energy ion-implant steps*.

5.1.1 Retrograde-Wells for Deep-Submicron CMOS

The medium-energy implant steps used to introduce well-dopants into the substrate in the diffused-well approach (50-100-keV), can be replaced with high-energy-implant steps (300-keV to 1-Mev). Such high-energy implants cause the peak of the implant to be located considerably deeper within the silicon substrate (at roughly 1-μm for the energies most frequently used in high-energy implant processes). Since the implant peak is now far beneath the silicon surface, the impurity concentration thus decreases as the doping profile approaches the wafer surface. Since the well-profile in this case is different from that of a conventional

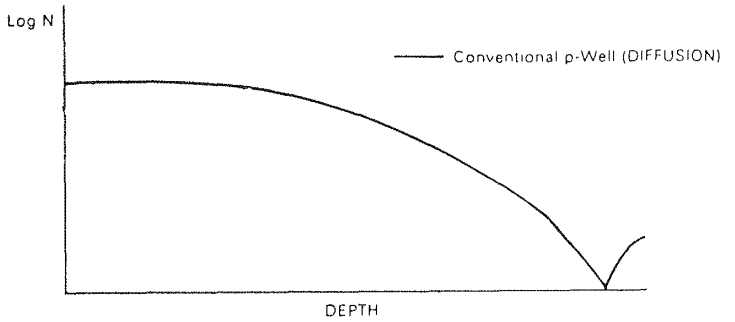


Fig. 5-1 Doping profile in a conventional well formed by a shallow implant and drive-in step.

diffused well (in which the doping concentration after drive-in is highest at the surface, and decreases monotonically with depth, as seen in Fig. 5-1), such deeply implanted wells are known as *retrograde-wells*.* Comparison of the doping profiles in conventional and implanted *p*-well profiles (obtained by simulation) is shown in Fig. 5-2.[2,3] The retrograde-doping-profile is retained by minimizing the temperature-cycles of later process steps (e.g., by using RTP instead of furnace anneals).

The retrograde-well approach to forming the twin-well CMOS-structure has several major advantages over the diffused-well process:

1. Much less lateral-dopant-spread occurs for a retrograde-well having the same junction depth as a diffused-well. This is because ion implantation is a more anisotropic dopant-distribution process than diffusion (i.e., the lateral-straggle for implant energies of 500-keV is only ~0.7-μm). In addition, little additional diffusion occurs after the implant is performed, since only a short anneal at a relatively low-temperature is needed to activate the implant. Thus, the total lateral-dopant-spread barely exceeds 1-μm after all thermal processing (compared to at least 2-μm for the same well-depth using a diffused-well process).[2] Figure 5-3 depicts the doping concentration profile of a high-energy implanted *p*-well. In this example, the implanted *p*-well is formed by boron implantation at 400-keV.

In addition, high-energy implanters can place dopant beneath a shallow-

* In order to avoid excessive spread of the active-device electrical parameters, the surface concentration after the retrograde implant should be at least five times smaller than its final value - which will be set by subsequently performed, shallower low-dose implants specifically designed for threshold and punchthrough control.

trench-isolation oxide or a LOCOS field oxide, allowing the CMOS twin-well structure to be implanted *after* the dielectric isolation features are formed. (In diffused-wells, the well-dopants must be implanted before the field-oxide, and driven-in either before - or during - the field-oxidation step). Thus, since the high-energy implanted wells are formed after the long, high-temperature cycle needed to grow the LOCOS field-oxide, lateral-spread of the well-dopants is further reduced. (A higher threshold voltage can also be achieved in the LOCOS field-regions of the p-wells if the boron implant is done following field-oxidation, because the boron does not segregate out into the oxide during the field-oxide growth-step.[1]) Hence, the 7-9-μm n^+-to-p^+ spacing required for a twin-well 1.2-μm CMOS process which uses conventional wells, can be reduced to ~5-μm if retrograde wells are used instead. Such overall reduced lateral-dopant-diffusion permits higher circuit-packing-densities.

2. By combining the high-energy-implant step with the other lower-energy MOSFET channel implants (V_T-adjust and punchthrough stopper) and since minimal thermal diffusion occurs, control of dopant levels throughout the depth of the well becomes possible with fewer tradeoffs (a task that has been dubbed "Well Engineering"). This permits optimization of such key

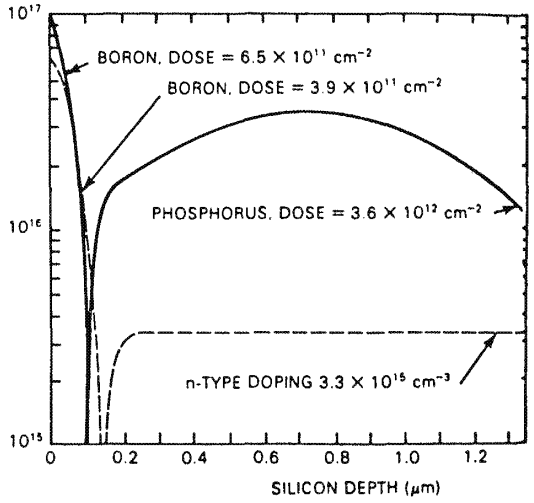


Fig. 5-2 Simulated doping profiles in PMOS channel regions: dashed lines, lightly doped substrate; solid lines, retrograde n-well.[3] (© 1982 IEEE).

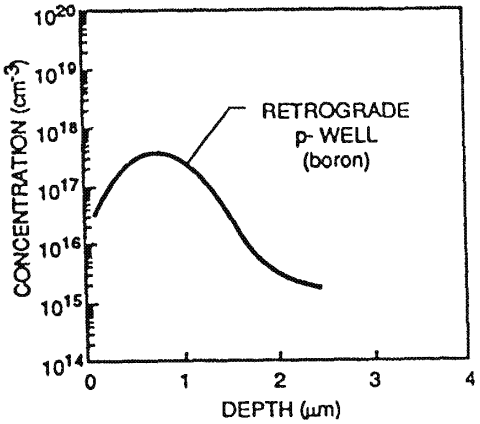


Fig. 5-3 Example of the doping profile of a retrograde *p*-well (boron) implant (400 keV).

MOSFET performance characteristics as: device speed (carrier mobility and, S/D capacitance), inter-device leakage-currents, V_T-control, punch-through suppression, parasitic-transistor effects, and soft-error and latch-up resistance. That is, low surface-doping levels increase mobility and decrease capacitance. Heavier doping below the surface (retrograde doping) suppresses minority carrier collection by decreasing depletion widths; thus reducing soft errors and latchup susceptibility. This degree of flexibility in optimizing device parameters is not possible with diffused-wells.

Besides the primary benefits of increased packing-density and greater process flexibility, retrograde-wells offer the following secondary advantages:[5]

1. A retarded electric field is created in the parasitic-vertical-bipolar-transistor, thereby reducing its current gain. This affords some extra protection against latchup.
2. Susceptibility to vertical-punchthrough is reduced without having to significantly raise the surface-doping-concentration.
3. The conductivity in the bottom of the well is increased, which also provides additional resistance against latchup.
4. The shallower well depth should also make the drain of the parasitic n^+-to-n-well (or p^+-to-p-well) parasitic-FETs shallower. This contributes to improvements in both the short-channel threshold-roll-off-effect and the in the punchthrough characteristics of these parasitic field-region devices (see Vol. 3, Chap. 6).

5. The formation of retrograde-wells with high-energy implants results in a reduced-doping-concentration under the bottom of the source and drain regions, which reduces the source/drain-to-body junction-capacitance.

However, some concerns associated with the retrograde process should also be mentioned. First, the process requires a thick photoresist as a masking material for the high-energy implant. Typical resist thicknesses are 2.0-2.5-μm, and special exposure and development techniques are required for this process. Implant shadowing arising from the use of such thick resists must also be addressed. (We will address the challenges associated with the use of thick resists as high-energy-implant masks in more detail at the end of this section.) Questions have also been raised about the possibility that the high-energy implants could cause damage that may degrade device and circuit performance.

5.1.2 Process Integration Issues Involving the Use of Retrograde Wells

Advanced twin-well CMOS processes with both retrograde *n*- and *p*-wells (using short [~5 min] RTP well-drive-in steps), were reported for submicron applications as early as 1991.[6] Further advances continued throughout the decade. Scaling CMOS to 0.5-μm and below reduces the well depths required to prevent vertical punchthrough and latchup. In turn, as device and isolation dimensions shrink with each device node, retrograde-well implant-energies for both DRAM and logic circuits are also gradually being reduced. However, the retrograde-well implants used for DRAM devices typically require higher doses than those used in logic devices (i.e., to achieve better minority-carrier suppression [soft-error rate]). This, in turn, requires a higher well-implant-energy (to avoid excessive dopant in the near-surface [channel] region). In addition, the well-depth and well-dopant-concentration can be reduced if additional, lower-energy implants are precisely located with respect to the transistor structure (to suppress punchthrough of the MOSFET channels and enhance the field-isolation characteristics).

The high-energy implant step used to form retrograde-wells in deep-submicron-CMOS is typically performed together with the V_T-adjust, punchthrough, and channel stop (isolation) implants. These implants are all done in the same implanter, without removing the wafers from the implant process chamber in between (unless channel engineering implants [that use high wafer-tilt-angle for precision dopant placement] are required - see Sect. 5.4). Such a "chained" implant sequence is performed by doing the highest-energy implant first, with the lowest-energy implant being done last. This ensures that no ions traverse the end-of-range (EOR) damage of a previous implant. Note that because

of economic requirements for high equipment utilization, this sequence must be carried out in an implanter that has the capability of doing not only high-energy implants (e.g., up to about 1-MeV for the retrograde n-well implant), but also lower-energy boron implants (down to about 10-keV, for the NMOS V_T-adjust implant).

An example of a "chained implant" retrograde twin-well CMOS process is given in Ref. 7. Here, a retrograde n-well is formed by a 600-keV, $3x10^{13}$ cm^{-2} phosphorus implant. Then, a second phosphorus implant (150-keV, $4x10^{12}$ cm^{-2}) is used to obtain both punchthrough prevention and intra-field isolation (channel stop), and a third phosphorus implant (70-keV, $6x10^{12}$ cm^{-2}) sets the threshold voltage. In an identical manner, three selective boron implants are placed in the p-well region. The retrograde p-well implant is boron at 300-keV, $3x10^{13}$ cm^{-2}. Then a punchthrough/channel-stop-implant at 150-keV, $4x10^{12}$ cm^{-2} and a threshold-adjust-implant at 30-keV, $5.5x10^{12}$ cm^{-2} are added. The process displays very good inter-well isolation characteristics, which allows a minimum p^+-*to*-n^+ spacing 40% lower than the same process without retrograde wells. It also achieves very good intra-well isolation, allowing n^+-to-n^+ (and p^+-to-p^+) device separation of 0.5-μm, without an additional channel-stop implant.

From a manufacturing perspective, the well-formation sequence thus appears to be a single implant step, reducing manufacturing complexity and cost. Figure 5-4 compares an advanced diffused-well process with a high-energy implanted process up to gate oxidation, showing the process simplification possible with "well-engineering." Eliminating process steps has other benefits besides reducing cycle time. These include a reduction in work-in-progress (WIP) inventory, less wafer handling, and fewer defects.

The cycle-time and cost savings obtained from such a "chained-implant" sequence was assessed by SEMATECH. They sponsored a comparison of a baseline diffused-well process with a high-energy retrograde-well process at two different chip manufacturers. The high-energy processes were engineered to achieve minimum cycle-time and maximum cost savings, while maintaining the same basic device and circuit parametric data (and thus were identical to the baseline processes after gate-oxidation). Significant improvements in cycle-time, cost, and device-performance were achieved by both companies.

By chaining these implants, the doping-profile in the well-regions of the MOSFETs has a *modulated* characteristic as depicted in Fig. 5-5 (where the term "modulated" means that the doping-concentration-profile has several maxima and minima).[7] The incorporation of such multiple implants to form a modulated well has become a firmly established processing approach in chip production.

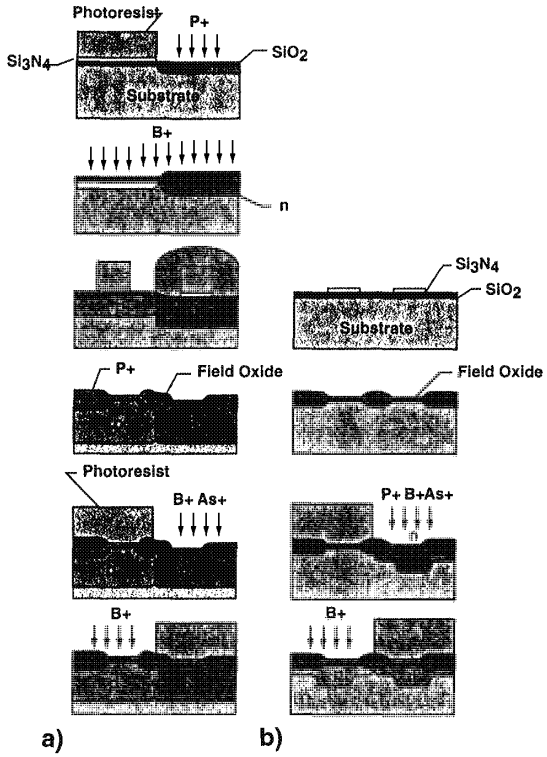


Fig. 5-4 Fabrication sequences for: a) conventional-diffused-well and; b) high-energy-implanted-well processes.

