



Nanoscale CMOS VLSI Circuits

Design for Manufacturability

Sandip Kundu
Aswin Sreedhar

Nanoscale CMOS VLSI Circuits

Design for Manufacturability

About the Authors

Sandip Kundu, Ph.D., is a professor in the Electrical and Computer Engineering Department at the University of Massachusetts at Amherst specializing in VLSI design and testing. Previously, he was a Principal Engineer at the Intel Corporation and Research Staff Member at the IBM Corporation. He is a fellow of the IEEE, has been a distinguished visitor of the IEEE Computer Society and associate editor of the IEEE Transactions on Computers and the IEEE Transactions on VLSI Systems. Dr. Kundu has published more than 130 technical papers and holds 12 patents.

Aswin Sreedhar, Ph.D., is a research associate in the Electrical and Computer Engineering Department at the University of Massachusetts at Amherst. His research interests are in statistical techniques for design for manufacturability of VLSI Systems and circuit reliability. Previously, he was a graduate intern at the Intel Corporation and AMD. He is a recipient of the best paper award for lithography based yield modeling (2009) at the DATE conference.

Nanoscale CMOS VLSI Circuits

Design for Manufacturability

Sandip Kundu

Aswin Sreedhar



New York Chicago San Francisco Lisbon
London Madrid Mexico City Milan New Delhi
San Juan Seoul Singapore Sydney Toronto

Copyright © 2010 by The McGraw-Hill Companies, Inc. All rights reserved. Except as permitted under the United States Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

ISBN: 978-0-07-163520-2

MHID: 0-07-163520-3

The material in this eBook also appears in the print version of this title: ISBN: 978-0-07-163519-6, MHID: 0-07-163519-X.

All trademarks are trademarks of their respective owners. Rather than put a trademark symbol after every occurrence of a trademarked name, we use names in an editorial fashion only, and to the benefit of the trademark owner, with no intention of infringement of the trademark. Where such designations appear in this book, they have been printed with initial caps.

McGraw-Hill eBooks are available at special quantity discounts to use as premiums and sales promotions, or for use in corporate training programs. To contact a representative please e-mail us at bulksales@mcgraw-hill.com.

Information contained in this work has been obtained by The McGraw-Hill Companies, Inc. (“McGraw-Hill”) from sources believed to be reliable. However, neither McGraw-Hill nor its authors guarantee the accuracy or completeness of any information published herein, and neither McGraw-Hill nor its authors shall be responsible for any errors, omissions, or damages arising out of use of this information. This work is published with the understanding that McGraw-Hill and its authors are supplying information but are not attempting to render engineering or other professional services. If such services are required, the assistance of an appropriate professional should be sought.

TERMS OF USE

This is a copyrighted work and The McGraw-Hill Companies, Inc. (“McGrawHill”) and its licensors reserve all rights in and to the work. Use of this work is subject to these terms. Except as permitted under the Copyright Act of 1976 and the right to store and retrieve one copy of the work, you may not decompile, disassemble, reverse engineer, reproduce, modify, create derivative works based upon, transmit, distribute, disseminate, sell, publish or sublicense the work or any part of it without McGraw-Hill’s prior consent. You may use the work for your own noncommercial and personal use; any other use of the work is strictly prohibited. Your right to use the work may be terminated if you fail to comply with these terms.

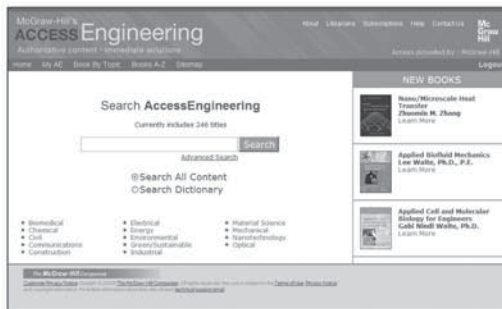
THE WORK IS PROVIDED “AS IS.” MCGRAW-HILL AND ITS LICENSORS MAKE NO GUARANTEES OR WARRANTIES AS TO THE ACCURACY, ADEQUACY OR COMPLETENESS OF OR RESULTS TO BE OBTAINED FROM USING THE WORK, INCLUDING ANY INFORMATION THAT CAN BE ACCESSED THROUGH THE WORK VIA HYPERLINK OR OTHERWISE, AND EXPRESSLY DISCLAIM ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. McGraw-Hill and its licensors do not warrant or guarantee that the functions contained in the work will meet your requirements or that its operation will be uninterrupted or error free. Neither McGraw-Hill nor its licensors shall be liable to you or anyone else for any inaccuracy, error or omission, regardless of cause, in the work or for any damages resulting therefrom. McGraw-Hill has no responsibility for the content of any information accessed through the work. Under no circumstances shall McGraw-Hill and/or its licensors be liable for any indirect, incidental, special, punitive, consequential or similar damages that result from the use of or inability to use the work, even if any of them has been advised of the possibility of such damages. This limitation of liability shall apply to any claim or cause whatsoever whether such claim or cause arises in contract, tort or otherwise.

McGraw-Hill's
AccessEngineering
 Authoritative content • Immediate solutions

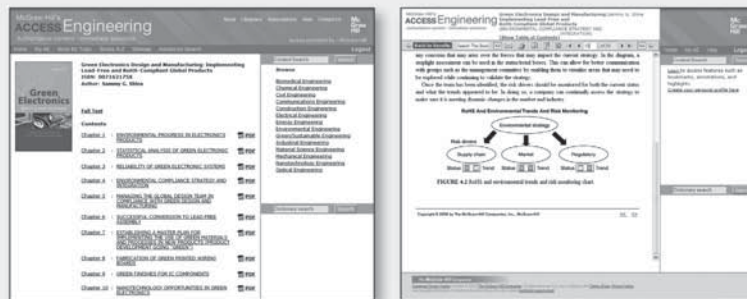
AccessEngineering offers the complete contents of hundreds of outstanding McGraw-Hill books, including *Marks' Standard Handbook for Mechanical Engineers*, *Perry's Chemical Engineers' Handbook*, and *Roark's Formulas for Stress and Strain*, with new books added biweekly. This dynamic source of world-renowned engineering content supports all levels of scientific and technical research in the corporate, industrial, government, and academic sectors.

Focused around 14 major areas of engineering, **AccessEngineering** offers comprehensive coverage and fast title-by-title access to our engineering collection in the following subject areas:

- / Biomedical
- / Chemical
- / Civil
- / Communications
- / Construction
- / Electrical
- / Energy
- / Environmental
- / Green/Sustainable
- / Industrial
- / Material Science
- / Mechanical
- / Nanotechnology
- / Optical



In addition, sophisticated personalization tools allow content to be easily integrated into user workflow. A science and engineering dictionary containing more than 18,000 terms is included in a fully searchable, taxonomically organized database.



For more information on individual and institutional subscriptions, please visit www.accessengineeringlibrary.com

Learn more.  Do more.

*To my late father Prof. Hari Mohan Kundu,
whose encouragement for pursuit of excellence still endures,
and to my mother Mrs. Pravati Kundu,
who has been a pillar of support and strength,
my loving wife Deblina and
my daughters Shinjini and Shohini*

—Sandip

*To my father Mr. Sreedhar Jagannathan
and my mother Mrs. Amirthavalli Sreedhar,
whose sacrifice and ever present support drives
my thirst for knowledge,
my loving wife Srividhya and
my brother Parikshit*

—Aswin

This page intentionally left blank

Contents

Preface	xiii
1 Introduction	1
Technology Trends: Extending Moore's Law	1
Device Improvements	3
Silicon on Insulator	4
Multigate Devices	5
Nanodevices	6
Contributions from Material Science	7
Low-K and High-K Dielectrics	7
Strained Silicon	9
Deep Subwavelength Lithography	10
Mask Manipulation Techniques	12
Increasing Numerical Aperture	14
Design for Manufacturability	15
Value and Economics of DFM	15
Variabilities	18
The Need for a Model-Based DFM Approach	23
Design for Reliability	23
Summary	24
References	25
2 Semiconductor Manufacturing	27
Introduction	27
Patterning Process	28
Photolithography	29
Resist Coat	30
Preexposure (Soft) Bake	30
Mask Alignment	31
Exposure	32
Postexposure Bake (PEB)	32
Development	32
Hard Bake	33
Etching Techniques	33
Wet Etching Techniques	33
Dry Etching Techniques	35
Optical Pattern Formation	36
Illumination	37
Diffraction	40

viii Contents

Imaging Lens	45
Exposure System	47
Aerial Image and Reduction Imaging	48
Resist Pattern Formation	51
Partial Coherence	53
Lithography Modeling	55
Phenomenological Modeling	56
Hopkins Approach to Partially Coherent Imaging	56
Resist Diffusion	57
Simplified Resist Model	57
Sum-of-Coherent-Systems Approach	58
Fully Physical Resist Modeling	59
Summary	60
References	61
3 Process and Device Variability: Analysis and Modeling	63
Introduction	63
Gate Length Variation	70
Patterning Variations Due to Photolithography	70
Proximity Effects	71
Defocus	74
Lens Aberration	75
Modeling Nonrectangular Gates (NRGs)	80
Line Edge Roughness: Theory and Characterization	82
Gate Width Variation	87
Atomistic Fluctuations	88
Thickness Variation in Metal and Dielectric	91
Stress-Induced Variation	96
Summary	99
References	99
4 Manufacturing-Aware Physical Design Closure	103
Introduction	103
Control of the Lithographic Process Window	108
Resolution Enhancement Techniques	113
Optical Proximity Correction	114
Subresolution Assist Features	118
Phase Shift Masking	120
Off-Axis Illumination	124