Another issue of using high-energy implants to introduce the well and isolation dopants involves implant shadowing. That is, as shown in Fig. 5-6a, thick photoresist films must be used as masks against the high-energy implant (e.g., 2.0-2.5 μm thick). The doping asymmetry resulting from high energy implants with a small (7°) tilt angle can have significant effects on retrograde well structures and inter-well isolation. Figure 5-7 depicts the simulated doping profiles from a 1 MeV phosphorus implant with a 7° tilt.[8] The simulation accounted for the fact that opposite corners of the photoresist will wear down and carbonize differently because of the asymmetry in the incoming ion beam. As the sidewall shape of the photoresist changes with time, this will affect the implanted profile in the silicon. The doping profiles in Fig. 5-7 clearly show a bulge of lightly doped material on the shadow side of the profile. More seriously, on the

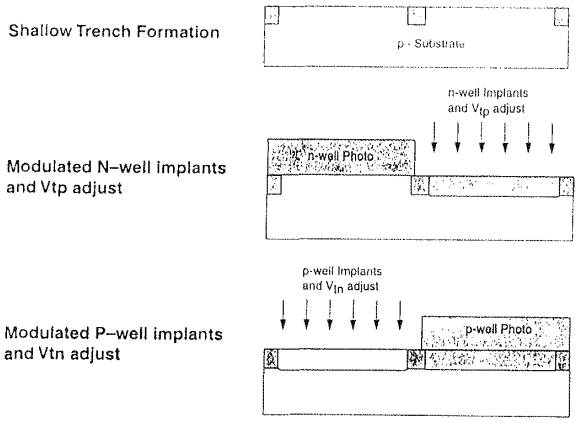


Fig. 5-5 Modulated *n*-well structure for advanced devices, showing the functions of each implant.[7]

incoming side, a channel doped to 20% of the peak-concentration extends from the projected range up to the surface. Doping variations such as these can adversely affect the mobilities and junction-capacitances of the resulting devices. This will also make it harder to maintain adequate isolation structures as isolation widths scale below 400-nm. These problems can be mitigated by reducing the tilt-angle to the 2-4° range, using 0°-tilt. However, using 0°-tilt-implants is not always possible, due to channeling effects (which are severe at high energies).

An alternative solution to the implant shadowing problem during high-energy

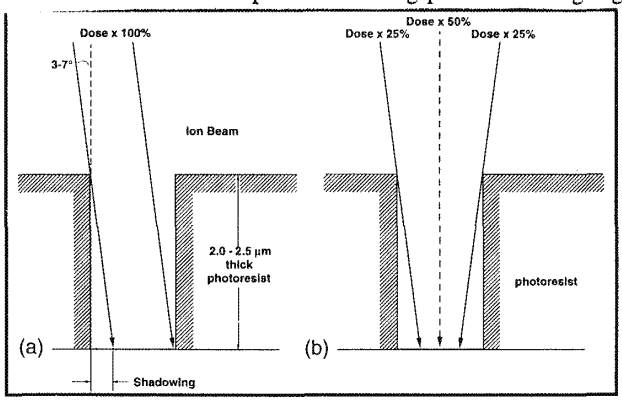


Fig. 5-6 The geometric effects of ion implantation into a narrow active region: a) low-tilt-angle shadowing and well misalignment; b) low-tilt quad implant provides both lateral and vertical (no channeling) uniformity.

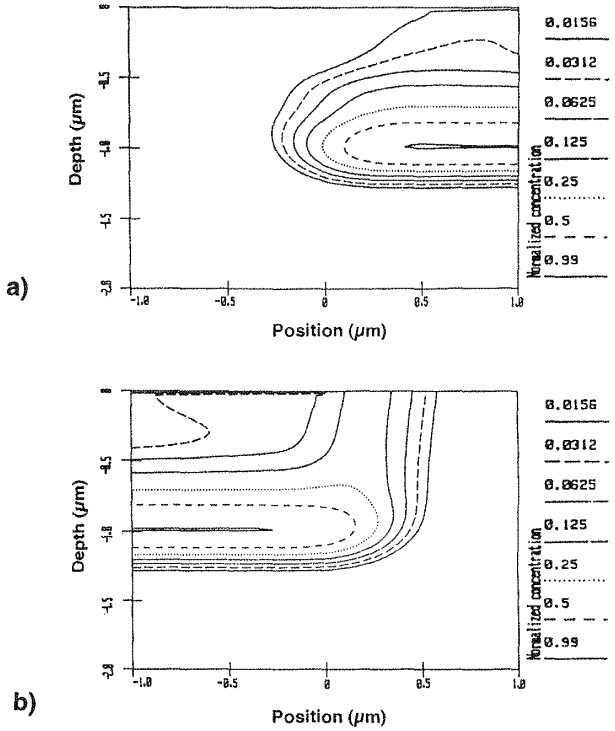


Fig. 5-7 Normalized doping profile of 1 MeV phosphorus implantation. a) On the shadow side (a) a lowly doped bulge is found (according to Monte Carlo simulation). b) On the incoming side, Monte Carlo simulation shows a trunk.[8]

7°-tilt implants is to reposition the wafers part-way through the implant process. If the wafers are repositioned four times, and 25% of the total dose is implanted with the wafer in each position (*quad implant*), all the areas of the exposed silicon surface will receive the same implant dose, eliminating the implant-shadowing problem (Fig. 5-6b). That is, well profiles are symmetric, both laterally and vertically, and shadowing effects near the surface are compensated.

A high-energy batch-implanter capable of performing low-tilt quad-implants without unloading the wafers from the disk addresses the isolation scaling problem with high productivity. Since it is not practical to scan a high-energy ion beam, implanters with fixed beams and batch end-stations using simultaneous disk spinning and scanning is the method-of-choice for process uniformity. The

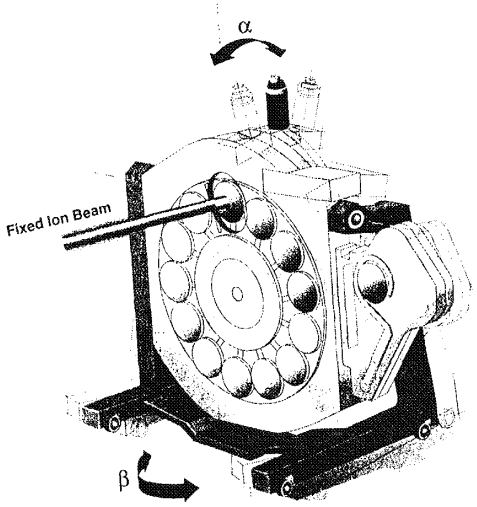


Fig. 5-8 *In situ* quad implants can be run on a batch end-station with orthogonal tilt axes (designated α and β).

quad-implant capability is achieved with a disk chamber as shown in Fig. 5-8.[9] Here the disk can be independently tilted in two orthogonal directions. Such batch high-energy implanters are sold by Axcelis Corp. Single-wafer high-energy implanters are also commercially available. A model offered by Applied Materials also permits tilting the wafer and repositioning it several times during an implant step. Higher tilt-angle control is claimed for such single-wafer high-energy implanters compared to the batch types.[10]

5.2 SUPERSTEEP RETROGRADE (SSR) CHANNEL DOPING PROFILES

The appropriate choice of substrate doping and engineering of doped layers under the gate are critical to optimum performance of deep-submicron MOSFETs (with respect to drain-to-source punchthrough, threshold voltage, and channel mobility). Specifically, the doping-density in the region between the source and drain must increase as the gate-length is reduced to prevent punchthrough. As shown in Fig. 5-9, the mean channel-doping-density needed to control punch-through when gate lengths scale below about 0.13-μm begins to exceed $5x10^{17}/cm^3$. However, to avoid reduction of carrier mobility in the channel, the doping at the surface must be kept below $5x10^{17}/cm^3$ to obtain the universal mobility curve.[11] To allow the surface concentration to be kept below this value, and still have a mean doping concentration in the channel above this number, a vertical *retrograde doping concentration profile* is needed in the channel region.

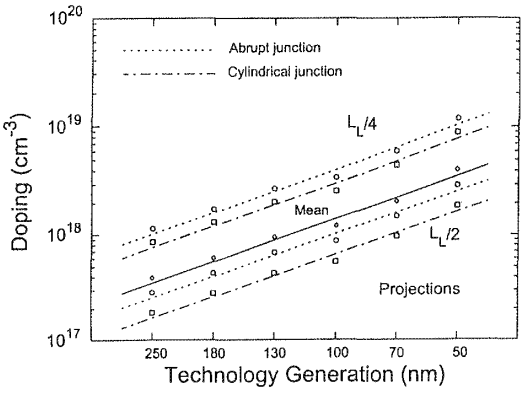


Fig. 5-9 Doping density required for punchthrough control.[11]

Assume that a lighter doped surface layer with a concentration of $1x10^{17}/cm^3$ exists down to a depth X_I (taken to be approximately 1/2 of the source/drain junction depth X_j), and that at that depth the doping abruptly changes to the value needed for the heavily doped punchthrough-control layer. Figure 5-10 shows the scaling required for X_I at each technology generation. Because the transition in doping concentration from the lighter doped region near the surface to the heavily doped punchthrough-control region below must occur over such a short distance, the gradient in the doping concentration must be very steep. Hence, such a doping concentration profile is termed a *super-steep retrograde* (SSR) profile.

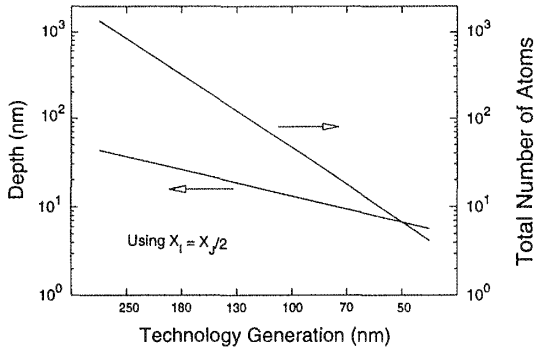


Fig. 5-10 Typical thickness of lighter-doped layer and total number of impurity atoms for $10^{17}/cm^3$ surface doping density.[11]

The formation of SSR channel-doping-profiles for technology nodes from 0.18-μm down to about 0.1-μm is done using ion implantation (in place of the conventional V_T-adjust implant step). In these implant processes, however, higher-mass dopants are substituted for the dopants typically used to carry out the V_T–adjust implants. That is, indium (In) is used instead of boron (B) for the NMOS devices, and either arsenic (As) or antimony (Sb) replace phosphorus (P) in PMOS devices. Heavier implant species provide steeper as-implanted profiles (due to their smaller straggle), and the heavier ions have reduced diffusivity (keeping the as-implanted profile from spreading out during subsequent thermal processes). Thus, well-profiles implanted with In and As (using implant energies in the 100-200-keV range) can be fabricated with sufficiently low doping (~$5x10^{16}/cm^3$) at the gate-oxide interface, and a higher peak doping level (~$1x10^{18}/cm^3$) 80-nm below the gate oxide. (Note that peak doping concentrations of indium cannot be made higher than about $10^{18}/cm^3$ because of the solubility limit of indium atoms in silicon.)

Figure 5-11a shows indium doping-profiles for various annealing temperatures, up to 1000°C.[12] It is evident that little change occurs in the as-implanted indium doping profile for annealing temperatures up to 1000°C, making it a promising candidate for forming SSR channel-doping-profiles. Indium implants for retrograde-channels in NMOSFETs have been reported in some cases to *improve* subthreshold-slopes relative to conventional boron implanted-channels,[12,13] while in another case it was found to have *degraded* the subthreshold-slope.[14] It is generally agreed that SSR doping-profiles are effective at reducing short-channel-effects and drain-induced-barrier-lowering, but it is not equally certain whether this translates into overall improvement in device performance (as this also depends on many other variables, including gate-oxide-thickness, gate-length, and junction-doping). Figures 5-11b and 11c, respectively, show the vertical concentration-doping-profile for an SSR and a conventional process for PMOSFETs (Sb for SSR, and P for the conventional process), and NMOSFETs (indium for SSR, and boron for the conventional process).[15]

The use of SSR channel-doping-profiles is entering some mainstream process flows, despite disagreements over their advantages compared to conventional channel-profiles. In late 1999, the successful fabrication of 100-nm-gate-length devices was reported. They were built using SSR channel-profiles, source/drain-extensions, and deep source/drain-regions for contact. A 100-nm-process yielded devices with 1.04-mA/μm (NMOS) and 0.46 -mA/μm (PMOS) on-currents at 1.5 V with 3-nA/μm off-current. This process uses indium and arsenic as the implant