Physical Design for DFM	126
Geometric Design Rules	127
Restrictive Design Rules	127
Model-Based Rules Check and Printability Verification	129
Manufacturability-Aware Standard Cell Design	131
Mitigating the Antenna Effect	135
Placement and Routing for DFM	138
Advanced Lithographic Techniques	141
Double Patterning	142
Inverse Lithography	148
Other Advanced Techniques	153
Summary	153
References	153
5 Metrology, Manufacturing Defects, and	
Defect Extraction	157
Introduction	157
Process-Induced Defects	161
Classification of Error Sources	162
Defect Interaction and Electrical Effects	164
Modeling Particle Defects	166
Defining Critical Area and Probability of Failure	167
Critical Area Estimation	169
Particulate Yield Models	172
Layout Methods to Improve Critical Area	173
Pattern-Dependent Defects	175
Pattern-Dependent Defect Types	176
Pattern Density Problems	178
Statistical Approach to Modeling Patterning Defects	179
Case Study: Yield Modeling and Enhancement for Optical Lithography	179
Case Study: Linewidth-Based Yield When Considering Lithographic Variations	181
Layout Methods That Mitigate Patterning Defects	184
Metrology	186
Precision and Tolerance in Measurement	187
CD Metrology	188
Scanning Electron Microscopy	188

x Contents

Electrical CD Measurement	190
Scatterometry	193
Overlay Metrology	195
Other In-Line Measurements	197
In-Situ Metrology	198
Failure Analysis Techniques	199
Nondestructive Techniques	201
Failure Verification	201
Optical Microscopy	201
X-Ray Radiography	202
Hermeticity Testing	202
Particle Impact Noise Detection	202
Destructive Techniques	203
Microthermography	203
Decapsulation	203
Surface Analysis	203
Summary	204
References	204
6 Defect Impact Modeling and Yield Improvement	
Techniques	207
Introduction	207
Modeling the Impact of Defects on	
Circuit Behavior	209
Defect-Fault Relationship	210
Role of Defect-Fault Models	211
Defect-Based Fault Models	212
Defect-Based Bridging Fault Model	214
Abstract Fault Models	215
Hybrid Fault Models	218
Test Flow	218
Yield Improvement	221
Fault Tolerance	222
Traditional Structural Redundancy	
Techniques	222
Nonstructural Redundancy	
Techniques	227
NAND Multiplexing	230
Reconfiguration	231
Comparison of Redundancy-Based Fault	
Tolerance Techniques	232
Fuses	233
Fault Avoidance	235
Summary	239
References	241

7	Physical Design and Reliability	243
	Introduction	243
	Electromigration	247
	Hot Carrier Effects	250
	Hot Carrier Injection Mechanisms	251
	Device Damage Characteristics	253
	Time-Dependent Dielectric Breakdown	254
	Mitigating HCI-Induced Degradation	255
	Negative Bias Temperature Instability	256
	Reaction-Diffusion Model	257
	Static and Dynamic NBTI	258
	Design Techniques	260
	Electrostatic Discharge	261
	Soft Errors	263
	Types of Soft Errors	263
	Soft Error Rate	263
	SER Mitigation and Correction for Reliability	264
	Reliability Screening and Testing	264
	Summary	265
	References	265
8	Design for Manufacturability: Tools and Methodologies	269
	Introduction	269
	DFx in IC Design Flow	270
	Standard Cell Design	271
	Library Characterization	272
	Placement, Routing, and Dummy Fills	274
	Verification, Mask Synthesis, and Inspection	275
	Process and Device Simulation	275
	Electrical DFM	276
	Statistical Design and Return on Investment	277
	DFM for Optimization Tools	279
	DFM-Aware Reliability Analysis	282
	DFx for Future Technology Nodes	283
	Concluding Remarks	284
	References	285
	Index	287

This page intentionally left blank

Preface

The purpose of this book is to introduce readers to the world of design for manufacturability and reliability. It is intended to be used as a text for senior-level undergraduates and for graduate students in their initial years and also to serve as a reference for practicing design engineers. Because there are entire conferences and journals devoted to this subject, it is impossible for any compendium to be complete or fully current. Therefore, we focus more on principles and ideas than on the granular details of each topic. There are references at the end of each chapter that direct the reader to more in-depth study. In order to understand this book, readers should have some knowledge of VLSI design principles, including cell library characterization and physical layout development.

This book is a result of the research interests of both coauthors, who have published actively in the area of design for manufacturability. Professor Kundu also introduced a new course on Design for Manufacturability and Reliability at the University of Massachusetts. Much of this book's organization is based on the structure of that course, which was developed for classroom instructions. Thus, it is hoped that students will benefit greatly from this book. The text also deals extensively with costs, constraints, computational efficiencies, and methodologies. For this reason, it should also be of value to design engineers.

The material is presented in eight chapters. Chapter 1 introduces the reader to current trends in CMOS VLSI design. It offers a brief overview of new devices and of contributions from material sciences and optics that have become fundamental for the design process achieving higher performance and reduced power consumption. The basic concepts of design for manufacturability (DFM) are reviewed along with its relevance to and application in current design systems and design flows. The chapter also explores reliability concerns in nano-CMOS VLSI designs from the perspective of design for reliability (DFR), computer-aided design (CAD) flows, and design optimizations to improve product lifetime.

Chapter 2 discusses the preliminaries of semiconductor manufacturing technology. The various steps—which include oxidation,

diffusion, metal deposition, and patterning—are explained. This chapter concentrates chiefly on patterning steps that involve photolithography and etching processes. Techniques are discussed for modeling the photolithography system so that the manufacturability of a given design can be effectively analyzed. The techniques are classified as either phenomenological or fully physical, and the methods are compared in terms of accuracy and computational efficiency.

The focus of Chapter 3 is variability in the process parameters for current and future CMOS devices and the effects of this variability. The main subjects addressed are variations in patterning, dopant density fluctuation, and dielectric thickness variation due to chemical-mechanical polishing and stress.

Chapter 4 explains the fundamentals of lithographic control through layout-based analysis as well as the important photolithography parameters and concepts. Lithographic variability control is illustrated by resolution enhancement techniques such as optical proximity correction, phase shift masking, and off-axis illumination. This chapter discusses components of the DRM manual, including geometric design rules, restricted design rules, and antenna rules. It also contains sections on the evolution of model-based design rules check and other changes to CAD tools for traditional physical design. The chapter concludes with a presentation of advanced lithography techniques, such as dual-pattern lithography, inverse lithography technology and source mask optimization, that are used to push the resolution limit.

Chapter 5 provides an in-depth look at various manufacturing defects that occur during semiconductor manufacturing. These defects are classified as resulting either from contaminants (particulate defects) or from the design layout itself (pattern dependent). The chapter describes how critical area is used to estimate yields for particle defects as well as how linewidth-based models are used to estimate yield for pattern-dependent defects. Metrology and failure analysis techniques—and their application to semiconductor measurement for process control—are also described.

In Chapter 6, particle defects and pattern-based defects are examined in terms of their impact on circuit operation and performance. The discussion covers the defect models and fault models used to effectively identify and predict design behavior in the presence of defects. This chapter also explores yield improvement for designs through fault avoidance and fault tolerance techniques.

The physics of reliability issues and their impacts are discussed in Chapter 7. Reliability mechanisms such as hot carrier injection, negative temperature bias instability, electromigration, and electrostatic discharge (ESD) are explained and illustrated. The mean time to failure for each of these reliability failure mechanisms are also discussed with design solution to mitigate their effects.

Finally, Chapter 8 addresses the changes to CAD tools and methodologies due to DFM and DFR approaches at different stages of the circuit realization process, including library characterization, standard cell design, and physical design. This chapter then delves into the need for statistical design approaches and model-based solutions to DFM-DFR problems. The importance of reliability-aware DFM approaches for future designs is also detailed.

The central theme of this book is that decisions made during the design process will affect the the product's manufacturability, yield, and reliability. The economic success of a product is tied to yield and manufacturability, which traditionally have been based solely on the effectiveness and productivity of the *manufacturing* house. Throughout this book, readers are shown the impact of the *design* methodology on a product's economic success.

Sandip Kundu
Aswin Sreedhar

This page intentionally left blank

Nanoscale CMOS VLSI Circuits

Design for Manufacturability

This page intentionally left blank

CHAPTER 1

Introduction

1.1 Technology Trends: Extending Moore's Law

Complementary metal oxide semiconductor (CMOS) technology has been the dominant semiconductor manufacturing technology for more than two decades. The demands for greater integration, higher performance, and lower power dissipation have been driving the scaling of CMOS devices for more than a quarter century. In 1965, Gordon Moore famously predicted that the number of transistors on a chip would double every 18 months while the price remained the same. Known as Moore's law, the prediction has held true until today. This has been made possible by advances in multiple areas of technology. Design technologies such as high-level design languages, automatic logic synthesis, computer-aided circuit simulation, and physical design have enabled increasingly larger designs to be produced within a short time. Manufacturing technologies—including mask engineering, photolithography, etching, deposition, and polishing—have improved continuously to enable higher levels of device integration.

Moore's law has been sustained by concurrent improvement across multiple technologies. There has always been a demand for high-performance circuits that are cheap to produce and consume less power. Historically, scaling of transistor feature size has offered improvement in all three of these areas. "High performance" means an increase in clock frequency with every new generation of technology. Such increases are possible only with commensurate increases in transistor drive current while parasitic capacitances remain low, thus reducing propagation delay. The transistor drive current is a function of the gate dimensions and the number of charge carriers and their mobility in the channel region. *Propagation delay* is the input-to-output signal transition time of a gate or a transistor. The propagation delay depends primarily on the intrinsic capacitances of the device: the threshold voltage and load capacitance. By making changes to the contributing parameters, an integrated circuit (IC) with higher clock frequency can be produced. With the foray of semiconductor chips into portable and handheld devices, power

2 Chapter One

consumption has also become a critical design parameter. Power consumption can be divided into dynamic power and static power. *Dynamic power* is the power dissipated during transistor operation. It is dependent on the supply voltage and frequency of circuit operation. It also depends on device and interconnect parasitic capacitances, which in turn depend on the manufacturing process and materials. *Static power* is the power consumed irrespective of the device use. It is chiefly dependent on the threshold voltage of a device, which in turn is related to process parameters such as dopant density in the gate poly and the transistor channel region.

Improvements in clock frequency have been achieved primarily by creating device of smaller effective channel length and reducing the threshold voltage of the device. Power consumption and some device reliability problems have been kept under control by reducing the supply voltage. Leakage control has been attained through selectively increasing the threshold voltage of transistors in noncritical gates. This is also known as multithreshold CMOS (MTCMOS) technology.