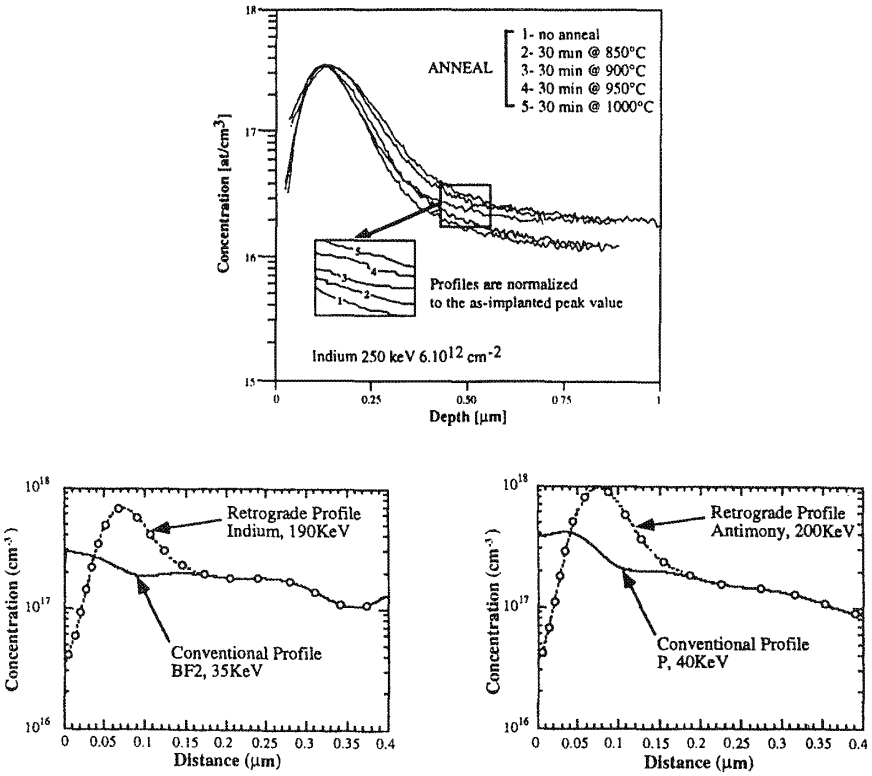


Fig. 5-11 a) SIMS data of indium profiles over a large range of annealing temperatures. Conventional and retrograde doping profiles (simulated) for b) NMOS and c) PMOS .[12]

species for the SSR-channel-doping-profiles, low-energy boron and arsenic source/drain-extensions, and high-angle pocket-implants.[16]

The source for indium ions in implant processes is $InCl_3$, a solid material at room temperature. The production of ion beams from solid $InCl_3$ presents some special challenges for ion implanters. First, $InCl_3$ is hygroscopic (i.e., it readily absorbs moisture), and thus may require excessive outgassing if not handled properly outside of the ion-source vaporiser. In addition, although $InCl_3$ is a solid, it has a low melting-point. Thus, indirect heating from a second vaporiser in the ion-source may cause unintentional vaporization. Both of these effects can lead to premature loss of the source charge, as well as the problem of indium

coating in and around the source area. Finally, even during normal operation, indium tends to drift down the beamline, covering the beamline components. This creates a risk of cross-contamination, since the indium atoms can be dislodged from these surfaces if they are struck by ion beams of other species.

Two ion-implanter features can adequately suppress such effects. First, active cooling in the source eliminates indirect heating, decreasing indium-loss and excess beamline-coating. Second, an electrostatic (or other) filtering element near the end-station greatly reduces the magnitude of such cross-contamination, allowing all species to be run on a single implanter. The latest generations of single-wafer, medium-current-implanters incorporate such features.[17]

5.2.1 Fabricating SSR Profiles for Sub-0.1 μm MOSFETs

When gate-lengths scale below 0.1-μm, it probably will no longer be possible to create SSR channel-doping-profiles using ion implantation (even if In and As are used as the implanted species). For example, if a lightly-doped surface-layer is established for a 50-nm-generation MOSFET (with a doping concentration of $1x10^{17}/cm^3$), the heavily-doped punchthrough-control layer ($10^{19}/cm^3$) will have begin at a depth of just 7-nm beneath the silicon surface. Since the width of the surface-inversion-layer is typically thinner than 7-nm, the channel electrons will still be located primarily in the lightly-doped surface-region. However, thin surface-layers with such an abrupt doping-concentration-increase cannot be fabricated with ion implantation (and its ancillary annealing processes). Other doping techniques, such as low-temperature selective-epitaxy (of perhaps Si-Ge layers[19]) will likely be needed for 50-nm and beyond MOSFETs.*

5.3 SOURCE/DRAIN ENGINEERING IN DEEP-SUBMICRON CMOS

In deep-submicron MOSFETs, the typical source/drain contact-structures consist of two regions: **1**) A shallow source/drain-extension-region (SDE), with a junction-depth X_j; and **2**) A deeper contact-source/drain-region, with a junction-depth X_{jC}. As also described in Chap. 1, when MOSFETs were first designed with such two-region source/drain-structures, the first-region had lighter doping than the second-region (and it was in fact called a *lightly-doped drain* (*LDD*) re-

* Figure 5-10 also shows a graph of N_T (the total number of impurity atoms in the lightly doped layer) versus technology generation. For 50-nm and below, this value approaches a small number (< 10), so that for cases where the doping density is important to device operation, any device parameter depending strongly on N_T will show large statistical fluctuations from device-to-device. If all parameters depend only very weakly on N_T then the layer may as well be undoped,[18] so that V_T-control relates to the positioning of the step (i.e., the increase in doping) and the doping in the heavily doped layer.

gion. This region was added to the single-region MOSFET drain-structures at about the time when MOSFETs were scaled below 1.5-μm. Its function was to reduce the peak electric-field (and thus hot-electron-injection) near the drain-end of the channel. However, as MOSFETs gate-lengths were scaled to the deep-submicron regime, the power-supply voltages also had to be reduced, and this resulted in a significant decrease in the peak electric-field near the drain. Hence, the LDD region was no longer needed to suppress hot-carrier-injection effects.

However, in deep-submicron MOSFETs the SDE-region remains a part of the drain structure. But, its role now is to reduce short-channel-effects. The drain contact-region is required to accommodate the thickness of the salicide contact-layer and minimize junction-leakage. Nevertheless, the junction depth X_{jC} of the contact-region should not be larger than that needed to meet these conditions. That is, the greater the junction-depth, the larger must be the width of the sidewall-spacer. This, in turn, will result in a longer SDE-region (which increases the value of the parasitic SDE-resistance, R_{SDE}). In Sect 5.3.1.3, we will see that to enable a MOSFET to deliver improved performance as the gate length is scaled, it is critical that the source/drain contact design should minimize both the parasitic series-resistance-effects and the salicide-substrate contact-resistance. In addition, to ensure that short-channel-effects are effectively controlled as the MOSFET is scaled, both X_j and X_{jC} will also have to appropriately scaled,.

Figure 5-12 shows the depths of the source/drain-region junctions and undoped channel-region as MOSFETs are scaled below 200-nm (where L_L is the lithographically-drawn gate-length), as taken from the ITRS 1999 study.[20] The design rules used in ITRS 2001 to calculate these same values have been relaxed slightly from those in the 1999 study: **1**) the source/drain-contact junction-depth is 1.1 L_L (±33%), compared to 0.8 L_L; **2**) the source/drain-extension junction-depth is 0.55 L_L (±25%), compared to 0.4 L_L; and **3**) the undoped region near the Si/gate-oxide interface in a super-steep retrograde doped channel (SSR) is 0.3 L_L (±25%), compared to 0.2 L_L. Creating and maintaining these ultra-shallow junctions remains a fundamental challenge for doping and metrology techniques.

In this section we will discuss two subjects related to the formation of the source/drain-structure in deep-submicron MOSFETs, namely: **1**) the issues involved in keeping the value of the resistance of the drain region R_S (which is the sum of the parasitic resistance components and contact resistance), to a value that is sufficiently low to meet the requirements of scaled-MOSFET device-design, and **2**) the process technology needed to form shallow SDE-junctions.

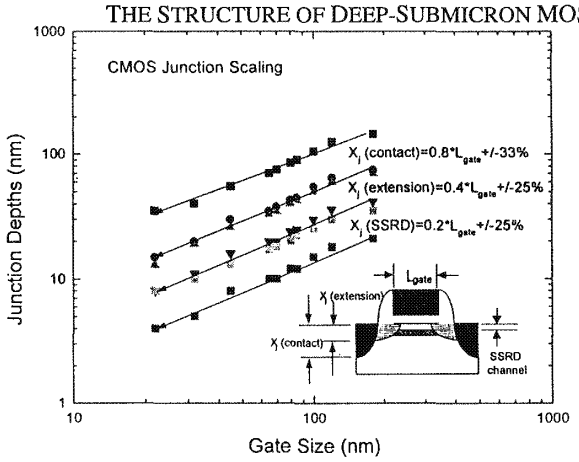


Fig. 5-12 Source/drain junction and undoped-channel depths for planar CMOS on bulk-Si.

5.3.1 Resistance of Deep-Submicron Source/Drains

In the planar MOSFET, the device-region in which transistor-action occurs is the *channel* under the gate. The current coming from the metal-interconnect lines must flow from the metal-Si-contact to the edge of the channel through the source or drain regions (Fig. 5-13). In the MOSFET, current enters the contact perpendicular to the wafer-surface, and then travels parallel to the surface in order to reach the channel. The total parasitic series-resistance, R_S, of the current path from the contact to the edge of the channel of a single-region drain (i.e., one

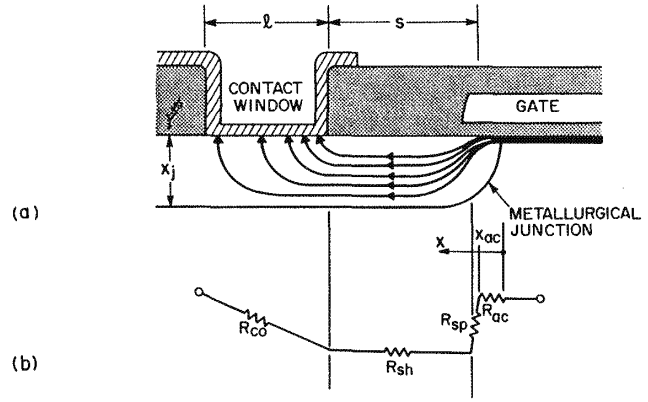


Fig. 5-13 Schematic drawing showing: a) current paths within the source and drain regions, of a planar MOSFET and; b) their corresponding representative series resistance components.[21] (© 1987 IEEE).

that does not have an LDD or a SDE) can be modeled by the sum of four components;[21]

$$R_S = R_{co} + R_{sh} + R_{sp} + R_{ac} \quad \textbf{(5.1)}$$

where R_{co} is the *contact resistance* between the metal and the source/drain region, R_{sh} is the *sheet resistance* of the bulk region of the source and drain, R_{sp} describes the *resistance of the current lines crowding near the channel end of the source*, and R_{ac} is the *accumulation-layer resistance.*

If the MOSFET drain has an SDE, the R_{sh} component of R_S can be replaced by R_{SDE}, as indicated in Fig. 5-14, so that R_S is now given by:

$$R_S = R_{co} + R_{SDE} + R_{sp} + R_{ac} \quad \textbf{(5.2)}$$

In any case, the value of R_S must be viewed from the perspective of how it affects the performance of a MOSFET. Specifically, its value must be compared to that of the channel-resistance R_{sat} of the MOSFET in its ON state:

$$R_{sat} = V_{DD}/I_{Dsat} \quad \textbf{(5.3)}$$

When the external resistance of the source + drain regions is equal to approximately 10% of the value of R_{sat}, the saturation-drain-current I_{Dsat} is reduced by about 8%. In most MOSFET design approaches, the goal is to keep I_{Dsat} from being reduced by no more than 10% due to such series-resistance of the source and drain regions. Thus, it is useful to analyze the components of R_S, so that it can be understood what factors can be adjusted to keep the value of R_S to the value which will meet the I_{Dsat}-degradation design goal.

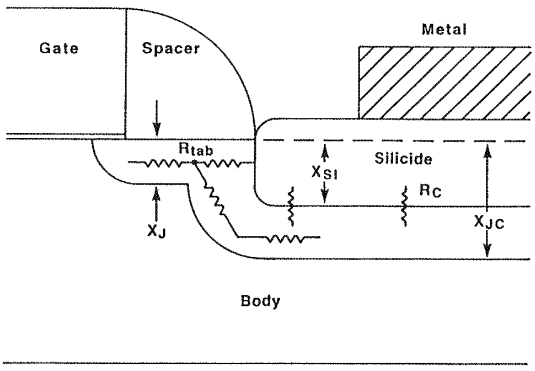


Fig. 5-14 Schematic drawing of the components of source or drain resistance in a deep-submicron MOSFET.

For future scaling, the value of R_{sat} (actually $R_{sat}W$, normalized per unit device width) is given by:

$$R_{sat}W \cong V_{DD}t_{ox}/(0.47v_{sat}\varepsilon_{ox}\,[V_{DD} - V_T]) \cong t_{ox}/(0.376v_{sat}\varepsilon_{ox}) \qquad \textbf{(5.4)}$$

Equation 5-4 indicates that since t_{ox} will continue to be made smaller as MOSFETs are scaled, the value of $R_{sat}W$ will also decrease as the devices are scaled. This implies that the value of R_S will also have to be made lower with each technology node. Before we examine in detail the implications of a decreasing R_{sat} (and therefore also a decreasing R_S), we first need to give a more thorough discussion of the physical meaning of the parasitic resistance components of the drain contact region, namely R_{ac}, R_{sp}, and R_{SDE}.

5.3.1.1 Physical Meaning of R_{ac}, R_{sp}, and R_{SDE}: Let us discuss the physical origins of the three parasitic resistance components of the drain/source-regions, with respect to deep-submicron NMOSFET structures. Figure 5-15 depicts the doping gradient of an n^+-p junction at the source (or drain) end of an NMOSFET. The dotted lines indicate contours of constant doping concentration. We define the end of the MOSFET channel as the location where the metallurgical junction occurs (i.e., where $N_D = N_A$ at the surface). When the MOSFET is turned ON, an inversion layer is induced along the length of the channel (i.e., in the substrate-region - where for this NMOSFET the doping is p-type). However, because the gate is an equipotential region, the positive bias needed to turn the MOSFET ON, will have a different effect on the silicon region under the gate if it is doped n-type. That is, in the region where the *drain doping* overlaps under the gate edge, an accumulation-layer (instead of an inversion-layer) is formed (dark region in Fig. 5-15).

The length of the accumulation-region is determined by the point where the induced (accumulated) carrier-concentration at the surface equals the background doping level. For example, a typical bias induces ~10^{19}/cm^3 carriers in the accumulation layer while the maximum doping level in the source/drain region is in the 10^{20}-10^{21}/cm^3 range. Therefore, the accumulation-layer has a more negligible effect in the more heavily doped region of the drain (i.e., further away from the channel edge). Nevertheless, there is still a component of resistance due to the resistance of the accumulation-layer in series with the channel-ON-resistance, and this component is referred to as the accumulation-resistance, R_{ac}. Both the accumulation-layer and the inversion-layer have about the same thickness (which is given by X_C,and to which we will refer again in the discussion on R_{sp}). *In any case, for deep-submicron MOSFETs, the length of the accumulation-layer is so short, that R_{ac} does not play a significant role in causing the value of R_S to*

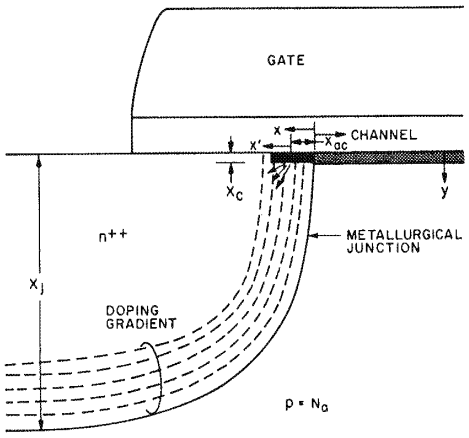


Fig. 5-15 Details of structures near the vicinity of the channel end. The dotted lines are contours of constant concentration. Dark region represents the accumulation layer.[21] (© 1987 IEEE).

rise as the devices are scaled. Hence, we will not consider this component any further in discussing the problem of the increase in R_S as the MOSFET is scaled.

Let us consider R_{sp} next, as this parasitic resistance component does play a role in the increase of R_S as the devices are scaled in the deep-submicron regime. Qualitatively, for a non-abrupt junction (i.e., where the doping contours are spaced over some finite distance, as depicted in Fig. 5-15), R_{sp} arises because after leaving the channel, the current does not immediately spread into the bulk of the source-drain regions (where the resistance effect is given by R_{SDE}). Instead, the carriers leave the accumulation layer and spread out as they move from the thin accumulation layer into the (laterally) uniformly-doped SDE-region of the source or drain SDE (Fig. 5-16).[22] This resistance component is called the spreading resistance, R_{sp}.

An analytical approximation for calculating the value of R_{sp} is given by:

$$R_{sp} \sim (2\rho/\pi W) \ln (\beta X_j/X_C) \qquad \textbf{(5.5)}$$

where ρ is the resistivity of the SDE-region, and X_C is the inversion-/accumulation-layer thickness, and β is a constant that varies from 0.37 to 0.9 (depending on the method of derivation). For the purpose of this discussion, an intermediate value of 0.58 will be used.[21]

The value of the series-resistance component R_{SDE} of the SDE-region as shown in Fig. 5-14 can be estimated from (i.e., $R = \rho l/A$):

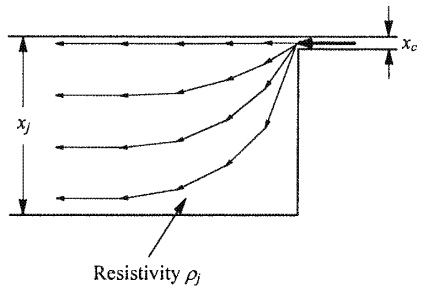


Fig. 5-16 Schematic diagram showing the resistance component associated with the injection region where the current spreads from a thin surface layer into a uniformly doped source or drain region.[22] (© 1983 IEEE)

$$R_{SDE} \cong \rho\, l_{SDE}/(WX_j) \qquad \textbf{(5.6)}$$

Without resorting to a two-dimensional analysis to calculate the length of the SDE region l_{SDE}, a reasonable estimate of its value is $L_L/2$ (where L_L is the lithographically drawn gate-length). Thus, since $X_j \sim 0.55\ L_L$ (as described in the introduction of this section), with scaling the ratio of l_{SDE} to the SDE junction-depth X_j will remain approximately 1.2.

The sum of R_{sp} and R_{SDE} will give the value of the parasitic series resistance component of the SDE region of the drain contact structure. That is, when Eq. 5.5 and 5.6 are added together we get:

$$(R_{sp} + R_{SDE})W \cong \rho\,(l_{SDE}/X_j)\left[1 + (2X_j/\pi\, l_{SDE})\ln(\beta X_j/X_C)\right] \qquad \textbf{(5.7)}$$

From Eq. 5-7 it can be seen that, in effect, R_{sp} multiplies the resistance of the SDE region by the quantity given in the brackets of Eq. 5-7. This quantity is expected to range from a factor of 2.09 at the 250-nm generation to about 1.48 at the 50-nm generation, which indicates the R_{sp} plays a significant role in the parasitic resistance of deep-submicron MOSFETs, and it would be beneficial if methods could be developed that could reduce its value.

5.3.1.2 Comparing the Resistance of the SDE Region (R_{sp} + R_{SDE}) to R_{sat}: With the equations derived in Sect. 5.3.1.1 it becomes possible to estimate the values of (R_{sp} + R_{SDE}) that will be required for MOSFETs scaled down to 50-nm. (Note that this analysis was adapted from Ref. 23.) The center curve of Fig. 5-17 shows the calculated values of the channel resistance $R_{sat}W$ (in Ω-cm) of MOSFETs from the 0.25-μm (250-nm) node down to the 50-nm node. The upper and lower curves show what the value of $20(R_{sp} + R_{SDE})W$ would be if the SDE region had sheet resistance values of 500-Ω/sq and 1000-Ω/sq, respectively. Note that

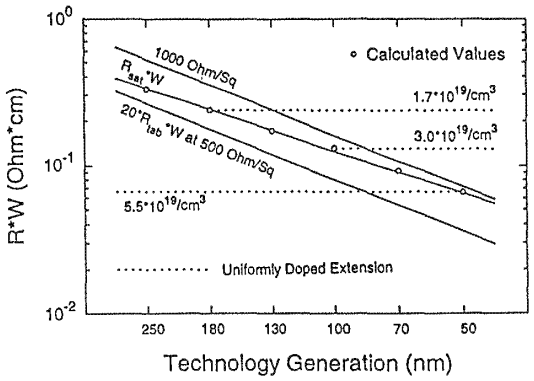


Fig. 5-17 Allowed values of drain (or source) extension resistance considering both bulk extension resistance R_{SDE} and extension spreading resistance R_{sp}.[23]

values are shown for *n*-channel devices, *p*-channel devices would be about 20% higher. In addition, $20(R_{sp} + R_{SDE})W$ is used because the source and drain contact regions each contribute to the resistance in series with $R_{sat}W$.

The data in Fig. 5-17 indicates that the sheet resistance values of $(R_{sp} + R_{SDE})$ will need to be about 650-Ω/sq for 250-nm MOSFETs and about 900-Ω/sq for 50-nm MOSFETs if the value of $20(R_{sp} + R_{SDE})W$ needed to limit the degradation in I_{Dsat} to 8-10% is to be achieved. However, since the junction-depths of the SDE-regions decrease as the gate-lengths are reduced, in order to keep the sheet-resistance-values in the 650-900-Ω/sq range, the average doping-densities in the SDE-regions will have to be increased. As shown in Fig. 5-17, the peak values of doping will have to exceed $10^{19}/cm^3$ (e.g., for the 50-nm node, the peak doping will have be $5.5x10^{19}/cm^3$).

By 2001 it was determined that junctions whose depth was 35-nm could be fabricated with 0.5-keV implants and spike-anneals, but that this was about as low as could be practically achieved (see Sect. 5.3.3). Thus, for 0.1-μm technologies, it will be possible to fabricate SDE-regions with adequately shallow junctions and sufficiently low-values of sheet-resistance. However, it is not clear if it will be possible to do this for 50-nm devices, since the junction-depth will have to decreased to about 20-25-nm, and this will mean that the sheet resistance will have to be significantly increased above 350-Ω/sq.

Also note that the model for R_{sp} given by Eq. 5-5 does not include the effect of the steepness of the lateral-doping-concentration-gradient of the SDE-implant.

As MOSFETs are scaled below 100 nm, this effect becomes the dominant factor that controls the value of R_{sp}. We discuss this issue further in Sect. 5.3.5.

Since increasing the concentration of dopants in the shallow SDE-regions is a difficult task, other ways to reduce the value of the resistance of the SDE-regions have also been sought. From Eq. 5-6 it can be seen that if the length of the SDE region could be reduced, the value of its resistance would also be decreased. While this can be accomplished by decreasing the width of the sidewall spacer, this is not practical in the conventional deep-micron MOSFET-structure. That is, since a deep contact-region is needed (to accommodate the reacted salicide-layer, see Sect 5.3.1.3), the lateral-straggle of the deeper contact-implant would penetrate under the gate edge too far if such a narrower sidewall-spacer is used. However, as we will discuss in Sect. 5.3.2, incorporating elevated-source/drains into the MOSFET structure would permit the use of such narrow sidewall-spacers, thus allowing the resistance of the SDE-region to be made smaller.

5.3.1.3 Parasitic-Resistance R_{co} of the Deeper Contact-Region of the Drain: Although the contact-region of the drain-structure has a deeper junction-depth than the SDE, it can also contribute significantly to the parasitic series resistance of the drain structure, namely through the *contact resistance*, R_{co}. That is, a salicide ($TiSi_2$ or $CoSi_2$) is formed on the surface of the contact-region by sputter-depositing a layer of Ti (or Co) onto the wafer. Then the Ti (or Co) is reacted with the silicon at elevated temperatures, forming the salicide layer. This salicide-layer (and the kinetics of its formation) play a role in the value of R_{co}.

A simple model for R_{co} (or $R_{co}W$) is given by:[21]

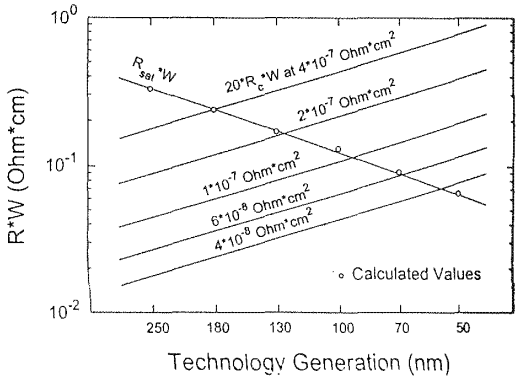


Fig. 5-18 Allowed values of drain (or source) contact-resistance.[23]

$$R_{co}W \geq \rho_c/2\, L_L \qquad (5.8)$$

where ρ_c is the specific contact resistivity of the silicon-to-silicide interface (in Ω-cm^2). This can again be compared with the channel-resistance R_{sat}. Figure 5-18 gives values of $20R_{co}W$ for a 10% contribution of (source + drain) contact resistance. To have a 10% or less contribution, the data in Fig. 5-18 indicates that ρ_c must be in the range of 10^{-7}-Ω-cm^2 for 0.25-μm MOSFETs, but must get down to around $4x10^{-8}$-Ω-cm^2 for 50-nm MOSFETs. It has been demonstrated that for a barrier height of about 0.5-eV, the required doping-density needed to achieve such a low value of ρ_c is above 10^{20}/cm^3. This means that the silicon regions in contact with the silicide will have be very-heavily doped, a condition that may not be possible to attain in 50-nm MOSFETs with conventional structures, for reasons we describe next.

When the silicide is formed on the source/drain contact-regions by the reaction of Ti (or Co) and silicon, some of the silicon in the substrate is consumed (i.e., when $TiSi_2$ is formed, the silicon consumed is about 75% of that of the thickness of the final silicide film, while for $CoSi_2$, even more silicon is consumed – about equal to the thickness of the $CoSi_2$ that is formed). This has two repercussions. First, the contact-region of the drain must deep enough so that the junction is not entirely lost as a result of the silicide-formation. To keep the implanted dopants from negating the benefits of the shallow SDE-junction (as a result of lateral-straggle that causes the implanted species to penetrate laterally under the implant-mask edge) a sufficiently wide gate-sidewall spacer must be used to mask the contact-region-implant. This impacts the resistance of the SDE-region, through an increase in l_{SDE}. Second, the silicon consumed is that nearest the surface, just where its doping concentration is the highest. The interface between the resultant silicide and the silicon is far below the initial silicon surface, where the implanted doping-concentration is considerably lower (which will impact the value of R_{co}). Furthermore, the dopant in the silicon (especially boron) is known to segregate into the silicide layer, further depleting the doping-concentration at the silicide-silicon interface. Since the contact-region of the drain cannot be made arbitrarily deep (due to the fact that lateral-straggle also increases as the projected-range of the implant is increased), it is not clear that it will be possible to make contact-regions with doping concentrations that are adequately high to achieve the required low-values of ρ_c. The use of elevated-source/drain regions (Sect. 5.3.2) would also allow a way around this obstacle.