In each generation of CMOS technology, new challenges have cropped up that required new solutions. In early days, when the layout size increased to about a thousand polygons, manual layout became impractical, so layout automation tools were required. Later, when the gate count increased beyond thousands, computer-aided logic synthesis became necessary. When the gate counts became still larger, high-level design languages were invented. At about 1.5 μm or so, interconnect delays became a concern. This led to development of interconnect resistance and capacitance (RC) extraction from circuit layout. With continued scaling, coupling capacitances created problems that required development of signal integrity tools. As the feature size became even smaller, considerations of device reliability required the introduction of electromigration rules and checks. As the transistor count increased and the frequencies became larger, power density became a critical concern—especially in mobile designs. The reduction of dynamic power demanded commensurate reductions in interconnect capacitances, which in turn required development of low-K interlayer dielectric between interconnect layers.

Collectively, these techniques have helped to move the industry from one technology node to the next. Future technologies will require innovation in (1) basic device structures, (2) materials, and (3) processing technologies.

The move into transistors feature widths close to 45 nanometers and below, unlike any other shift in technology, has led to the simultaneous investigation of many changes to design and manufacturing that attempt to better satisfy the three technology requirements just listed. As identified in the *International Technology Roadmap for Semiconductors* (ITRS) report,¹ the basic transistor patterning technology will become a critically important factor in

enabling future technologies. Photolithography, the core of transistor and interconnect patterning, has been found wanting at lower feature sizes. The main concerns are materials for mask and projection optics, temperature of operation, and wavelength of light source. (These issues are detailed in Chapter 3.) Lithography techniques that help produce devices of smaller effective gate lengths are central to technology scaling. The emergence of new, structurally modified metal oxide semiconductor field-effect transistor (MOSFET) devices aim at increasing the channel region's surface area in order to improve the device's drive current is also impelling the need for better manufacturing techniques. These devices also need the help of lithography to be printed on wafer. Newer materials that can provide higher performance (through increased mobility) and better variability control (through reduced intrinsic and interconnect capacitances) also aim to help break the performance and power consumption barriers. We address all these issues in the following sections.

1.1.1 Device Improvements

Conventional CMOS scaling involves scaling of multiple aspects of a transistor; these include feature length and oxide thickness as well as dopant density and profile. As we approach atomic scales for transistors, scaling of these aspects presents a new set of challenges. For example, scaling oxide thickness increases tunneling leakage through oxide. Increased channel doping increases source-drain leakage. Increased source-drain doping increases band-to-band direct tunneling leakage to bulk. Increased source-drain doping also increases the source-drain capacitance, compromising the performance of transistors. It is widely recognized that conventional scaling of bulk CMOS will encounter these difficulties and that further scaling of transistors will require modification to the conventional MOS transistor. Several alternative devices are now under investigation or actual use. They include silicon-on-insulator (SOI) MOSFET, whose design seeks to mitigate the source-drain capacitance and transistor body effects. The FinFET (an FET with a finlike vertical rather than a planar structure) and tri-gate transistors being developed seek to increase transistor ON current without increasing the OFF current. Transistors based on carbon nanotube (CNT) technology offer another alternative to device scaling. However, it is not yet patternable using current lithography processes.

Several foundries are currently manufacturing SOI-based MOSFETS used in high-performance ICs. Although the wafer cost for SOI has come down significantly, cost and yield still remain as barriers to wider adoption of SOI devices. Furthermore, FinFET, tri-gate, and other multigate devices are in the early stages of development. Once they become practical to manufacture in high volume, such devices will impel changes to conventional circuit design and optimization methods. Carbon nanotubes have been touted as a potential

4 Chapter One

replacement technology for silicon transistors, but assembly, performance, and reliability remain open issues for such devices. The following sections provide an overview of these developing device technologies.

1.1.1.1 Silicon on Insulator

The SOI process uses a silicon-insulator-silicon substrate in place of a doped silicon substrate as the bulk material in MOSFET devices. A thin layer of buried SiO_2 isolates the channel regions of transistors, which enables the transistor body to float without risk of substrate coupling noise. This, in turn, reduces capacitance to body and improves both circuit performance and power attributes. Figure 1.1 depicts an SOI-based MOSFET device. From a circuit perspective, one difference between an SOI and a bulk transistor is that the body of an SOI transistor forms an independent fourth terminal of the device. This terminal is typically left unconnected, but it can be fixed at a potential to control the threshold voltage. The main advantage of using SOI over bulk CMOS is the increase in performance due to lower junction capacitance, lower body effect, and improved saturation current. Other advantages include better control of device leakage over the entire die, reduced susceptibility to soft error, and low temperature sensitivity. A partially depleted SOI process is often used to deliver both high performance and reduced power consumption, where “partially depleted” means that the channel inversion region of the SOI device does not completely consume the body region.^{2,3} Partially depleted SOI uses the same materials, tools, processes, and parameter specifications as the older bulk technology. It also uses the same design technology and computer-aided design (CAD) tools, and changes in CAD tools for circuit simulation and physical design are easily accommodated within current framework. This fact facilitates quick adoption of this technology.

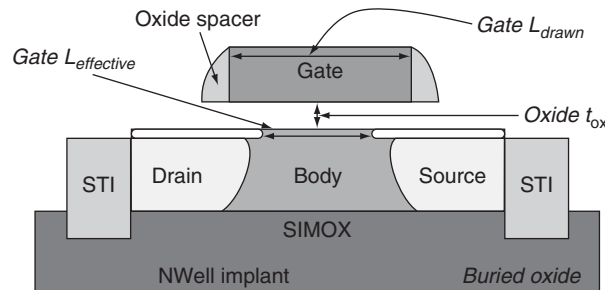


FIGURE 1.1 Partially depleted SOI MOSFET.

1.1.1.2 Multigate Devices

A group of multigate devices that includes FinFETs, DG-FETs (“DG” denotes “double-gate”), and tri-gates has been touted to replace the traditional MOSFET. Multigate devices are MOSFETs that have multiple gate terminals. The gate terminals can be joined together to form a single connection that performs the same operation as a MOSFET, or they can be controlled independently to offer circuit designers greater flexibility. In this latter configuration the devices are simply termed *independent gate* FET devices. Multigate transistors are being manufactured solely to create ever-smaller transistors that can provide higher performance with the same or smaller chip area. These devices can be classified based on the direction of gate alignment. Horizontally aligned multigate devices are called *planar* gate devices, which can be double-gate or multibrIDGE transistors with common or independent gate controls. Both FinFETs and tri-gates are *vertical* gates, which (for manufacturing reasons) must all have the same height. This constraint forces all transistors to have the same width. Consequently, current approaches to device sizing and circuit optimization techniques must be tweaked to accommodate discrete transistor sizes. Planar double-gate devices form a natural extension to SOI technology, and they can be manufactured using any of three conventional techniques: (1) the layer-on-layer process, (2) wafer-to-wafer bonding, and (3) the suspended channel process. The channel region of the device is sandwiched between two independently fabricated gate terminals with oxide stacks. Figure 1.2 shows a planar double gate fabricated using the layer-on-layer technique.

A FinFET is a nonplanar vertical double-gate transistor whose conducting channel is formed by a thin polysilicon “fin” structure that wraps around the body of the device.^{4,5} The dimensions of the fin dictate the channel length of the device. Figure 1.3(a) illustrates the structure of a FinFET. As shown, the gate region is formed over the silicon that connects the drain and source regions. Transistors with effective gate length of 25 nm have been reported.⁵ Tri-gate transistors were devised by Intel⁶ as its trademark new device for future technologies that attempt to extend Moore’s law for higher performance and lower leakage. They are quite similar to FinFETs

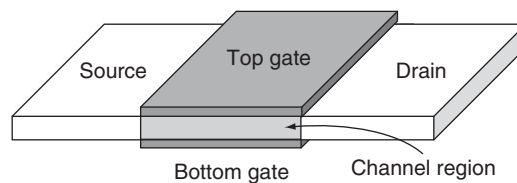


FIGURE 1.2 A planar double gate.

6 Chapter One

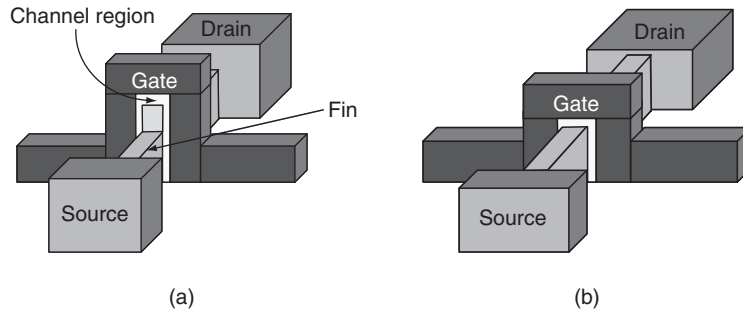


FIGURE 1.3 Structure of (a) FinFET and (b) Tri-gate.

and are also considered as nonplanar vertical multigate devices (see Figure 1.3(b)). These devices provide increased surface area for the channel region, thus creating higher drive currents by wrapping a single gate in place of multiple gate structures.^{7,8}

The new multigate devices described here provide greater control over gate threshold voltage and also increase the surface area of the channel for improved drive current, thereby producing faster chips.

1.1.1.3 Nanodevices

Nanodevices are created using materials other than silicon. These materials are used to realize nonconventional devices capable of mimicking the operation of a MOSFET. Not only are nanodevices an order of magnitude smaller than conventional MOSFETs produced today, they are also unique in terms of materials and the manufacturing technology used. MOS transistors operate on the basis of movement of charge carriers across the channel region, whereas the operation of nanodevices is based on quantum mechanical principles. Nanodevices can be classified, in terms of their working mechanism, as molecular and solid-state devices. *Molecular* devices use a single molecule (or a few molecules) as the switching device, and they can be as small as 1 nm.⁹ Examples include switches using catenanes¹⁰ or rotaxanes¹¹ as well as DNA-strand-based devices.¹² Molecular computing systems are highly sensitive to electrical and thermal fluctuations. They also require large-scale changes to current design practices in order to accommodate significantly higher failure rates and power constraints.