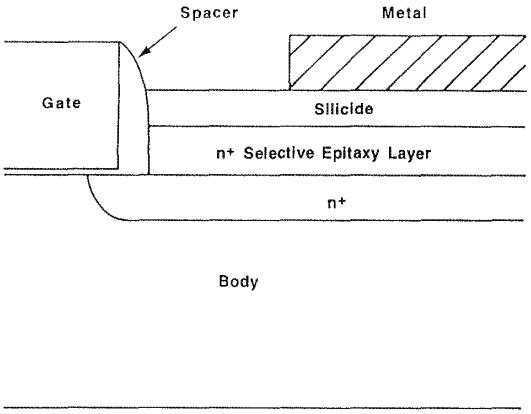


Fig. 5-19 Schematic of elevated-source/drain contacting structure.

5.3.2 MOSFETs with Elevated-Source/Drains

One proposed method for reducing the series-resistance of the source/drain structures in deep-submicron MOSFETs is to modify the basic device structure by adding so-called *elevated* (or *raised*) *source/drains* (Fig. 5-19). In this method, the heavily-doped contact is elevated above the original substrate surface by forming a selectively-grown epitaxial-Si (or perhaps Si-Ge) layer.

The MOSFET with elevated-source/drains has several advantages for devices when they are scaled to 50-nm and beyond. First, the width of the sidewall-spacer can be made much smaller. This can greatly reduce the resistance of the SDE region by decreasing its length l_{SDE}. Second, the n^+ contact layer can be sufficiently thick to contain the salicide layer without junction-leakage problems. Third, the selective-epitaxy-layer can potentially be heavily-doped during growth to obtain a lower contact-resistance. However, two selective-epitaxy processes will be needed since both n^+ and p^+ contacts must be made. The Si-epitaxial-growth steps occur after the SDE-implant. To keep the dopant of the SDE-region from diffusing too-far under the gate-edge, the selective-epitaxial-processes will have to be carried out as sufficiently low temperatures (i.e., probably below 850°C). A technique involving such a low-temperature epitaxial-process for forming elevated-source/drains was described by Miyano *et al.*[24]

In late-2001, Intel announced that the deep-submicron-MOSFET structure it intends to use for the next decade is designed with elevated-source/drains (and an SOI-structure that also employs a high-k dielectric [ZrO_2], SSR-channel-doping, and pocket-implants for punchthrough-control).[58]

5.3.3 Shallow-Junction-Formation for Deep-Submicron Source/Drains

The main thrust for achieving shallow SDE region junction depths is to decrease the energy of the SDE-implant. For the source/drain-structures of the 0.13-μm and 0.1-μm technology nodes, single-digit keV energies are expected to be needed for the implants that will form the drain contact regions, and sub-keV energies will be needed for the SDE-region-implants.

5.3.3.1 Ultra-Low-Energy Implants: There are three reasons that reducing the energy of the implant leads to a decrease in the junction-depth. First, the projected range R_p of the implant becomes smaller as the energy is reduced. Second, the straggle and ion-channeling effects both scale with R_p (which is especially important for B implants). Third, TED that occurs during the implant-annealing step is reduced (see Sect. 5.3.3.2). It turns out that for sub-keV B-implants the channeling-effects become relatively unimportant (compared to the broadening of the implant doping profile caused by TED during the annealing step). For this reason, elaborate measures to reduce channeling (such as performing a pre-amorphization-implant prior to the low-energy SDE-implant) are not likely to be used for MOSFETs with gate lengths of 0.1-μm and smaller.

The difficulty with implementing such low-energy-implant processes in manufacturing applications is achieving the productivity that the industry has been accustomed to getting at higher-implant-energies. For example, at energies greater than 10-keV, a high-current-implanter with a beam-current between 10- and 25-mA can implant 230 wafers per hour with the high dose of $4x10^{15}/cm^2$. (Note that beyond 0.18-μm, the doses of the SDE will be in the $1\text{-}2x10^{15}/cm^2$ range, and the doses of the contact-regions of source/drains will be even higher.) At sub-keV energies, the maximum beam-current available on commercial implanters in the mid-1990s timeframe was in the 1-2-mA range. Even after a reduction in dose (made possible by the decreasing energy), the productivity is only fraction of what it previously was. The decreased beam-current at lower energies follows from Child's law, which states that the maximum ion-current that can be extracted from any cathode, drops off as $V^{3/2}$ where V is the extraction voltage. While this basic constraint can be somewhat circumvented (by extracting a larger current at a higher energy and then decelerating the beam

to a final lower energy; i.e.., what implant vendors refer to as "decel"), the effective beam-current still decreases with the final energy. With decel, care must also be taken to minimize energy-contamination caused by ion-neutralization before (or in) the decelerating field. That is, ions neutralized while being transported in the beamline will not be counted as part of the dose. Furthermore, they will be implanted at much higher energy, possibly causing deeper junctions (thus defeating the central purpose of using an ultra-low-energy implant).

There are two commercially available high-current, low-energy-implanter configurations: **1)** *batch* (or *multiwafer*) implanters; and **2)** *single-wafer* implanters. Batch type implanters are sold by Axcelis (GSD/200E® model) and Applied Materials (Quantum LEAP®). A single-wafer, high-current, low-energy-implanter (with a scanned ribbon beam) is offered by Varian Semiconductor (VIISta-80®). The vendors of the "batch" type low-energy-implanters argue that the throughput is significantly higher than for single-wafer machines. The batch systems use a fixed beam, while the single-wafer systems use a scanned beam. The tuning time for a fixed beam is always shorter than that needed for a scanned beam, Hence, the set-up time for a batch system is smaller. Capacity modeling studies have shown that fewer batch low-energy-implant-systems are required (e.g., 3 to 4 for a fab that starts 5000-wafers per week with a 0.18-μm process, versus 5 to 9 single-wafer-implanters), primarily because of the fast tuning in the batch machines. This reduces the overall capital-cost and cost-per-wafer.

The advantages claimed for the low-energy, single-wafer-implanters are the following: **1)** the angle of the implant with respect to the wafer surface can be controlled more accurately in a single-wafer-implanter; **2)** sputtering of metals from the spinning disk (the largest source of contamination in batch-implanters) is eliminated; and **3)** a 70°-angle analyzing-magnet allows the VIISta-80® to essentially eliminate the problem of energy contamination.

While a significant effort was being expended to improve the productivity by increasing the beam current,[25,26] the basic capability of being able to implant at an ultra-low-energy (i.e., sub-keV) already existed by 2000. (That is, by 2000 implanters were available that could provide 4-mA at 0.2-keV.) This allowed researchers to investigate the limits of shallow SDE-junction-depths with sub-keV implants and conventional RTA methods. Drain extensions with junction-depths that meet 0.1-μm technology requirements were demonstrated.[27] It appears, however, that it may not be possible to improve on these results unless either the extension-resistivity-specification is significantly relaxed or new annealing technologies are developed which provide a smaller thermal budget

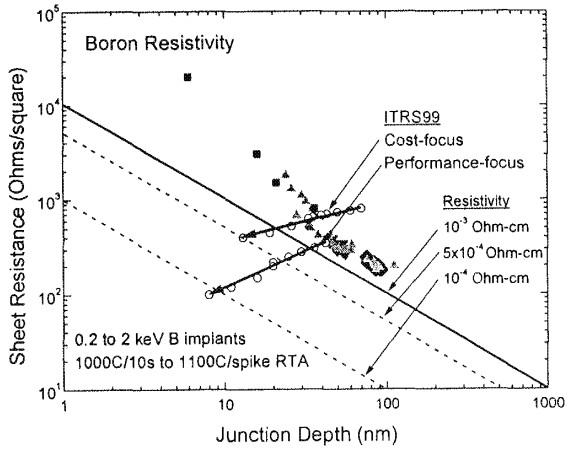


Fig. 5-20 Sheet resistance and source/drain junction depth expectations for CMOS transistors according to the ITRS99 study for cost and performance-focused product and gate sizes from 180 nm towards deeply-scaled devices with gates ~20 nm. Also shown are experimental results for sub-keV B^+ implants with "spike" RTA in CZ-Si materials.[59]

and are compatible with CMOS processing. That is, in order to meet the decreased resistivity requirements of the scaled MOSFETs, it is necessary to implant a larger dose or increase the annealing temperature to activate more of the smaller dose implant. The issue is illustrated in Fig. 5-20, where the ITRS99 expectations for the source/drain-extension junction-depth and sheet-resistance are compared to experimental results for boron-doping with sub-keV B-implants and a "spike"-RTA process (with a high-temperature thermal-cycle of 1050°C for 1 sec). All of the experimental results are above the ~10^{-3}-Ω-cm resistivity "wall" (the dark downward-sloping line in this figure). The shaded squares and triangles in Fig. 5-20 are actual experimental data, whereas the open circles are expected results that have not yet been attained. The resistivity requirements for "high performance" MOSFETs are ~10^{-4}-Ω-cm for 20-nm MOSFETs.

As compared to the SDE-regions of the source/drain, higher doping-concentrations are needed for the contact source/drain-regions. This is because these regions are used to contact the silicide-layer, and high doping-concentrations help to reduce the value of R_{CO}. Since these regions are not adjacent to the channel, they do not have as great an impact on short-channel behavior as the SDE-junctions. Consequently, doses and implant-energies can be higher than those permissible for SDE-regions, even when B-implants are used.

By 2001 the trend of using lower implant energies for the SDE-implant for each generation appears to have reached a practical limit in terms of device performance (especially for *p*-channel transistors). That is, 0.5-keV boron implants appear to be the lower useful limit for shallow junctions, not just for 0.13-μm but for all technologies.[28] Below 0.5-keV, the implant becomes less productive, more costly, and the final junction performance (i.e., the simultaneous achievement of junction-depth and resistivity) is the same. The performance limit with a 0.5-keV boron-implant and spike-anneal is a junction-depth of about 35-nm and a sheet resistance of about 350-Ω/sq. It appears that to further reduce either value, the other must be increased.

5.3.3.2 Transient-Enhanced Diffusion: The second key reason for developing dopant implant processes with energies in the sub-keV range (besides reducing the as-implanted depth), is to minimize the transient-enhanced diffusion (TED) of the dopant. When high-energy ions enter the silicon lattice, they displace silicon atoms from their lattice sites, creating pairs of vacancies and interstitials. When the post-implant anneal is initiated, most of these interstitials and vacancies recombine, but some interstitials remain behind (in an amount about equal to the quantity of implanted atoms – most of which now occupy the sites that were formerly filled by the still-displaced silicon interstitials). The fact that each implanted atom produces about one residual interstitial after a very short anneal is called the "*+1 approximation.*" TED can be explained by relating its enhanced-diffusion effects to the presence of these excess interstitials for a wide range of low- and medium-dose implants, and low- and medium-energies.

The residual excess interstitials rapidly coalesce into such unstable extended-defects as {311}'s and also into more-stable extended-defects, such as dislocation loops. While this coalescence occurs because there is a lower system-energy when the interstitials populate extended-defects (compared to when they exist as individual interstitials), the defects are still metastable. Thus, they dissolve with continued annealing (especially the {311}-defects). During their dissolution they emit interstitials into the silicon. The rate at which these defects dissolve depends on the activation-energy and annealing-temperature. Nevertheless, as long as these defects exist, an interstitial supersaturation is maintained (i.e., a higher concentration of interstitials is present than would exist in the absence of the extended defects). Since boron diffuses by an interstitial mechanism, its diffusivity is enhanced by the excess interstitials - with the time-averaged diffusivity-enhancement equal to the time-averaged interstitial-supersaturation. In any case, both the interstitial-supersaturation and the

diffusivity-enhancement soon end after the defects have dissolved. The enhancement is thus transient.

The *increase* in depth of a shallow SDE-junction ΔX_j due to TED can be expressed in terms of the initial-supersaturation of interstitials and the time of the interstitial defect dissolution as:

$$\Delta X_j^2 \propto N R_p \exp [- (-1.4 \text{ eV})/kT] \quad \textbf{(5.9)}$$

where N is number of interstitials initially trapped in the defects (approximately equal to the implanted dose, as noted above) and R_p is the projected ion range (which is where the peak concentration of the excess interstitials is initially located). The linear dependence of N simply represents the fact that the larger the dose, the larger the supersaturation, and hence the greater the diffusivity enhancement. The linear dependence on R_p has been demonstrated experimentally. The nearer the excess interstitials are to the surface, the faster they reach the surface (where they are annihilated). The linear dependence on R_p explains why use of ultra-low-energy implants reduces TED effects. That is, if an ultra-low-energy implant is used, R_p will be small. Thus, the location of the interstitials will be closer to the surface, allowing them to escape there faster. Thus, the duration of the transient is shortened, making the value of ΔX_j smaller.[29] (Later work indicated that for implants below 1-keV in energy, TED is even further reduced because there is less lattice damage per ion.[30])

Another important dependence contained in Eq. 5-9 is the *negative* activation-energy of ΔX_j^2, which arises from the fact that the interstitial-supersaturation due to the presence of the extended-defects is larger at a lower temperature. The negative-activation-energy implies that if the defects are annealed out at a low temperature, the final junction will be deeper than if the defects are annealed out at a higher temperature. This is one of the main reasons why the S/D anneals are currently carried out in an RTA tool rather than in a conventional batch furnace (that has a ramp-up rate of a few degrees per minute). In an RTA chamber, significantly less time is spent during the temperature ramp-up at a lower temperature (at which the diffusivity-enhancement is larger).

5.3.3.3 Rapid-Thermal-Annealing of Shallow Junctions: To form the shallowest junction possible using ultra-low-energy implantation, it is necessary to minimize the thermal-budget after implant. (Thus, as noted above, annealing such implants is carried out in RTA tools to realize good-quality shallow junctions.) In the ideal case, the as-implanted-dose would spend "zero" time at the temperature that would electrically-activate the dopants and heal the lattice damage from the implant. (Exposure to a high-temperature for a very-short time is effective

because the activation-energy for dopant-activation is typically higher than that for diffusion, thus electrical activation is favored while limiting diffusion. Reported values for boron are 5.4-eV and 3.4-eV for activation and diffusion, respectively.) Attempts to reach this ideal annealing condition have led to the development of so-called "spike" anneals, which consist of very fast ramp-up and ramp-down of temperature, with virtually zero-time at the peak temperature. Experiments have shown that the difference in the percentage-of-activation and sheet-resistance between spike-anneals and those that keep the wafer at temperature for 10-sec is negligible. On the other hand, for the 10-sec-at-temperature anneals, the junction-depth increases with temperature.[88] This indicates that spike anneals are more effective at higher peak-temperatures.

In practice, spike anneals have an effectively finite-time at temperature. Also, in reality, while ramp-up rates as fast as about 200°C/sec are possible, the ramp-down is limited to approximately 50-80°C/sec. As a result of these deviations from ideality, it is useful to determine what is a sufficiently-fast ramp-up rate. The negative-activation-energy of TED seems to indicate that as high a ramp-rate that can be attained, should be used during a spike anneal. However, experiments show that for such ultra-fast ramp-rates, the advantage gained by going from moderately fast (e.g., 100°C/sec) to ultra-fast ramp-up rates (200-400°C/sec) is relatively small for ultra-low-energy implants. Such a small advantage must be balanced against the loss of process control (e.g., across-wafer thermal-uniformity and temperature-repeatability) which inevitably occurs when such ultra-fast ramp-up-rates are used. The reason the advantage is not very great is that TED is already sufficiently minimized at 100°C/sec.[32] This reduction in thermal-budget during ramp-up is negligible compared to the portion of the thermal budget which cannot be reduced (the finite-dwell-time at temperature and the relatively low ramp-down-rate).

The data shows that the largest reduction in junction-depth with increased ramp-up-rates occur for the lowest-energy-implants (see Fig. 5-21).[31] Since TED decreases with decreasing implant-energy, it can be inferred that the advantage is primarily due to the overall thermal-budget and not just of the TED. It is therefore evident that without the ability to increase the ramp-down-rate (when it is relatively smaller than the ramp-up-rate), the advantage from increasing the ramp-up-rate will quickly saturate at ramp-up-rates greater than the ramp-down-rate. In addition, it has been determined that the ramp-rate is not as good a measure of the quality of a spike-anneal as is the time spent diffusing the dopant. That is, since about 90% of the dopant-activation occurs in the short-duration from approximately 50°C below the peak-temperature to the peak-temperature

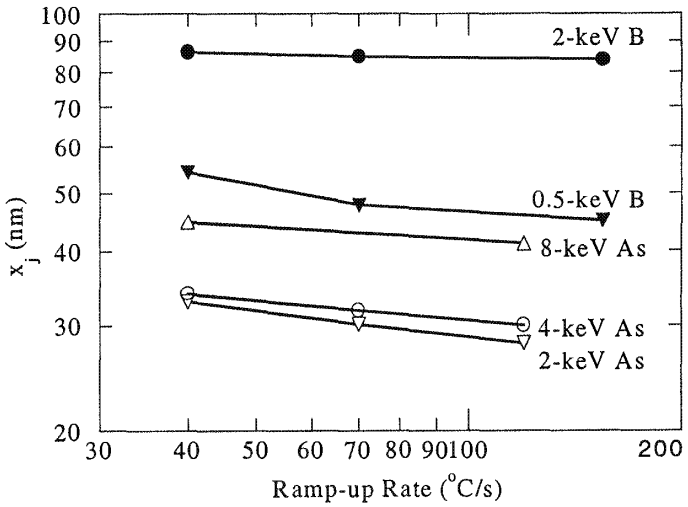


Fig. 5-21 Junction-depth, measured at a concentration of $1x10^{18}/cm^3$ after spike-annealing at various ramp-up-rates. The ramp-down-rate was the same in all cases. For both As and B the advantage from a faster ramp-up is largest for the shallowest implant. This implies that the reduction in TED (known to decrease with implant-depth) is not the dominant mechanism responsible for the larger reduction in junction-depths observed for the shallower-implants with increasing ramp-up-rates.[31]

and back to 50°C below, this interval of time is really a better specification of the quality of a spike-anneal.[28]

Figure 5-22 is a survey of published spike-anneal data that was presented in Ref. 32. It can be seen that many of the data points are coincident, despite having been taken from different implant-and-annealing conditions (energy, dose, and ramp-rate). It is evident that more than one set of process conditions can be used to arrive at the same junction properties.

A variety of RTA-tool configurations are commercially available, including the following: **a**) linear lamps, double-sided heating; **b**) axisymmetric lamps, single-sided heating; **c**) arc lamp; **d**) hot-plate; and **e**) hot-wall RTP. Several general references are available that give detailed descriptions of these types of RTP tools.[33,34,35] Data on shallow junctions formed with the various RTP-tool types can be found in the following references: lamp-based RTP systems (36); hot-plate (susceptor – 37); and hot-wall RTP (38).

New ways to increase the value of the electrically-active-dopant in shallow-junction regions are also being actively being researched. Some of these novel techniques include: **1**) laser anneals;[39] **2**) strain-compensation implants

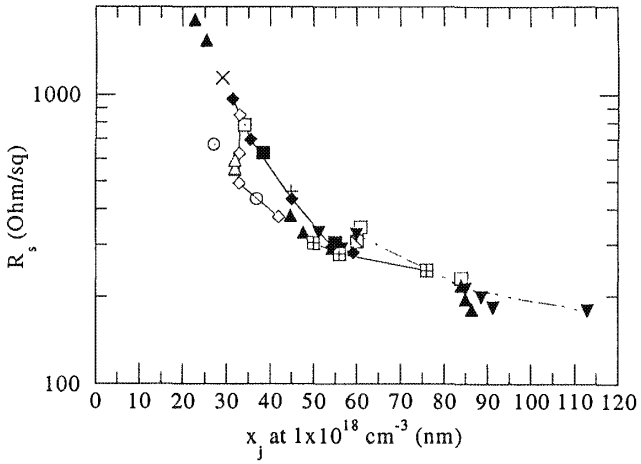


Fig. 5-22 Comparison of *p*-type sheet-resistance and junction-depth data for spike-anneals using various implant-and-annealing conditions. The data set includes a 400°C/sec ramp-up-rate and 0.2-keV implant-energy. Note that all the data points fall in a rather narrow band. No one has been able to make a 30-nm junction with a 300-Ω/sq sheet-resistance using ion-implantation and conventional rapid-thermal-annealing.[32]

(e.g., by adding Ge-atoms to the silicon lattice, their larger size compensates for strain caused by the smaller B atoms. This increases the solubility of the boron atoms in the silicon.);[40] and **3**) using RTP-tools which have a short lamp-response-time, which increases the ramp-down-rate.[41]

5.3.3.4 Dopant Loss for sub-keV Implants from Self-Sputtering, Out Diffusion, and Surface Oxide: Several factors associated with ultra-low-energy implants give rise to loss of some of the species that are implanted, and they are discussed in this section. The effects that lead to such dopant loss are: self-sputtering, out-diffusion, and the presence of a surface oxide during the implant process. When the dose-loss from all three of these effects are combined, the loss can be large enough to pose a significant problem for ion-implant processes carried out at energies of 0.5-keV or lower.

The first phenomenon associated with dopant loss in ultra-low-energy implant processes is *self-sputtering*. That is, when any solid-surface (the target) is struck by an ion-beam, sputtering of atoms from the target will occur. The number of atoms sputtered for each bombarding ion rises as the energy of the ion decreases. In particular, when the ions being implanted have energies in the 0.2-1-keV range, significant sputtering of the target will occur. For example, when 0.5-

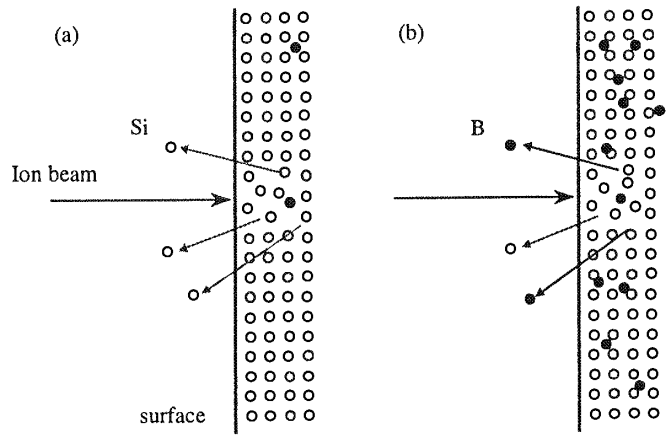


Fig. 5-23 The dose-loss due to self-sputtering depends on how much dopant has already been implanted. (a) At the start of the implant most of the sputtered atoms are silicon. (b) As the composition of the near surface layer becomes altered, with an increasing fraction of them being boron, more boron atoms are sputtered.[59]

keV B-ions strike a silicon target, about one Si-atom is sputtered for every four implanted B-atoms. Since at such low-energies the implanted species also come to rest near the Si surface, some of these implanted B-atoms themselves will be subject to sputtering. In fact, it is expected that with increasing B-dose, more and more of the implanted B-atoms will be sputtered from the surface. (This phenomenon is depicted in Fig. 5-23.) Since additional B-implanted-atoms will be sputtered with increasing dose, this implies that the fraction of implanted dose that is retained in the silicon wafer will decrease with increasing dose. Given the difficulty associated with achieving a large beam-current at such low-energies, a further reduction in the retained dose is very undesirable. That is, while the dopant-loss could be compensated by increasing the implant dose, this will further decrease throughput (which is already too low).

Figure 5-24 plots the retained-dose compared with the nominal-dose for various energies.[32] It indicates that at 2-keV essentially the entire dose is retained. But at implant-energies of 0.5-keV and 0.2-keV, a significant fraction of the implanted-dose has been sputtered out, with the missing fraction rising with increasing dose. For a nominal-dose of $1x10^{15}/cm^2$, sputtering causes a 10% reduction in the retained dose at 0.5-keV, and a reduction of 20% of the retained dose at 0.2-keV.

If BF_2^+ is used to implant B at low-energies, the sputtering problem is expected to be even worse than if B is implanted. That is, in BF_2^+-implantation, for each implanted B-atom that contributes to the retained dose, there are now 3 atoms (one B and 2 F) which can cause sputtering (compared to just one B-atom for an implant process that uses B^+-ions instead). For example, for a nominal dose of $1x10^{15}/cm^2$, we stated that a 0.5-keV B^+-implant will cause a 10% reduction in retained dose. However, if a BF_2^+ implant is carried out so that the B-atom in the BF_2^+-ion has the same energy as in the 0.5-keV B^+-process (i.e., the BF_2^+-ions have energies of 2.2-keV), the reduction in retained-dose increases to ~20%. The above circumstances imply that caution must be exercised when comparing junction data from the literature for different sub-keV implants or for different species. A correction needs to be applied for the different retained-doses to allow meaningful comparisons to be made.

The second effect that gives rise to dose-loss is *out-diffusion.* Out-diffusion of implanted species occurs during the annealing steps that are performed after the implant. As the projected-range R_p of the dopant is decreased (as a result of implanting at lower energies), the out-diffusion grows worse. The reduction in retained-dose after a 1000°C/15-sec-anneal is approximately 10% for an implant energy of 0.5-keV, but increases to 15% for 0.2-keV energies.

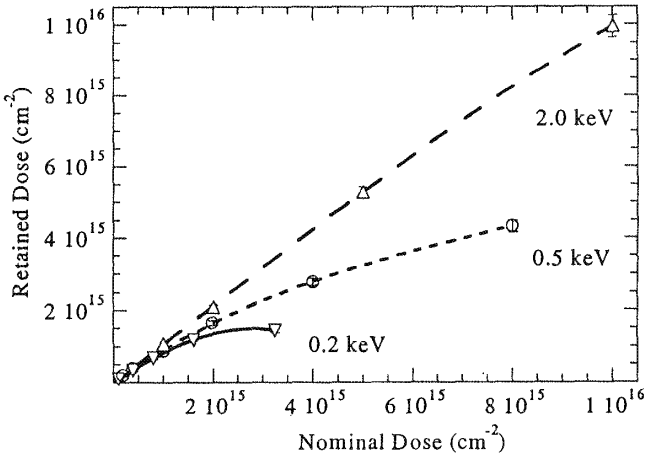


Fig. 5-24 Retained B-dose as a function of nominal B-dose implanted into (100)-silicon wafers with an as-received native oxide. The trend lines are 3rd order polynomial spline-fits. For a nominal B-dose of $1x10^{15}/cm^2$, sputtering leads to a reduction of ~10% in the retained dose at 0.5-keV, and a reduction of ~20% at 0.2-keV.[52]

The final cause of dopant-loss in sub-keV implants is due to the presence of an *oxide on the silicon-surface during implant.* Such an oxide is often deliberately grown to protect the silicon surface from metal contamination during the implant process (see Chap. 10, Vol. 1, 2nd Ed.). While this protective oxide layer can be made quite thin, it nevertheless has a finite thickness. It is clear that as the implant energy is reduced, a larger fraction of the as-implanted dose will be located in the protective oxide. Since the oxide is stripped off at the end of the implant step, the dopant that remains in the oxide cannot contribute to the formation of the junction. Approximately 25% of the B implanted at 0.5-keV is lost by removal of a 15-nm-thick surface-oxide, while more than 40% is lost for 0.2-keV implants. With these results, it is possible to estimate the equivalent-dose needed at lower energies (such as 0.2-keV) to make up for the larger loss as compared to 0.5-keV implants. For example, it has been calculated that to successfully implant a dose of $4x10^{14}/cm^2$ B-atoms at 0.2-keV into the silicon beyond the protective oxide, the beam-current would have to be as much as 60% greater than the beam-current at 0.5-keV to obtain an equivalent throughput process.[32] To circumvent the oxide dose-loss problem, some ultra-low-energy implant processes are being carried out without the use of a protective oxide.