Solid-state nanodevices that have been investigated with an eye toward forming logic circuits of reduced density include (1) carbon nanotubes, (2) quantum dots, (3) single-electron transistors, (4) resonant tunneling devices, and (5) nanowires. Without delving into details of these devices, we list the underlying mechanism of conduction in each. Carbon nanotubes and silicon nanowires are further along in terms of manufacturing developments. These devices use “ballistic transport” (i.e., the movement of electrons and holes are

unhindered by obstructions) of charge carriers as the charge-conducting mechanism. Quantum dots interact with each other based on Coulomb forces but without actual movement of electrons or holes.¹³ Resonant tunneling diodes exhibit negative differential resistance characteristics when a potential is applied across the device, so they can be used to build ultrahigh-speed circuitry.^{14,15} Finally, single-electron transistors are three-terminal devices whose operation is based on the “Coulomb blockade,” a quantum effect whereby the gate voltage determines the number of electrons in a region.¹⁶ Nonconventional manufacturing techniques (i.e., not based on lithography) are being used to reduce the cost of fabricating these devices, but large-scale manufacturing is not yet practical.

1.1.2 Contributions from Material Science

Except for SOI, all the new devices described so far are still in their nascent stages and have not yet been manufactured in large quantities. In order to obtain consistent results while scaling transistors, process improvements were made in the materials domain. New materials that target specific process stages and device regions have been suggested either to improve performance or reduce leakage and so allow extension of CMOS scaling. Examples include strained silicon materials, low- K and high- K dielectrics, metal gates, and copper conductors. A discussion of the purpose and properties of these developments is presented next.

1.1.2.1 Low- K and High- K Dielectrics

Dielectrics (oxides) form an integral part of CMOS IC manufacturing today. A liner composed of silicon dioxide (SiO_2) or silicon oxynitride (SiON) has traditionally been at the heart of transistor operation, providing high impedance control for the conduction path between source and drain of a transistor. Similarly, SiO_2 is used as a barrier layer between active regions of devices and also between layers of metal interconnects. The drive current (and hence the speed) of a transistor is directly proportional to the gate oxide capacitance. The gate oxide capacitance depends on the oxide thickness t_{ox} and the dielectric constant ϵ_{ox} of the material used as oxide:

$$I_D = \mu C_{\text{ox}} \frac{W}{L} V; \quad C_{\text{ox}} = \frac{\epsilon_{\text{ox}}}{t_{\text{ox}}} \quad (1.1)$$

In order to increase the oxide capacitance, the thickness of the oxide is scaled in tandem with transistor scaling until the oxide thickness reaches only a few layers of molecules. Oxide thickness is

8 Chapter One

measured both in optical/microscopy and electrical terms. Because of various field effects, thickness as measured in terms of capacitance tends to be slightly higher than thickness observed via microscopy. At a thickness of about 20 Å, tunneling leakage through gate oxide becomes a serious problem (see Figure 1.4).¹⁷ Clearly, the tunneling current for SiO₂ is many orders of magnitude higher than other gate oxides at this thickness.

Thicker oxides are required to reduce tunneling leakage. However, thicker oxides reduce both gate capacitance and transistor drive current. This necessitates a high-*K* oxide material to maintain higher gate capacitance with thicker oxides. Hafnium oxide ($\epsilon=25$) has reportedly been used as a high-*K* gate dielectric. With the introduction of high-*K* gate oxides, threshold voltages tend to increase significantly, which is addressed by changing the gate electrode materials. Unlike conventional SiO₂ gates, high-*K* gate dielectrics need metal gates and a complex gate stack structure. Required properties of high-*K* materials include (but are not limited to) high dielectric constant, low leakage current density, small flat-band voltage shifts, low concentration of bulk traps, and reliability comparable to that of current SiO₂ dielectrics. Table 1.1¹⁸ lists some high-*K* dielectrics along with their dielectric constants and compatibility with silicon substrate. The crystal structure and stability of the Si substrate together determine possible defect levels for a given type of oxide.

As the transistor count increases, the interconnect length increases even faster. In today's chips, interconnect capacitance dominates gate

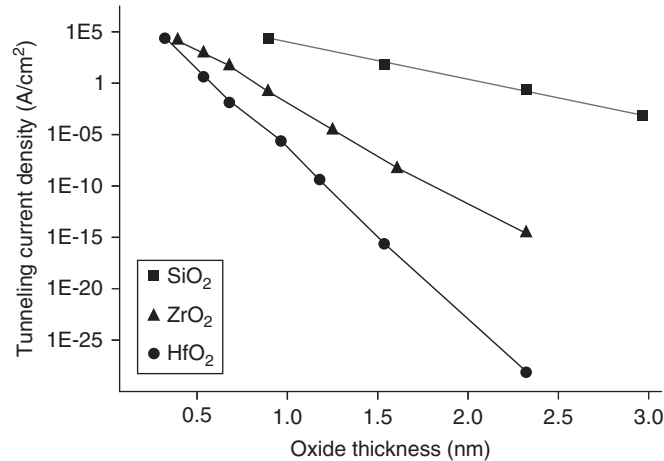


FIGURE 1.4 Gate oxide tunneling currents for three oxide types at different oxide thicknesses.

Material	ϵ	Crystal structure	Stable on silicon?
SiO ₂	3.9	Amorphous	Yes
Si ₃ N ₄	7.8	Amorphous	Yes
Y ₂ O ₃	15	Cubic	Yes
TiO ₂	80	Tetragonal	No
HfO ₂	25	Monoclinic, tetragonal	Yes
Ta ₂ O ₅	26	Orthorhombic	No
Al ₂ O ₃	9	Amorphous	Yes

TABLE 1.1 High-K oxide materials with dielectric constants and silicon substrate compatibility

capacitance and is the greatest source of active power dissipation. Because power has emerged as the most significant barrier to transistor usage, reducing power dissipation has become a shared goal for both process and design engineers. At manufacturing level, this is addressed by reducing the dielectric constant for the dielectric material between metal layers, which directly reduces the interconnect capacitance and contributes to power reduction. Organic materials and porous SiO₂ have been explored as possible alternatives. The interlayer dielectric (ILD) must meet a thermal specification to transport heat effectively; today, $K \leq 2.5$ is used.

1.1.2.2 Strained Silicon

The movement of charge carriers, such as electrons in an n -channel device and holes in a p -channel device, cause current to flow from the source to the drain of the transistor. Under the influence of an electric field, the speed at which the carriers move is called *mobility*. It is defined as $\mu = v/E$ where v is the velocity of charge carriers and E is the applied electric field. The strength of the drain current (I_D) is proportional to the mobility (μ_n, μ_p) of the carriers. This mobility is a function of temperature and is also a function of crystal stress. The latter property is used by modern processes to improve mobility by straining the atoms in the channel. *Straining* refers to the technique through which the interatomic distance between the atoms in the channel is increased or decreased. This causes an increase in the mean-free path of the charge carriers present in the channel. For nMOS devices, a tensile stress improves electron mobility; for pMOS devices, a compressive stress improves hole mobility. The source and drain regions of the nMOS transistor are doped with silicon-germanium atoms to induce a tensile stress on the channel region.

A layer of compressive or tensile nitride liner can be applied over the gate region. Typically, this is done at a higher temperature. Differential coefficients of thermal expansion produce strain upon cooling. Higher stress may contribute to crystal defects and reliability issues. Higher mobility may also contribute to increased transistor leakage. These factors limit the amount of stress that can be applied to the device. See Sec. 3.6 for more details on strain engineering.

1.1.3 Deep Subwavelength Lithography

Manufacturing small MOSFET devices and interconnect wires today requires printing of polygons that can have feature widths of less than a quarter wavelength of the light source. Photolithography is at the heart of the semiconductor manufacturing process; it involves multiple steps that lead to the formation of device and interconnect patterns on the wafer. Without photolithography it would not have been possible to assemble billions of transistors on a single substrate. A simple photolithography setup involves an illumination system consisting of a UV light source; a mask that carries the design patterns; the projection system, which comprises a set of lenses; and the wafer. With everyday lighting equipment, an object whose width is smaller than the wavelength of the light being used to project it will not be projected with good resolution and contrast. The resolution of the patterns being printed is defined as the minimum resolvable feature on the wafer for a given illumination source and projection system parameters. Thus the *resolution* R depends on the numerical aperture (NA) of the lens system and the wavelength λ of the light source. The equation that describes this relation is known as *Rayleigh's equation* and is given as follows:

$$R = k_1 \frac{\lambda}{\text{NA}} \quad (1.2)$$

The minimum resolvable linewidth on a particular mask is also referred to as the critical dimension (CD) of the mask. The numerical aperture of a lens system is the largest diffraction angle that a lens system can capture and use for image formation. Mathematically, it is the sine of the maximum angle incident on the lens multiplied by the refractive index (n) of the medium:

$$\text{NA} = n \sin \theta \quad (1.3)$$

Because air is the medium in optical systems, the limit value of numerical aperture is 1. Manufacturing limitations are such that the NA limit has not been achieved. But new inventions using water as medium have increased the NA above 1, since the refractive index of water is higher than that of air. Numerical aperture will continue to

play an important role because the higher the NA, the better the resolution of the system.

Another important parameter that controls the robustness of patterns printed on wafer is the depth of focus (DOF), which is defined as the maximum vertical displacement of the image plane such that an image is printable within the resolution limit. This is the total range of focus that can be allowed if the resulting printed image on the wafer is to be kept within manufacturing specifications. The maximum vertical displacement is given by

$$DOF = k_2 \frac{\lambda}{NA^2} \tag{1.4}$$

Since this focus tolerance depends inversely on the square of the numerical aperture, there is a fundamental limit on extremely high NA processes. Improvement in resolution means a reduction in R in Eq. (1.2). One way to improve resolution is to use a light source with wavelength less than or equal to the required minimum feature width of the mask. Figure 1.5¹⁹ shows the historical trend in wavelength of the illumination system. For a light source to be used in lithography, it should be of single frequency with nearly no out-of-band radiation, coherent in phase, and flicker-free with minimum dispersion. In addition, the lens system must be available to focus light at that frequency. Ordinary glass tends to be opaque to UV rays and is unsuitable for lithography²⁰.