In summary, the dose-loss effects are significant for sub-keV implant processes, and the loss increases as the implant-energy is reduced. For the extreme case of 0.2-keV implants, as much as 40% of the B-dose can be lost due to the combined effects of self-sputtering, out-diffusion, and native-oxide removal for the high doses needed for an SDE-implant.

5.3.4 Effect of the Overlap of the SDE-Region and Gate-Edge on MOSFET Performance

In addition to being able to carefully set the junction-depth X_j and length of the SDE region l_{SDE}, it has also been reported that the amount of overlap of the SDE-region with the gate-edge is also critical for optimum device performance.[42,57] That is, if the overlap of the SDE-region and the gate-edge is too great, the deep-submicron MOSFET will suffer excessive leakage (due to larger short-channel effects that arise if the channel-length is too small). Excessive leakage means that the off-state current exceeds its maximum design value (e.g., 1 nA/μm). In addition, the gate-to-(drain/source) overlap-capacitance increases, which negatively impacts device switching speed.

On the other hand, if the SDE-to-gate-edge-overlap is too small, the ON-state saturated-drain-current I_{Dsat} will decrease because the SDE-region and the end of the channel are not electrically well-coupled. Thus, an excessively-high-resist-

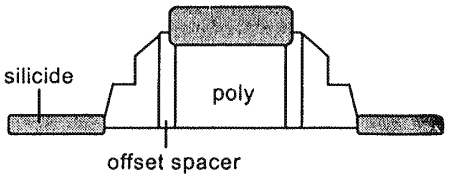


Fig. 5-25 Schematic of a deep-submicron MOSFET with an offset-sidewall spacer.[57]

ance-region separates them (and this increases the parasitic resistance of the source/drain regions, which leads to a decrease in the value of I_{Dsat}).

Controlling the position of the edge of the SDE region with respect to the gate edge can be done in the following way. First, the implant-and-annealing processes for forming the SDE (and contact) region of the drain structure are determined, and the drain structure is formed using these process conditions. However, an offset-spacer is fabricated on the side of the gate after the gate-poly is etched (but before the SDE implant step is carried out, Fig. 5-25).[57] The width of this off-set spacer is varied from zero (no-spacer) to some upper-bound (e.g., 50-nm). Measurements are then made for the MOSFET having different offset-spacer-widths. The offset-spacer-width that yields optimum device characteristics (i.e., the best combination of off-current and I_{Dsat}) is taken to be the value that produces the correct overlap of the SDE-region and the gate edge.

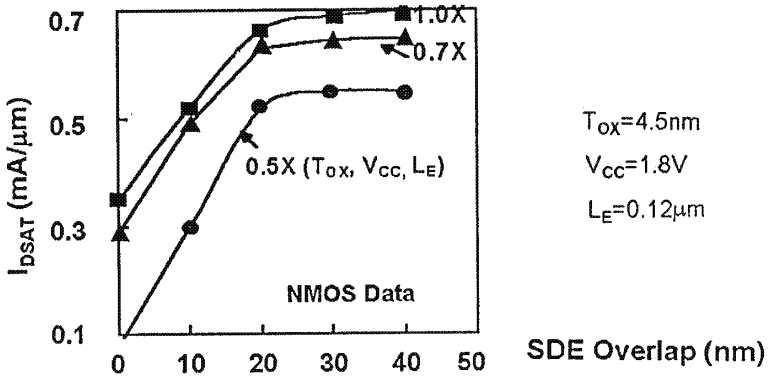


Fig. 5-26 I_{Dsat} versus SDE overlap (NMOS).[42]

The results of such a study are shown in Fig. 5-26. In this case, the effects of SDE-overlap on 0.25-μm NMOSFETs were analyzed. The arsenic-implanted SDE junction-depth X_j was measured to be 50-nm. Since the lateral diffusion of the SDE junction under the gate-edge is estimated to be 0.6-0.7 times the vertical depth (about 30-nm), minus the width of the offset spacer, the SDE-overlap for each offset spacer width can be determined. Fig. 5-26 shows that if the SDE-overlap is less than 20-nm, the value of I_{Dsat} falls off rapidly. Thus, a minimum overlap of ~20-nm is recommended by this study. Information as to how to form the offset spacer was not given, but a method (such as re-oxidizing the poly-silicon gate-structure after it has been etched) is one possible way of doing so.

5.3.5 SDE-Junction Abruptness

The SDE-region must not only have the correct junction-depth, average-doping-concentration, and position under the edge of the gate, but recent studies have shown that there must also be a highly-abrupt lateral-doping-gradient at the interface of the SDE and the MOSFET channel-regions.[43,44] Not only does such an abrupt-doping-gradient help limit the lateral-encroachment of the extension-depletion-regions into the channel, but more importantly it minimizes R_{sp}. The design rules for lateral-junction-abruptness are that the doping-falloff must be at least one decade in a distance equal to ~3% of the size of the gate-length (for a spreading-resistance less than 5% of R_{sat}). For 100-nm MOSFETs this means that the junction-abruptness must be ~3 nm/decade, which is about the profile slope of as-implanted B at 0.2-keV.[45] Since the abruptness will have to increase even further as smaller MOSFETs are designed, this implies that new techniques will have to be developed that result in final junction-doping-profiles that are even steeper than the as-implanted profiles.

The design of the implanter can play a role in the lateral-abruptness of the SDE as-implanted doping profile. That is, unlike mid- to high-energy ion-implant applications, channeling-effects are minimal for low-energy, high-current implants (because of increased channeling acceptance angle). Nevertheless, there is still a significant difference between single-wafer and batch-implanters due to implant shadowing effects. Ideally, a 0°-implant will give the most-abrupt lateral-junction because any tilt-angle larger than 0° induces shadowed areas at the vicinity of the gate pattern (leading to laterally-graded doping-profiles). If the incident beam has an angular-spread deviating from the ideal parallel beam, each of the ions having different incidence angles will make different shadows. The result is a loss of dose- and junction-abruptness at the outskirts of the gate. A single-wafer, parallel-beam-implanter minimizes the incident-angle-variations.

Spinning-disk batch-implanters use a stationary-spot ion-beam. The wafers pass through the beam due to the rotation of the disk. Such batch-implanters use a concave-shaped disk to secure wafers in their respective sites during the high-speed motion of the disk. Centrifugal force exerted on the disk surface holds them in place. Since the conical disk has a curved surface and the wafer-surface is flat, this leads to an incident-beam angle-variation across the wafer during the implant process. Depending on the dimensions of the disk, cone-angle, and wafers, this can amount to 1.5°-variation across a wafer.[46]

Recently, the abruptness of the SDE-lateral-doping was measured using both scanning capacitance microscopy (SCM)[47] and scanning tunneling microscopy (STM).[48] The latter measurement detected a tail-region of dopant at the end of the doping profile, spreading from the extension-region in the horizontal direction near the gate oxide. The cause of this anomalous diffusion tail was not understood.

5.4 ANTI-PUNCHTHROUGH STRUCTURES IN DEEP-SUBMICRON CMOS

Lateral tailoring of the channel doping profile can provide some additional flexibility to control V_T and punchthrough. Recently, such lateral tailoring of the channel profile has been carried out with the use of *large-angle-tilt implanted punchthrough stopper* (LATIPS), which are also referred to as *pocket-implants* (and are depicted in Fig. 5-27). Such a pocket-implant puts additional dopants in the channel close to the source/drain-junctions for punchthrough control, and leaves much of the channel region with a lighter doping, as needed for a low V_T.* The peak doping values of the pocket-implants must be somewhat larger than the doping-concentration values needed for a punchthrough-control layer that is uniform in the lateral channel direction between the source and drain.

These punchthrough-suppression pocket-implants are done after the source/drain-extension-implant (and before the formation of the spacer used to define the edge of the deeper contact-source/drain-implant). This ensures that the pocket-implant-dopant is self-aligned to the gate. The dopant is placed just below the active channel, adjacent to the source and drain, and is of the same dopant-type as the well. Precise control of the placement and dose of this implant within the transistor-structure is needed to obtain the intended enhancements. As noted

* As pointed out in Chap. 1, if an implant with a beam directed normally to the wafer surface is used to introduce dopants to create a punchthrough-control structure, instead of one whose direction is tilted at a large angle to the wafer surface, the resulting punchthrough-control structure is termed a *halo* structure. Such a halo structure is described in Chap. 1, and is depicted in Fig. 1-2b.

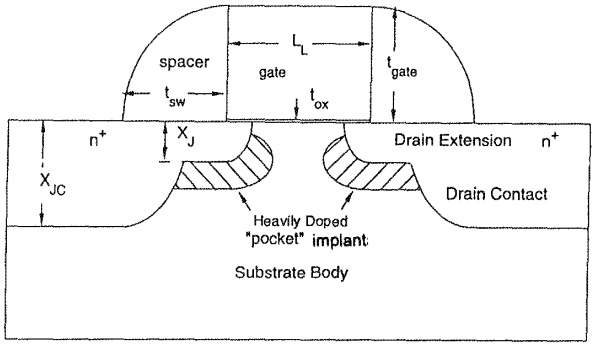


Fig. 5-27 MOSFET with heavily-doped pocket-implants for punchthrough-control.

above, the use of high wafer-tilt-angles when performing such pocket-implants offers an additional, controllable processing parameter for producing complex channel structures, without diffusion or additional masks. (The presence of the pocket-implanted-species also helps keep the source/drain-extension [SDE] junction-depths shallow, because they compensate some of the dopant in the SDE region [which, in effect, squeezes the SDE-junction closer to the surface].)

As introduced by Hori, high tilt-angles are needed for achieving greater

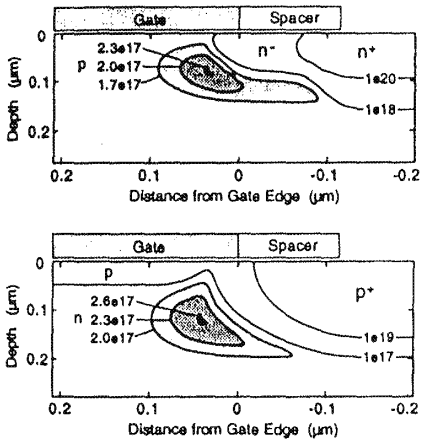


Fig. 5-28 Simulated 2-D impurity profiles of typical LATIPS structures for (a) NMOSFET with *n*-LDD region and; (b) Buried-channel PMOSFET.[49]

lateral-penetration of the pocket-implant than the adjacent source/drain-extension implant without thermal diffusion.[49] That is, the species implanted by the pocket-implant should reside outside of the final source/drain-extension-doping-profile after implant and annealing. Typically, tilt-angles are 20°-45° from the wafer normal. Figure 5-28 shows the resulting (simulated) doping profile in the channel region and in the adjacent well of a LATIPS device.[49]

To maintain device symmetry with high tilt-angle, the implants must be split into segments. Since most circuits contain transistor gates placed 90° from each other, high tilt-angle implants are typically done in four segments of equal dose ("quad-implants") with the wafer rotated 90° between segments. (As gate-widths shrink, less undercutting of the gate is required to produce the punchthrough stop effect. Consequently, tilt-angle requirements for this application are steadily decreasing below the 0.25-μm node, and are expected to reach the "low-tilt-angle" threshold of 10° by the 0.1-μmtechnology node.)

High-tilt-angle punchthrough-suppression-implants improve device performance. Junction-capacitance can be reduced by up to 65%; while the subthreshold slope of the device becomes steeper (which suppresses short-channel effects).[50] The dramatic improvement in punchthrough-resistance does not depend on extreme scaling of the S/D junction-depth or n-well background-doping. The reduced subthreshold-slope and junction-capacitance directly result from carefully controlled doping underneath the channel at the gate edge and the near elimination of all other undesired doping that results from less controllable doping methods.

Because of their lower-diffusivity and as-implanted-straggle, higher mass dopants can be placed more precisely in the channel. Consequently, indium (In) and antimony (Sb) are being increasingly used for this application.[51,52] Punchthrough-stop implants were the primary production application of In and Sb in ion-implantation in the 2000 timeframe. The lower susceptibility of these species to TED also helps reduce the reverse-short-channel-effect (the slight increase in threshold-voltage with decreasing effective-channel-length that occurs before punchthrough-effects dominate the smallest channel-lengths).