Another method for improving resolution is to reduce the k_1 factor, which depends on processing technology parameters. This factor can be written as

$$k_1 = R_{\text{half-pitch}} \frac{NA}{\lambda} \tag{1.5}$$

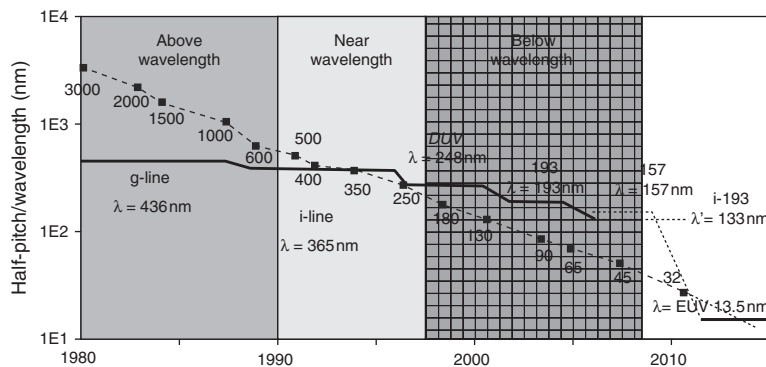


FIGURE 1.5 Progress trend of illumination light source across technology generations.

Here R is defined by the patterning rules of the technology. It is typically referred to as the *minimum half-pitch* used in the technology node or process. Factors that determine k_1 are the imaging system's numerical aperture (NA), wavelength (λ), and half-pitch (R). As seen in Figure 1.6, the k_1 factor has been progressively reduced by technology scaling to produce smaller features on the mask. With current technology using 193-nm light source for printing 45-nm features, the theoretical limit for the k_1 factor can be obtained from Eq. (1.5) as $k_1=0.25$. The theoretical limit is calculated by assuming a value of 1 for the numerical aperture. In practice, however, attaining a k_1 anywhere near 0.25 with the current single-exposure systems requires the use of high-index fluids having NA close to or greater than 1. Another approach is to use light sources of smaller wavelengths. These two options are still being investigated, and neither has been shown to perform reliable image transfer.²¹

Double-pattern lithography has been seen as a viable technique to improve the k_1 factor below 0.25. This is accomplished by increasing the pitch size while holding constant the minimum resolvable dimension of patterns. More details on double patterning are provided in Sec. 4.5.1.

1.1.3.1 Mask Manipulation Techniques

Resolution enhancement techniques (RETs) are methods used to improve the resolution of the lithography system by manipulating various parameters of the system. Most RETs aim to manipulate the patterns on the mask. The resolution of a feature being printed depends on neighboring features and the spacing between them.

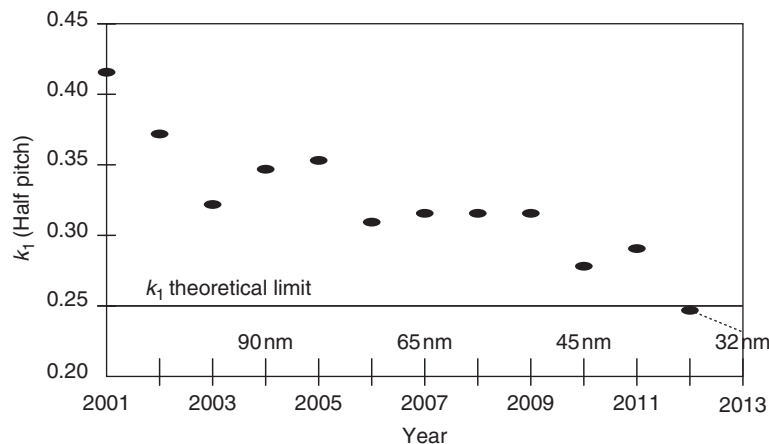


FIGURE 1.6 Trend in k_1 reduction for printability improvement.

The principle of diffraction governs the interaction of light waves that pass through the mask patterns on the mask while being projected onto the wafer. As shown in Figure 1.7, RET modifications to the mask improve the resolution of the features being printed.

Resolution enhancement techniques include optical proximity correction, phase shift masking, off-axis illumination, and multiple exposure systems. Optical proximity correction (OPC) changes the shape of the feature by adding extra jogs and serifs to improve the resolution (see Sec. 4.3.2). Phase shift masking (PSM) utilizes the superposition principle of light waves to improve resolution by creating phase changes in spaces between the features; see Sec. 4.3.3 for more details. Off-axis illumination (OAI) is based on the principle that, if the light rays are incident at an angle on the mask, then higher-order diffraction patterns can be made to pass through the lens and thereby improve resolution. This method and the type of lenses it uses are described in Sec. 4.3.4.

Another promising technique that improves the resolution of patterns being printed is multiple exposure systems. In one

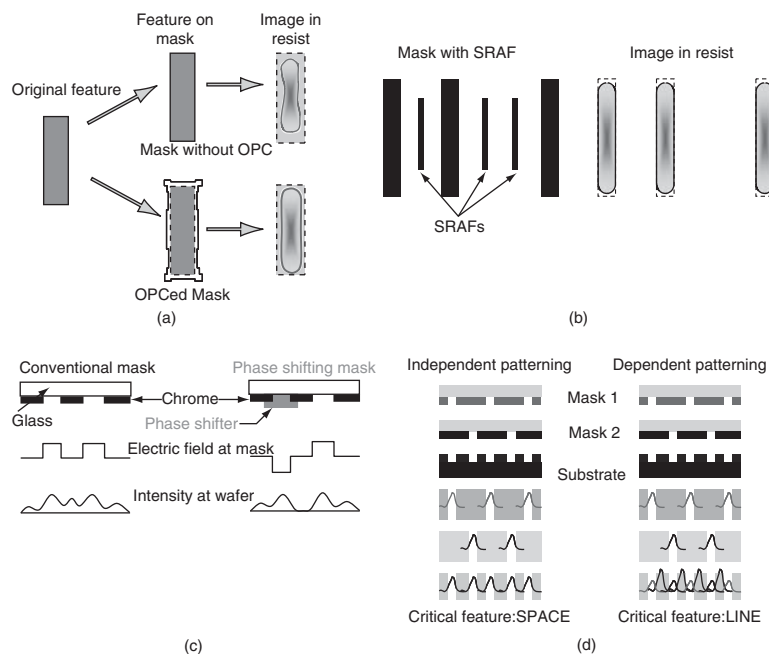


FIGURE 1.7 RET mask manipulation to improve pattern transfer: (a) optical proximity correction; (b) SRAF insertion; (c) phase shift masking; (d) double patterning.

embodiment of such a system, the pattern is carried by two separate masks that are exposed in separate steps, leading to the final image on wafer. This is known as double-pattern lithography, and it is being used to print critical masks by decomposing them into two masks. This method increases the spacing between metal lines and hence reduces the minimum resolvable feature that can be printed on the wafer; see Sec. 4.5.1 for details about this technique. Ongoing research seeks to use triple and quadruple patterning techniques to further push the resolution barrier. One negative consequence of such patterning systems is that fabrication throughput and the overall process yield may decrease, leading to an increase in product cost.

1.1.3.2 Increasing Numerical Aperture

As Eq. (1.2) indicates, increasing the system’s numerical aperture will improve the resolution. The numerical aperture is given by Eq. (1.3), where n is the refractive index of the medium between the projection system and the wafer and θ denotes the maximum angle of incidence for a ray passing through the projection lens. Air, with refractive index 1, is typically used as the medium. Figure 1.8¹ illustrates goals for future techniques that aim to improve NA. Immersion lithography (Figure 1.9) uses a liquid medium to increase the refractive index n . Water, with refractive index of 1.3, is currently being used as the immersion fluid, although other high-index fluids have been suggested as possible replacements. Immersion can lead to process issues such as spot defects from water molecules and error in handling wafers. Additional suggestions for improving numerical aperture include the use of high-index lenses, lenses with increased curvature, and high-index resist materials.

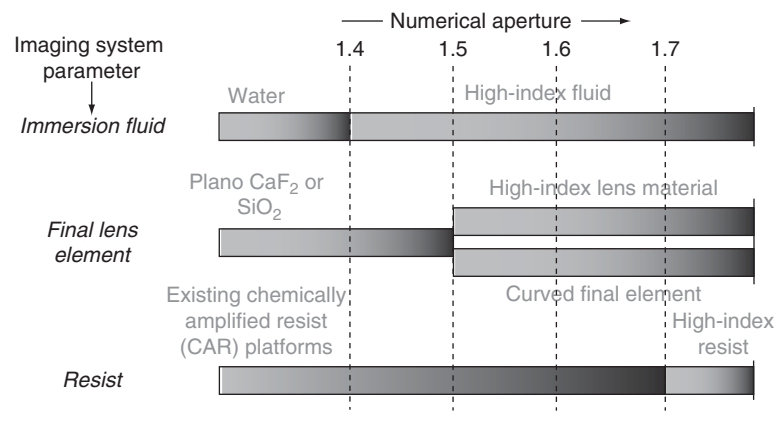


FIGURE 1.8 Future trends in techniques to improve numerical aperture.

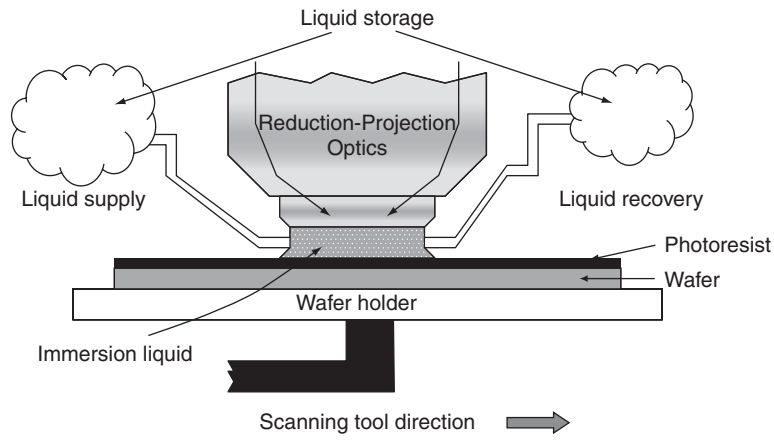


FIGURE 1.9 Immersion lithography technology.

1.2 Design for Manufacturability

Design for manufacturability (DFM) in the current context refers to the new design techniques, tools, and methodologies that ensure printability of patterns, control the parametric variation, and enhance yield. A broad definition of DFM encompasses various methodologies from the starting point of design specification to the product launch of an IC, which include circuit design, design optimization, mask engineering, manufacturing metrology—to name just a few technologies that aim to manufacture chips with repeatability, high yield, and high cost effectiveness. The two most important metrics by which all DFM methodologies are assessed are the cost of the entire process and the cumulative chip yield loss due to irregularities at various manufacturing steps.

With continued scaling, patterning is conducted almost at the resolution limit, while the transistor count is growing exponentially. When lithographic patterning is pushed to its limit, a single defect may invalidate a chip consisting of millions of transistors. This underscores the importance of DFM. Concerns about manufacturability have become so pervasive that DFM considerations—once dealt with entirely by design rules and mask engineering—are moving upstream in the design process. The next few sections will examine the economic value of DFM, current parameter uncertainties, traditional DFM approaches, and the requirement for model-based DFM techniques.

1.2.1 Value and Economics of DFM

In Sec. 1.1 it was noted that scaling from one technology node to the next often required changes in design methodologies, CAD tools, and

process technology. Design for manufacturability is one of those necessities that became a design concern with the advent of subwavelength lithography. Advances in this technology have translated into more intrusive changes in design methodology. Long ago, optical diffraction effects were handled through design rules check (DRC), and any remaining issues were handled in mask preparation. Since the introduction of subwavelength lithography, rule-based DRC has been supplanted by model-based DRC: simple rules have been replaced with quick and approximate optical simulation of layout topology. With an increase in the field of optical influence (aka optical diameter), the DRC models became more complex. The true value of DFM is not yet understood by many designers still using older process technologies. However, as designs move to 45 nm and below, DFM steps are becoming more critical in the design process. In 130-nm to 65-nm technologies, DFM issues were mostly handled by postprocessing of the layout (i.e., resolution enhancement techniques). But postprocessing alone cannot ensure manufacturability when deep subwavelength lithography is involved. If one-pass postprocessing proves insufficient, then iteration between physical layout generation and RET steps becomes necessary. When all is said and done, DFM methodologies increase the number of tools that are run through the design as well as the number of iteration cycles, thus increasing the design's time to tape out (TTTO). Designers are increasingly concerned about the fact that improving their designs using DFM tools requires more in-depth knowledge of the process techniques. Designers already juggle multiple design targets, which include area, performance, power, signal integrity, reliability, and TTTO. Of course, adding new design objectives will affect existing ones. Thus, the effectiveness of any DFM methodology must be judged in terms of its impact on other design objectives. In addition to the imaging system parameters already discussed and to the associated processes, DFM also pertains to gate CD and interconnect CD variations, random dopant fluctuations, mobility impacts, and other irregularities that are subjects of TCAD (technology CAD) studies. Once the design netlist rolls over to the DFM step (see Figure 1.10), the outcome may be (1) changes to the physical design netlist such that the desired parameters are within specifications; (2) information feedback to designers regarding areas of design where such DFM changes cannot be incorporated automatically; or (3) The parametric impact of DFM on the design process, including static timing analysis as well as signal integrity and reliability. With outcome (2), where a one-pass DFM step does not succeed, the remaining issues may be addressed by automatic layout tools or may require manual intervention. This is an area in which tools and methodologies are still evolving.

The first outcome results in a modified version of the design, incorporating anticipated postsilicon effects. The techniques used

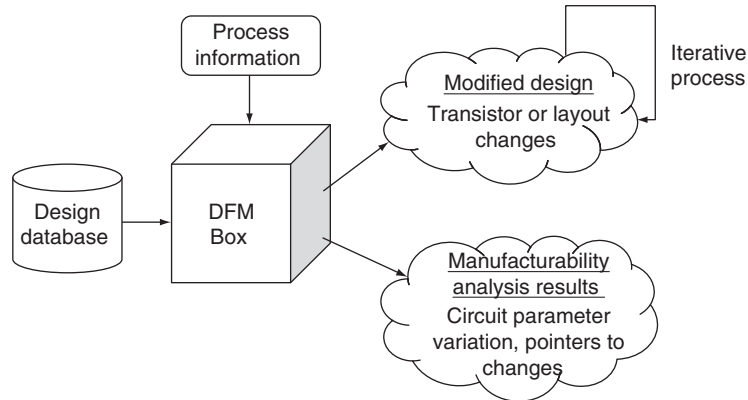


FIGURE 1.10 Purpose of design for manufacturability (DFM) methodologies.

include, for example, circuit hardening, resolution enhancement, and dummy fills. The third outcome involves analyzing the given design and providing design-specific parameter variability ranges. Instead of providing yield benefits with DFM-based suggestions, these results help designers perform more effective optimization with the promise that postsilicon circuit variability will be minimized and will fall within specifications.

Other than the value of DFM perceived by the designer, DFM also bears an important economic aspect. The economics of DFM aims to establish a cost-benefit metric for each DFM methodology by assessing its return on investment. Most such methodologies seek to improve the overall design yield by taking different approaches to optimizing parameters of the design and manufacturing process. As described by Nowak and Radojcic,^{22,23} the economics of DFM can be classified into three areas of potential profit or loss: (1) DFM investment cost; (2) design re-spin costs; and (3) DFM profit (see Figure 1.11).²³ The costs of investing in DFM tools and methodology are incurred during the phases of product concept, design, optimization, and tape out. The benefit of this investment is realized after tape out, with improvements in the yield and reliability of the chip. This means that the yield curve of a process that includes DFM-based methodologies is sharper.

Quantification of DFM benefits requires silicon feedback. Given the high cost of such direct observation techniques as microscopy, the benefits of DFM are usually assessed via indirect silicon feedback such as manufacturing yield and parametric yield. Indirect measurements are contaminated with multiple parameters, so decorrelating these parameters requires carefully constructed test structures. Design for manufacture is important not only for continual product yield improvement but also for enabling future technology nodes.

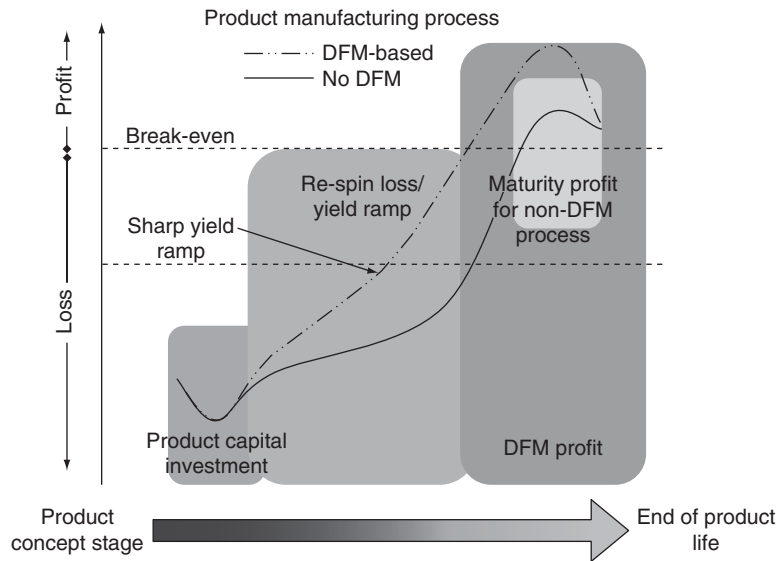


FIGURE 1.11 The value and economics of design for manufacturability (DFM).

1.2.2 Variabilities

Parametric variation has emerged as a major design concern. For *correct-by-construction* methodology, circuit models need to be accurate and model parameters need to be correct; otherwise, the behavior of the design cannot be reliably predicted before construction. In reality, a design may vary from model parameters owing to variations in manufacturing process parameters. Current designs may consist of billions of transistors, so when these variations become large there is always the possibility of circuit failure, which can significantly reduce yield. Also, current design practice is to assume that the underlying hardware continues to be correct during the product lifetime. However, the relentless push for smaller devices and interconnects has moved the technology closer to a point where this design paradigm is no longer valid.²⁴⁻²⁷ For example, with the advent of 90-nm technology, negative bias temperature instability (NBTI) became a major reliability concern,²⁸ since a pMOS device degrades continuously with voltage and temperature stress. For nMOS devices fabricated using 45-nm technology, positive bias temperature instability (PBTI) is likewise becoming a concern.²⁹ Windows XP failure data compiled by Microsoft Research also points to increased occurrences of hardware failures.³⁰ According to ITRS, these problems

are expected to worsen in future technologies,³¹ as designs are more likely to experience failure due to what designers call PVT issues—that is, process corner, voltage, and temperature issues.

Table 1.2 categorizes these variations from both a source and impact point of view. Columns 1 and 2 form the first source and effect relationship for variations in semiconductor manufacturing processes. As mentioned previously, variations in the manufacturing process lead to variations in the properties of the device and interconnect. Manufacturing variations can be categorized as irregularities in equipment and processing, such as in lithography, and chemical processing. Other sources of variations include mask imperfections caused during mask manufacturing, mask mishandling, tilting, and alignment issues. Additional sources of variation are improper focal position and exposure dose of the imaging system and variation in photoresist thickness. Sources of device and interconnect variation include such process steps as dopant implant onto the source, drain, or channel regions of devices on the wafer and planarization of metal lines and dielectric features during chemical-mechanical polishing (CMP).