One proposed structure for MOSFETs with an effective gate length (L_E) of 0.25-μm is shown in Fig. 5-29.[53] This device has a very heavily doped source/drain next to a very heavily doped pocket-implant. Since the gate-length is so small, the pocket-implanted-region actually doubles for the traditional retrograde-well-implant. This aggressive device design shows the challenges of doping a 25-nm MOSFET. Very-high doping levels and very-steep doping gradients are required – with a minimum of thermal processing.

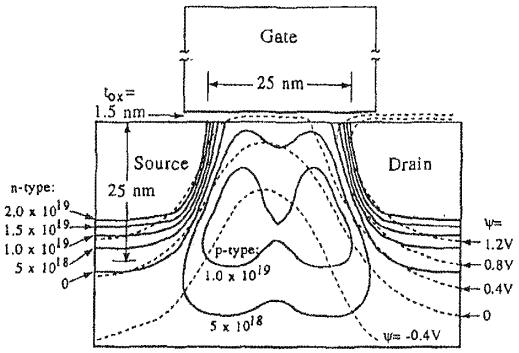


Fig. 5-29 Doping contours in 25-nm PMOSFET.[53]

An issue associated with the use of In as a pocket-implant species has recently been reported. That is, an In implant at the doses needed for the pocket-implant process causes a buried amorphous layer to form. This layer has an interface with the single-crystal silicon. The silicon at these interfaces is still crystalline, but is heavily damaged. The indium segregates to these damaged regions and becomes

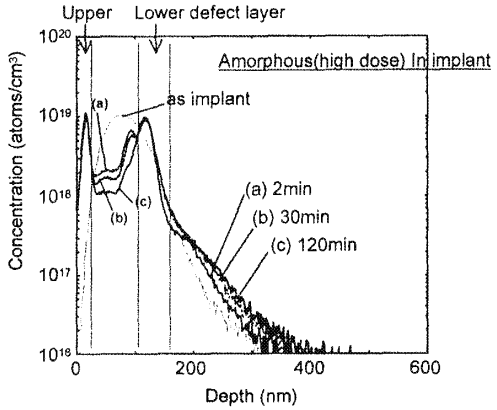


Fig. 5-30 Indium SIMS profiles for $1x10^{14}/cm^2$ implants at 200 keV annealed at 750°C for 2, 30, and 120 min in nitrogen. This figure shows large dose loss for indium. The depth of the defect layer (measured by XTEM) is shown (upper & lower defect layers).[54]

electrically de-activated (negating its ability to provide punchthrough control).[54,55] Figure 5-230 shows a SIMS profile of the In as-implanted, and after annealing. The loss of indium from its peak is seen. A double-implanted extension region (As + Sb) has been reported to be able to control the damaged region near the In pocket, mitigating the indium segregation and de-activation.[56]

REFERENCES

1. R.D. Rung, C.J. Dell'Oca, and L.G. Walker, "A Retrograde *p*-Well for High Density CMOS," *IEEE Trans. Electron Dev.,* **ED-28**, p. 1115, 1981.

2. D.M. Brown *et al.,* "Trends in Advanced CMOS Process Technology," *Proceedings of the IEEE,* December 1986, p. 1646.

3. G.J. Hu *et al.*, "Design and Fabrication of *p*-channel FET for 1-μm CMOS Technology," *IEDM Tech. Dig.* 1982, p. 710.

4. R. Martin and J.Y. Chen, "Optimized Retrograde *n*-Well for 1 μm CMOS," *Proc. Custom Integ. Circuits Conf.,* 1985, p. 199.

5. J.Y. Chen, "CMOS - The Emerging Technology," *IEEE Circuits and Devices Magazine,* March 1986, p. 16.

6. Y. Mochizuki, "Special Report: High Energy Ion Implantation Technology," Nikkei Microelectronics (Dec. 1991), p. 92.

7. K. Tsukamoto *et al.,* "High Ion Energy Implantation for ULSI: Well Engineering and Gettering," *Solid State Technology,* June 1992, p. 35.

8. R. Wijberg *et al.,* "Strong Asymmetric Doping Profiles at Mask Edges in High-Energy Ion Implantation, "*IEEE Trans. Electron Dev.,* **37-1**, January 1990, p. 79.

9. T.M. Parrill and M.J. Namaroff,: "Ion Implantation Productivity for Advanced Applications," *European Semiconductor*, July 1999, p. 44.

10. B. Adibi and V. Meniailenkov, "Transistor Performance: The Impact of Implant Doping Accuracy," *Solid State Technology,* January 2001, p. 88.

11. S. Takagi *et al.,* "On the Universality of the Inversion Layer Mobility in Silicon MOSFETs: Part I-Effects of Substrate Impurity Concentration," *IEEE Trans. Electron Dev.*, **41**, 2357 (1994).

12. P. Boullion *et al.,* "Re-Examination of Indium Implantation for a Low Power 0.1 μm Technology," *Tech. Dig. IEDM*, 1995, p. 897.

13. S.C. Williams *et al.,* "Scaling Trends for Device Performance and Reliability in Channel-Engineered NMOSFETs," *IEEE Trans. Electron Dev.*, **45**, January 1998, p. 254.

14. I. De and C.M. Osburn, "Impact of Super-Steep-Retrograde Channel Doping Profiles on the Performance of Scaled Devices," *IEEE Trans. Electron Dev.*, **46**, August 1999, p. 1711.

15. S. Venkatesan *et al.,* "Device Drive Current Degradation Observed with Retrograde Channel Profiles," *Tech. Dig. IEDM* 1995, p. 419.
16. M. Harlan and K. Petry, *Ion Implantation Technology*, 1998, June 22-26, Kyoto, Japan.
17. T. Ghani *et al.,* "100 nm Gate Length High Performance/Low Power CMOS Transistor Structure," *Tech. Dig. IEDM,* 1999, p. 415.
18. T. Mizuno, *IEEE Trans. Electron Dev.,* **47**, 756 (2000).
19. Ohguro *et al.,*"Undoped Epitaxial Si Channel in NMOSFET Grown by UHV-CVD with Preheating," *IEEE Trans. Electron Dev.,* **45**, March, 1998, p. 710.
20. International Technology Roadmap for Semiconductors, 2001 Edition, Semiconductor Industry Association, San Jose, CA.
21. K.K. Ng and W.T. Lynch,"Analysis of the Gate-Voltage-Dependent Series Resistance of MOSFETs," *IEEE Trans. Electron Dev.*, July 1986, p. 965.
22. G. Baccarani and G.A. Sai-Halasz, "Spreading Resistance in Submicron MOSFETs," *IEEE Electron Dev. Lett.,* **EDL-4**, 1983, p. 27.
23. W.C. Holton *et al., "*Overview of Semiconductor Devices,*"* Ch. 1 in *Handbook of Semiconductor Manufacturing Technology,* Y. Nishi and R. Doering, Eds., Marcel Dekker, New York, 2000, p. 1.
24. K. Miyano *et al.,* "Low Thermal Budget Elevated Source/Drain Technology Utilizing Novel Solid Phase Epitaxy and Selective Vapor Phase Etching," *Tech. Dig. IEDM* 2000, p. 433.
25. D.W. Duff and L.M. Rubin, "Ion Implant Equipment Challenges for 0.18 μm and Beyond," *Solid State Technology,* June 1998, p. 83.
26. T. Parrill *et al.,* "A New Era for High-Current, Low-Energy Ion Implantation," *Solid State Technoloy,* November 2000, p. 103.
27. H.J. Gossmann *et al.,* "Junctions for Deep-Sub-100-nm NMOS: How Far Will ion Implant Take Us?" *Symp. on Front End Processing* at the *Spring 2000 Meeting of the Materials Research Soc.,* April, 2000.
28. L. Peters, "Ramping the 0.13 μm Generation," *Semiconductor International*, January 2001, p. 56.
29. A. Agarwal *et al.,* "Reduction of Transient Diffusion from 1-5 keV Si^+ Ion Implantation Due to Surface Annihilation of Interstitials," *Appl. Phys. Letts.*, **71**, p. 3141 (1997).
30. A. Agarwal *et al., Mat. Res. Soc. Symp. Proc.*, **568**, p. 19 (1999).
31. A. Agarwal *et al.,* "Ultra-Shallow Junction Formation Using Ion Implantation and Rapid Thermal Annealing: Physical and Practical Limits," *Abs. Mtg. Electrochem. Soc.,* Spring 2000, Abs. 526.
32. A. Agarwal *et al.,* "Ultra-Shallow Junctions by Ion Implantation and Rapid Thermal Annealing: Spike Anneals, Ramp-Rate Effects," *Mater. Res. Soc. Symp.*, **19** p. 568 (1999).

33. P.J Timans *et al.,* "Rapid Thermal Processing," Ch. 9 **in** *Handbook of Semiconductor Manufacturing Technology,* Y. Nishi and R. Doering, Eds., Marcel Dekker, New York, 2000, pp. 201-86.
34. *Advances in Rapid Thermal and Integrated Processing,* F. Roozeboom Ed., Klewer Academic Publishers, Dordrecht, NL, 1996.
35. R.B. Fair, *Rapid Thermal Processing,* Academic Press, San Diego, CA, 1993.
36. M.A. Foad and D. Jennings, "Formation of Ultra-Shallow Junctions by Ion Implantation and RTA," *Solid State Technology,* December 1998, p. 43.
37. A. Atanos *et al.,*"Susceptor-Based Rapid Thermal Processing for Ultra-Shallow Junctions," *Solid State Technology,* April 2000, p. 50.
38. M. Ameen and J. Hebb, "Implanter, RTP Systems Issues for Ultrashallow Junction Formation," *Solid State Technology,* September 2001, p. 77.
39. B. Yu *et al.,*"70-nm MOSFET with Ultra-Shallow, Abrupt, and Super-Doped S/D Extension Implanted by Laser Thermal Process (LTP)," *Tech. Dig. IEDM* 1999, p. 509.
40. S. Banerjee,"High Ramp Rate Thermal Annealing for Ultra-Shallow Junctions," *Abs. Mtg. Electrochem. Soc.,* Spring 2000, Abs. 521.
41. D.M. Camm and M. LeFrancois,"Spike Thermal Processing Using Arc Lamps," *Abs. Mtg. Electrochem. Soc.,* Spring 2000, Abs. 567.
42. S. Thompson *et al.,* "MOS Scaling: Transistor Challenges for the 21st Century," *Intel Technology Journal,* 3rd Quarter 1998.
43. C. Osburn *et al.,* "Design and Integration Considerations of End-of-the-Roadmap Ultrashallow Junctions," *J. Vac. Sci. Technol.***B-18**, p. 338, 2000.
44. T. Ghani *et al., Dig. Tech. Papers, Symp. VLSI Technol.* 2000,
45. M.I. Current *et al.,* "Process Integration Issues for Doping of Ultra-Shallow Doping Junctions," *J. Vac. Sci. Technol.* **B-18**, p. 468, 2000.
46. U. Jeong *et al.,* "Devices Dictate Control of Implant-Beam Incident Angle," *Solid State Technology,* October 2001, p. 79.
47. R.N. Kleiman *et al., J. Vac. Sci. Technol.*, **B-18**, p. 2034 (2000).
48. H. Fukutome *et al.,* "Anomalous Diffusion in the Extension Region of Nanoscale MOSFETs," *Tech. Dig. IEDM,* 2001, p. 67.
49. T. Hori, "A 0.1 μm CMOS Technology with Tilt Implanted Punchthrough Stopper (TIPS)," *Tech. Dig. IEDM* 1994, p. 75.
50. M. Rodder *et al.,* "A Scaled 1.8 V, 0.18 μm Gate Length CMOS Technology: Device Design and Reliability Concerns," *Tech. Dig. IEDM* 1995, p. 415.
51. K. Miyashita *et al.,* "Optimized Halo Structure for 80-nm Physical Gate CMOS Technology with Indium and Antimony Highly Angled Ion Implantation," *Tech. Dig. IEDM* 1999, p. 645.
52. Y. Momiyama *et al.,* "Indium Tilted Channel Implantation Technology for 60 nm NMOSFET," *Dig. Tech. Papers, Symp. VLSI Technol.* 1999, p. 67.

53. Y. Taur, "25 nm CMOS Design Considerations," *Tech. Dig. IEDM,* 1998, p. 789.
54. T. Noda, "Modeling of End-of-Range (EOR) Defects for Channel Engineering," *Tech. Dig. IEDM*, 2001, p. 839.
55. T. Noda *et al.*,"Effects of End-of-Range Dislocation Loops on Transient-Enhanced Diffusion of Indium Implanted in Silicon," *J. Appl Phys.*, **88**, 4980 2000.
56. H.C.-H. Wang *et al.*,"Antimony Assisted Arsenic S/D Extension Engineering for Sub-0.1 μm NMOSFETs: A Novel Approach to Steep and Retrograde Indium Pocket Implants," *Tech. Dig. IEDM* 2001, p. 63.
57. S. Narasimba *et al.*, "High Performance Sub-40 nm CMOS Devices on SOI for the 70 nm Technology Node," *Tech. Dig. IEDM* 2001, p. 625.
58. R. Chau *et al.*, "A 50-nm Depleted Substrate CMOS Transistor (DST)," *Tech. Dig. IEDM,* 2001, p. 621.
59. M.I. Current *et al.*, "Ultrashallow Junctions or Ultrathin SOI?" *Solid State Technology*, September 2000, p. 66.