The effects of such manufacturing variations are observed through changes in the circuit parameters. The most important among them are the parameters of the active devices, notably transistors and diodes. Variations in circuit parameters have engendered several modeling and analysis techniques that attempt to predict parameter behavior after fabrication. Among physical features, circuit

Manufacturing process	Circuit parameters	Circuit operation	CAD analysis
Mask imperfections	Channel length	Temperature	Timing analysis
Alignment, tilting	Channel width	Supply voltage	RC extraction
Focus, dosage	Threshold voltage	Aging, PBTI/NBTI	I-V curves
Resist thickness, etch	Overlap capacitance	Coupling capacitance	Cell modeling
Doping	Interconnects	Multiple input switching	Process files
Chemical mechanical polishing			Circuit simulations

TABLE 1.2 Variations in IC Manufacturing and Design: Sources and Impacts

performance is more sensitive to postlithography channel length and interconnect width than any other. Consequently, they are known as critical dimensions (CD). Poly-CD variation leads to change in effective channel length. Circuit delay tends to increase linearly with increasing channel length, whereas leakage current tends to increase exponentially with decreasing channel length. As shown in Figure 1.12, a 10 percent variation in gate CD induces large variation in threshold voltage (V_T) and delay. Because the V_T of devices can fall below the minimum allowable for leakage control, poly-CD control has become a critical aspect of overall process control. Interconnect CD variation leads to changes in path delay, coupling capacitance effects, increased susceptibility to electromigration, and spot defects.

Although the manufacturing variations have always existed and the manufacturing tolerances have generally improved with successive generations of technology, the impact on circuit parameters has been otherwise. This is reflected in terms of wider variation in circuit performance and power dissipation due to leakage.

Finally, manufacturing sources of variation—which include mask imperfections, wafer handling, alignment, and tilting—lead to errors in overlay and dielectric thickness. Focus, dose, and resist thickness variation are factors that lead to CD variations on wafer. Many modeling methods apply statistical techniques to predict the effect of

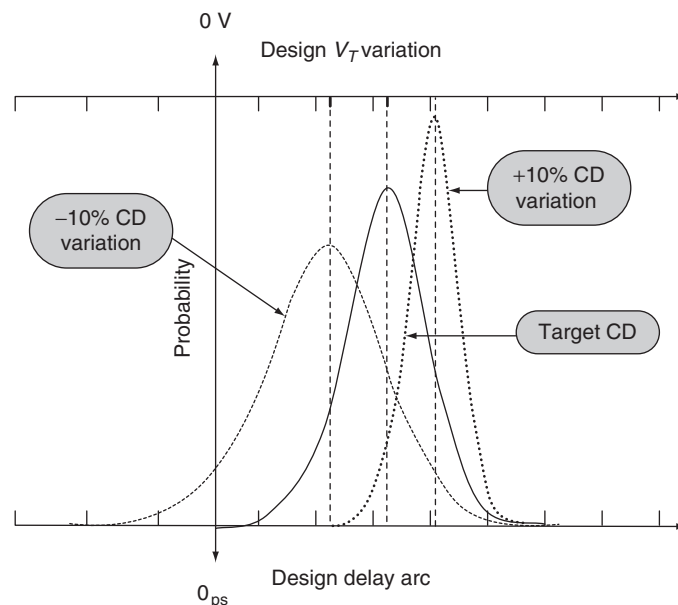


FIGURE 1.12 Design V_T and delay variation due to change in critical dimensions (CD).

these variations on electrical parameters, layout printability, and die yield. The etching process is used to remove parts of the material not covered by protective layers. It can lead to pattern fidelity issues because wet, chemical, and plasma etch processes cannot be error-free. The most important effects of etching problems are line edge roughness (LER), which refers to the horizontal deviation of the feature boundary, and line width roughness (LWR), which refers to random deviation in width along the length of the polygon. One effect of LER on transistor is changes in threshold voltage V_T , as shown in Figure 1.13.³² Fluctuation in dopant density also induces such variation in V_T . Figure 1.14³³ illustrates three devices that have an equal number of dopant atoms in the channel but have different V_T values. Chemical-mechanical polishing is used to planarize the wafer after deposition

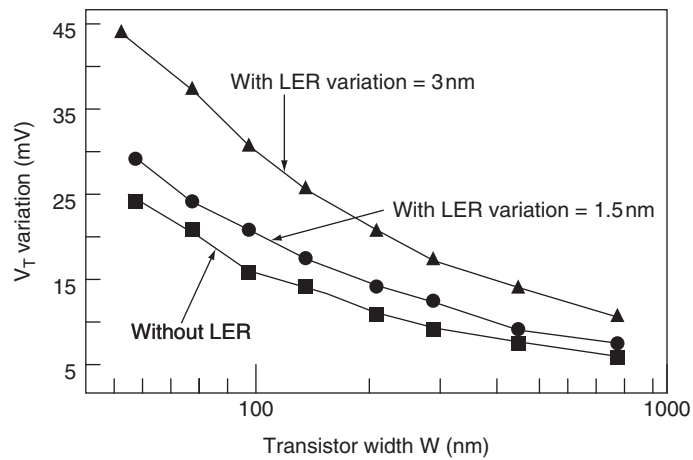


FIGURE 1.13 V_T variation due to LER (produced with 45-nm gates using predictive technology models).

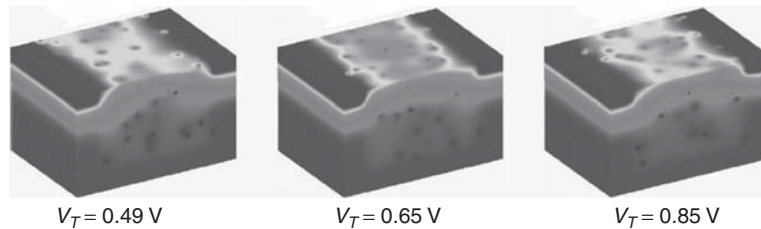


FIGURE 1.14 V_T variation due to random dopant fluctuation (RDF).

of the metal and dielectric layer material. Pattern density on the wafer causes CMP to create surface roughness, defined as vertical deviation of the actual surface from an ideal one. Such changes in the surface lead to focus changes during subsequent lithography steps, contributing to further CD variation (see Figure 1.15).

Circuit operation can be affected by several sources other than variation in manufacturing process and circuit parameters. For example, environmental factors, which include supply voltage and temperature variation, affect the amount of current that flows through a device. Temperature has an effect on circuit reliability (i.e., aging). Circuit reliability effects, such as electromigration, NBTI, and hot carrier degradation, change interconnect and gate delays over time. These effects are related to interconnect width and thickness, which in turn depend on the effectiveness of patterning and CMP (respectively). Thus, a link can be seen between physical design, patterning structures in the surrounding regions, and the circuit aging process.

At each step of the circuit realization process, CAD tools are used to predict circuit performance. As the realization gets closer to the transistor and physical levels, the model parameters become progressively more accurate to better predict circuit performance. Initial performance prediction models do not consider variation. In a typical design environment, interconnect RC extraction may be based on nominal process parameters, while transistor models may take parametric variation into account. Subsequent to manufacturing, if silicon fails to meet performance expectations, such unlisted variations have been identified as sources of errors. Considering all the possible sources of variations is an expensive proposition in terms of design optimization and timely convergence of design. Thus, a company must be constantly evaluating new techniques for its DFM arsenal.

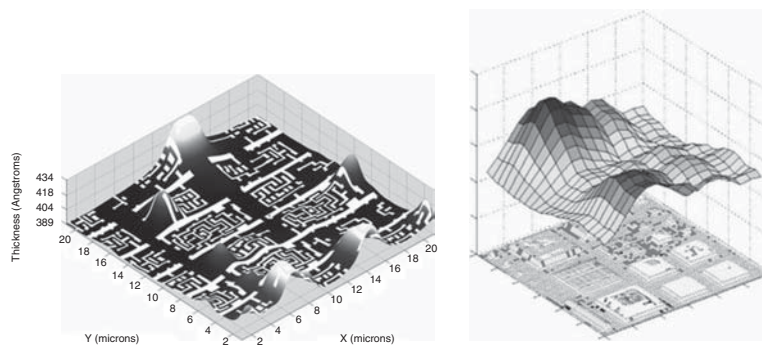


FIGURE 1.15 Design-dependent chip surface undulations after CMP. (Courtesy of Cadence Design Systems.)

1.2.3 The Need for a Model-Based DFM Approach

Design for manufacturing has been in use since the late 1990s. Traditional DFM relied on design rules and guidelines for the polygons and shapes present in an IC layout. Rules suggested by tools were based on interaction between two adjacent metal line or two adjacent poly line features. If a design layout passed all specified design rules and abided by all suggested guidelines, then it was set to produce a high yield. All traditional DFM methodologies were applied at the full chip level, with corner-based functional analysis and parametric yield-loss analysis predominating.

With the advent of subwavelength lithography, design rules check alone is not sufficient to ensure high yield. This fact has been chiefly attributed to the printability problems introduced by subwavelength lithography. Printing of features whose width is less than half the wavelength of the light source creates diffraction-induced pattern fidelity issues. The interaction between polygons has been found to extend well beyond adjacent features. This region of influence on neighboring features is called *optical diameter*. As the number of polygons increase, it is impossible to bring about rule checks for each type of polygon-polygon interaction. The number of DRC rules has increased exponentially to a point where it has become virtually impossible to produce an optically compliant layout by rule based DRC alone. Since the introduction of subwavelength lithography, rule-based DRC has been supplanted by model-based DRC, wherein simple rules are replaced with quick and approximate optical simulation of layout topology. As the optical diameter increased, these models became more complex. Because of this complexity, model-based DFM methodologies typically limit themselves to smaller regions of the circuit. These models have evolved over time to incorporate multiple effects, including diffraction, CMP-induced wafer surface modulations, random dopant fluctuations, and LER.

As fabrication moves into the 32-nm technology node, layout modifications based on phase shift masking and double patterning will have to consider interactions of second and third alternative neighbors. Another new aspect for DFM methodologies is the need for model-based techniques to predict DFM impact on timing, power, signal integrity, and reliability issues. There is also a need to provide early-design-stage feedback so that correct circuit operation within the process variation window is assured. Standard cell methodologies today incorporate model-based postlithography analyses that yield highly compact, printable, and functional cells on silicon.

1.3 Design for Reliability

Transistors and interconnects are known to fail during their life time under circuit operations. Some of the known failure mechanisms

include gate-oxide shorts, interconnect voids or blobs caused by electromigration, V_T shift during the lifetime of transistor operation that is due to negative and positive bias temperature instability, and other mechanical, chemical, or environmental factors associated with manufacturing. When such failures are modeled correctly, product lifetime can be improved by design changes that involve device and interconnect sizing and well as floorplanning to reduce thermal hotspots. Collectively, this process is known as design for reliability (DFR). Although DFR is distinct from DFM, the two may be integrated from the perspective of design methodology because the correction mechanisms are similar.

Design for reliability comprises the techniques, tools, and methodologies employed to analyze, model, and predict the reliability of a given device or circuit. Reliability parameters are known to evolve over process maturity. Nonetheless, it is important to establish a relation between circuit parameters and product reliability so that clear targets can be set during the DFR process. The reliability models use information about failure mechanisms and how they relate to circuit design parameters in order to model a product's reliability; the models aim to predict the mean time to failure (MTTF) of a device. The MTTF is a function of manufacturing parameters and also of the parameters associated with circuit operation, such as the device's supply voltage and temperature.

Design for reliability is an exercise in circuit and layout sizing, floorplanning, and implementing redundancies to address failures. Redundancies could be added at the circuit, information, time, and/or software levels.^{30,31} Dual-rail encoding and error-correcting codes are examples of information redundancy. Spare circuits and modules are examples of circuit redundancy. Modern memories often incorporate spare rows and columns to improve yield; in addition, spare processor cores, execution units, and interconnects have been used in commercial circuits. Multisampling latches enable time redundancy, and software redundancy includes redundant multi-threads (RMT); many of these features are now found in commercial systems.

1.4 Summary

An effective DFM-DFR methodology provides early feedback to the design during its nascent stage. In this chapter we introduced the reader to current trends in the design of nanoscale CMOS and very large-scale integration (VLSI) circuits, explaining the various changes that have been incorporated toward the end of achieving the two principal goals of higher performance and lower power consumption. We also provided a brief overview of new device structures in the 22-nm technology node that have been touted as replacements for traditional MOSFET devices. We discussed the role of material science

and optics in improving device operation, printability, and design reliability. Also discussed were the applicability of DFM in the presence of process and design parameter variability as well as the process of integrating design and manufacture. We examined the need for newer, model-based DFM methodologies given the use of subwavelength lithography and higher density of layout patterns. Finally, we mentioned some important reliability concerns in nanoscale CMOS VLSI design and described the DFR-based CAD tools that can help increase the anticipated lifetime of designs. In short, we have described the trends in technology and the rising importance of DFM and DFR.

References

1. "Lithography," in *International Technology Roadmap for Semiconductors Report*, <http://www.itrs.net> (2007).
2. K. Bernstein and N. J. Rohrer, *SOI Circuit Design Concepts*, Springer, New York, 2000.
3. K. K. Young, "Analysis of Conduction in Fully Depleted SOI MOSFETs," *IEEE Transactions on Electron Devices* **36**(3): 504–506 (1989).
4. D. Hisamoto, T. Kaga, Y. Kawamoto, and E. Takeda, "A Fully Depleted Lean-Channel Transistor (DELTA)—A Novel Vertical Ultra-Thin SOI MOSFET," in *Technical Digest of IEEE International Electron Device Meeting (IEDM)*, IEEE, Washington, DC, 1989, pp. 833–836.
5. X. Huang, W. C. Lee, C. Kuo, D. Hisamoto, L. Chang, J. Kedzierski, E. Anderson, et al., "Sub-50nm FinFET:PMOS," in *Technical Digest of IEEE International Electron Device Meeting (IEDM)*, IEEE, Washington, DC, 1999, pp. 67–70.
6. R. S. Chau, "Integrated CMOS Tri-Gate Transistors: Paving the Way to Future Technology Generations," *Technology @ Intel Magazine*, 2006.
7. F. L. Yang, D. H. Lee, H. Y. Chen, C. Y. Chang, S. D. Liu, C. C. Huang, T. X. Chung, et al., "5 nm-gate nanowire FinFET," in *Technical Digest of IEEE Symposium on VLSI Technology*, IEEE, Dallas, TX, 2004, pp. 196–197.
8. B. Doyle, B. Boyanov, S. Datta, M. Doczy, J. Harelend, B. Jin, J. Kavalieros, et al., "Tri-Gate Fully-Depleted CMOS Transistors: Fabrication, Design and Layout," in *Technical Digest of IEEE Symposium on VLSI Technology*, IEEE, Yokohoma, 2003, pp. 133–134.
9. K. Sandeep, R. Shukla, and I. Bahar, *Nano, Quantum and Molecular Computing: Implications to High Level Design and Validation*, Springer, New York, 2004.
10. C. P. Collier, G. Mattersteig, E. W. Wong, Y. Luo, K. Beverly, J. Sampaio, F. M. Raymo, et al., "A Catenana-Based Solid State Electronically Reconfigurable Switch," *Science* **289**(5482): 1172–1175, 2000.
11. W. R. Dichtel, J. R. Heath, and J. F. Stoddart, "Designing Bistable [2]Rotaxanes for Molecular Electronic Devices," *Philosophical Transactions of the Royal Society A: Mathematical Physical and Engineering Sciences* **365**: 1607–1625, 2007.
12. N. B. Zhitenev, H. Meng, and Z. Bao, "Conductance of Small Molecular Junctions," *Physical Review Letter* **88**(22): 226801–226804, 2002.
13. C. S. Lent, P. D. Tougaw, and W. Porod, "Quantum Cellular Automata: The Physics of Computing with Arrays of Quantum Dot Molecules," in *Proceedings of Physics and Computation*, IEEE, Dallas, TX, 1994, pp. 5–13.
14. J. P. Sun, G. I. Haddad, P. Mazumder, and J. N. Schulman, "Resonant Tunneling Diodes: Models and Properties," *Proceedings of the IEEE* **86**(4): 641–660, 1998.
15. K. Ismail, B. S. Meyerson, and P. J. Wang, "Electron Resonant Tunneling in Si/SiGe Double Barrier Diodes," *Applied Physics Letters* **59**(8): 973–975, 1991.
16. K. K. Likharev, "Single-Electron Devices and Their Applications," *Proceedings of the IEEE* **87**(4): 606–632, 1999.

17. F. Sacconi, J. M. Jancu, M. Povolotskyi, and A. Di Carlo, "Full-Band Tunneling in High- κ Oxide MOS Structures," *IEEE Transactions on Electron Devices* **54**(12): 3168–3176, 2007.
18. S. K. Ray, R. Mahapatra, and S. Maikap, "High-k Gate Oxide for Silicon Heterostructure MOSFET Devices," *Journal of Material Science: Material in Electronics* **17**(9): 689–710, 2006.
19. R. Doering, "Future Prospects for Moore's Law," in *High Performance Embedding Computing Workshop*, MIT Press, Lexington, MA, 2004.
20. Harry J. Levinson (ed.), *Principles of Lithography*, SPIE Press, Bellingham, WA, 2005.
21. Chris A. Mack (ed.), *Fundamental Principles of Optical Lithography*, Wiley, West Sussex, U.K., 2007.
22. M. Nowak, "Bridging the ROI Gap between Design and Manufacturing," *SNUIG*, Santa Clara, CA, 2006.
23. M. Nowak and R. Radojic, "Are There Economic Benefits in DFM?" in *Proceedings of Design Automation Conference*, IEEE, Anaheim, CA, 2005, pp. 767–768.
24. S. Y. Borkar, "Designing Reliable Systems from Unreliable Components: The Challenges of Transistor Variability and Degradation," *IEEE Micro* **25**(6): 10–16, 2005.
25. J. M. Carulli and T.J. Anderson, "Test Connections—Tying Application to Process," in *Proceedings of International Test Conference*, IEEE, Austin, TX, 2005, pp. 679–686.
26. P. Gelsinger, "Into the Core...," *Stanford EE Computer Systems Colloquium*, <http://www.stanford.edu/class/ee380/Abstracts/060607.html> (2006).
27. J. Van Horn, "Towards Achieving Relentless Reliability Gains in a Server Marketplace of Teraflops, Laptops, Kilowatts, & 'Cost, Cost, Cost' ... (Making Peace between a Black Art and the Bottom Line)," in *Proceedings of the International Test Conference*, IEEE, Austin, TX, 2005, pp. 671–678.
28. M. Agostinelli, S. Pae, W. Yang, C. Prasad, D. Kencke, S. Ramey, E. Snyder, et al., "Random Charge Effects for PMOS NBTI in Ultra-Small Gate Area Devices," in *Proceedings of International Reliability Physics Symposium*, IEEE, San Jose, CA, 2005, pp. 529–532.
29. M. Denais et al., "Interface Trap Generation and Hole Trapping under NBTI and PBTI in Advanced CMOS Technology with a 2-nm Gate Oxide," *IEEE Transactions on Device and Materials Reliability* **4**(4): 715–722, 2004.
30. B. Murphy, "Automating Software Failure Reporting," *ACM Queue* **2**(8): 42–48, 2004.
31. G. Groseneken, R. Degraeve, B. Kaczer, and P. Rousel, "Recent Trends in Reliability Assessment of Advanced CMOS Technology," *Proceedings of IEEE 2005 International Microelectronics Test Structure* **18**: 81–88, 2005.
32. W. Zhao and Y. Cao, "New Generation of Predictive Technology Model for Sub-45nm Design Exploration," *IEEE Transactions on Electron Devices* **53**(11): 2816–2823, 2006.
33. A. Asenov, A. R. Brown, J. H. Davies, S. Kaya, and G. Slavcheva, "Simulation of Intrinsic Parameter Fluctuations in Decanometer and Nanometer-Scale MOSFETs," *IEEE Transactions on Electron Devices* **50**(9): 1837–1852, 2003.